



Universidade de Brasília
Instituto de Exatas – IE
Estágio Supervisionado 2

Estudo sobre o perfil dos estudantes ingressantes da Universidade de Brasília em 2012

Ilka Oliveira Torres

Relatório Final

Orientadora: Prof^a. Maria Teresa Leão Costa

**Brasília
Junho de 2013**

ILKA OLIVEIRA TORRES

**ESTUDO SOBRE O PERFIL DOS ESTUDANTES INGRESSANTES DA
UNIVERSIDADE DE BRASÍLIA EM 2012**

Trabalho apresentado junto ao Curso de Estatística da Universidade de Brasília, para conclusão do curso de Bacharelado em Estatística.

Orientadora: Prof^a. Maria Teresa Leão Costa.

**Brasília
Junho de 2013**

Agradecimentos

Agradeço à minha família, pelo apoio prestado durante a graduação, principalmente à minha tia, com quem vivi a maior parte da graduação, e à minha irmã que esteve presente incentivando e motivando na reta final desta jornada.

Agradeço também ao grupo de pesquisa do Observatório da Vida Estudantil da Universidade de Brasília pelo fornecimento da base de dados utilizada no presente trabalho. E por fim, à professora e orientadora Maria Teresa Leão Costa, com quem tive a honra de cursar a primeira disciplina do curso de Estatística.

Sumário

1. INTRODUÇÃO	1
2. OBJETIVOS	2
2.1. Objetivo Geral	2
2.2. Objetivos Específicos	2
3. REFERENCIAL TEÓRICO	3
3.1. A Universidade de Brasília	3
3.1.1. Os <i>Campi</i>	4
3.1.2. Formas de admissão	5
3.2. Análise de Dados Categorizados	6
3.2.1. Comparação entre categorias	7
3.3. Regressão Logística	10
3.3.1. Regressão Logística Simples	10
3.3.2. Interpretação do Modelo	11
3.3.3. Inferência para Modelos de Regressão Logística	13
3.3.4. Intervalo de confiança	14
3.3.5. Seleção de variáveis do modelo	14
3.3.6. Construção do Modelo de Regressão	16
4. METODOLOGIA	17
5. ANÁLISE DESCRITIVA	21
5.1. Perfil do aluno por semestre	21
5.2. Perfil do aluno por <i>Campus</i>	27
5.3. Perfil por Turno do Curso	34
5.4. Perfil do aluno segundo Sistema de Ingresso	38
5.5. Perfil do aluno segundo Tipo de Escola que Concluiu o Ensino Médio	43
6. MODELAGEM	50
6.1. Primeiro semestre (1/2012)	51
6.2. Segundo semestre (2/2012)	55
7. CONCLUSÃO	60
REFERÊNCIAS BIBLIOGRÁFICAS	62

Lista de gráficos

Gráfico 1 - Distribuição das Instituições de Ensino Superior por Categoria Administrativa, Brasil - 2011	3
Gráfico 2 – Box-plot da idade dos alunos da UnB por semestre de ingresso - 2012	23
Gráfico 3 – Distribuição dos Registros de Alunos por Campus na Universidade de Brasília – 1/2012	25
Gráfico 4 – Distribuição percentual dos registros de alunos por campus e sexo na Universidade de Brasília - 2012	27
Gráfico 5 – Box-plot da idade dos alunos da UnB por campus - 2012	28
Gráfico 6 – Distribuição dos alunos da UnB por semestre de ingresso, <i>campus</i> e tipo de instituição que concluiu o ensino médio - 2012	32
Gráfico 7 - Box-plot da idade dos alunos da UnB por turno do curso - 2012	34
Gráfico 8 – Box-plot da idade dos alunos da UnB por sistema de ingresso e semestre - 2012	40
Gráfico 9 – Distribuição dos alunos por tipo de escola que estudou no ensino médio segundo o semestre de ingresso na UnB.	44
Gráfico 10 – Box-plot da idade dos alunos da UnB por natureza da institucional da escola que o aluno concluiu o ensino médio - 2012	46
Gráfico 11 – Resíduos de Pearson para o modelo ajustado em 1/2012	53
Gráfico 12 – Deviance Residual para o modelo ajustado em 1/2012	53
Gráfico 13 – DfBetas: Influência na estimativa do parâmetro nível de escolaridade da mãe: analfabeta	54
Gráfico 14 - Resíduos de Pearson para o modelo ajustado em 2/2012	57
Gráfico 15 – Deviance Residual para o modelo ajustado em 2/2012	57

Lista de tabelas

Tabela 1 - Pontuação atribuída pela ABEP a bens, facilidades e serviços para classificação econômica.....	18
Tabela 2 - Pontuação atribuída pela ABEP ao grau de instrução do chefe da família para classificação econômica.....	18
Tabela 3 - Pontuação atribuída a bens, facilidades e serviços para classificação econômica.....	19
Tabela 4- Pontuação atribuída ao grau de instrução do chefe da família para	19
Tabela 5 - Faixas de pontuação estabelecidas pela ABEP para classificação econômica.....	19
Tabela 6 – Percentual de alunos da UnB por semestre de ingresso segundo sexo, estado civil, cor, com quem reside e faixa etária - 2012.....	22
Tabela 7 - Percentual de alunos da UnB por semestre de ingresso segundo nível de escolaridade do pai e mãe, tipo de instituição que cursou o ensino médio, forma de ingresso e ocupação - 2012.	24
Tabela 8 – Percentual de alunos da UnB por semestre segundo tipo de instituição que concluiu o ensino fundamental e médio, forma e sistema de ingresso, escolha do curso e ocupação do aluno – 2012	26
Tabela 9- Percentual de alunos da UnB por semestre e campus segundo sexo, estado civil, cor e faixa etária - 2012.	29
Tabela 10 – Percentual de alunos da UnB por semestre e <i>campus</i> segundo nível de escolaridade do pai e mãe, classe econômica e ocupação - 2012.	31
Tabela 11 – Percentual de alunos da UnB por semestre e <i>campus</i> segundo tipo de instituição que concluiu o ensino fundamental e médio e ocupação do aluno – 2012	33
Tabela 12 – Total de alunos da UnB por turno e campus e semestre de ingresso - 2012	34
Tabela 13 – Percentual de alunos da UnB por semestre e turno do curso segundo estado civil, cor, faixa etária e tipo de instituição que concluiu o ensino médio – 2012.	35

Tabela 14 – Percentual de alunos da UnB por semestre e turno do curso segundo nível de escolaridade do pai e mãe, classe econômica e local de residência - 2012.	36
Tabela 15 – Percentual de alunos da UnB por semestre e turno do curso segundo tipo de instituição que concluiu o ensino fundamental e médio, forma e sistema de ingresso e ocupação do aluno – 2012.....	37
Tabela 16 – Percentual de alunos da UnB por semestre e sistema de ingresso segundo sexo, estado civil, cor e faixa etária - 2012.....	39
Tabela 17 - Percentual de alunos da UnB por semestre e sistema de ingresso segundo nível de escolaridade do pai e mãe, classe econômica e local de residência- 2012.....	41
Tabela 18 – Percentual de alunos da UnB por semestre e turno do curso segundo tipo de instituição que concluiu o ensino fundamental e médio, forma e sistema de ingresso e ocupação do aluno – 2012.....	42
Tabela 19- Evolução do IDEB das turmas da 3ª série do ensino médio segundo dependência administrativa.....	43
Tabela 20 – Percentual de alunos da UnB por semestre tipo de escola que cursou o ensino médio segundo sexo, estado civil, cor e faixa etária - 2012.....	45
Tabela 21 – Percentual de alunos da UnB por semestre tipo de escola que cursou o ensino médio segundo nível de escolaridade do pai e mãe, local de residência e classe econômica - 2012.....	47
Tabela 22 - Percentual de alunos da UnB por semestre tipo de escola que cursou o ensino médio segundo tipo de instituição que concluiu o ensino fundamental, forma e sistema de ingresso área do curso escolhido e ocupação do aluno – 2012	48
Tabela 23 - Estimativas e Estatísticas de Wald do Modelo Logístico (1/2012)	52
Tabela 24 - Estimativas das Razões de Chances para as Variáveis Significativas do Modelo Logístico (1/2012).....	54
Tabela 25 - Estimativas e Estatísticas de Wald do Modelo Logístico (2/2012)	56
Tabela 26 - Estimativas das Razões de Chances para as Variáveis Significativas do Modelo Logístico (2/2012).....	58

Resumo

Este trabalho é um estudo observacional com o objetivo de traçar o perfil do estudante que ingressou na Universidade de Brasília (UnB) em 2012. Foi delineado o perfil do aluno por semestre de ingresso, campus, turno, sistema de ingresso (cotas raciais) e a natureza institucional da escola no ensino médio. Visto a obrigatoriedade da implantação de cotas para estudantes concluintes em escola pública em todas as universidades federais, com a sanção da Lei nº 12.711/2012, optou-se por ajustar um modelo de regressão logística com o objetivo identificar as principais características dos estudantes provenientes da rede pública que ingressaram na UnB antes da implantação da nova política de cotas. Para a realização da pesquisa foi utilizada a base de dados coletada pelo Observatório da Vida Estudantil obtida através da aplicação de questionário a todos os ingressantes na UnB em 2012.

Palavras-chaves: Regressão Logística, Perfil do aluno, UnB, Escola Pública.

1. INTRODUÇÃO

A Universidade de Brasília foi inaugurada no dia 21 de Abril de 1962 e completou 50 anos em 2012. Ao longo de sua história foram registrados fatos marcantes e inéditos no Brasil, como a implantação do Programa de Avaliação Seriada (PAS) e do sistema de cotas para negros e pardos na forma de ingresso.

No primeiro semestre de 2012 a Universidade era composta por 28.570 alunos de graduação e 6.304 de pós-graduação, 2.630 técnico-administrativos e 2.445 professores. É formada por 26 institutos e faculdades e 21 centros de pesquisa especializados. A UnB está dividida em 4 *campi*, sendo que os criados recentemente estão situados na Ceilândia e no Gama e foram construídos com o auxílio do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI). O programa criado pelo Governo Federal em 2007 visa melhorar a infraestrutura e qualidade nas Universidades Federais através da construção e reforma de instalações, criação de mais vagas, compra de novos equipamentos e contratação de professores e funcionários.

O REUNI colaborou com o aumento do número de cursos oferecidos pela instituição, atualmente são oferecidos 109 cursos de graduação, sendo 31 noturnos e 10 à distância. Em 2008, antes da implantação do REUNI, a UnB oferecia 64 cursos, sendo 15 noturnos e 8 à distância. Conseqüentemente o número de alunos na Universidade aumentou, mas esse aumento também foi motivado pelo aumento no número de vagas em cursos já existentes.

As principais fontes de informações disponíveis a respeito dos alunos da UnB eram as constantes no registro (processo em que o aprovado no processo seletivo apresenta documentação para a confirmação de ocupação da vaga) e do questionário sócio cultural do CESPE, em que os candidatos respondiam no ato da inscrição do vestibular. Alguns estudos pontuais sobre o perfil do estudante da UnB já foram realizados por alguns departamentos da universidade. Este trabalho visa atualizar as informações sobre o tema, assim como utilizar técnicas estatísticas para gerar resultados.

No desenvolvimento desse trabalho foram considerados os dados coletados pelo Observatório da Vida Estudantil da Universidade de Brasília, grupo de pesquisa formado por membros do corpo docente da Universidade de Brasília

que visa entre outros projetos, descrever as características sociodemográficas, econômicas e culturais do corpo discente da universidade, e acompanhar sua trajetória estudantil.

2. OBJETIVOS

2.1. Objetivo Geral

Este trabalho tem por objetivo principal avaliar o perfil dos estudantes de graduação que ingressaram na Universidade de Brasília em 2012.

2.2. Objetivos Específicos

Fazer um estudo comparativo dos perfis dos estudantes, usando técnicas estatísticas, segundo as seguintes variáveis:

- *Campus*;
- Turno do curso (Diurno/Noturno);
- Sistema de ingresso (Cotas para negros e pardos ou não cotistas);
- Tipo de Escola que estudou no Ensino Médio (Pública ou Particular);
- Identificar que características estão associadas aos estudantes que ingressaram na Universidade de Brasília vindos de instituição de ensino médio pública.

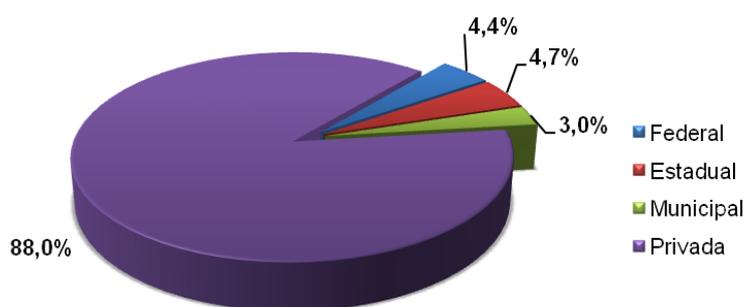
3. REFERENCIAL TEÓRICO

Este capítulo está dividido em três tópicos essenciais. Primeiramente será apresentada serão expostas informações relevantes sobre a Universidade de Brasília (UnB). Nos tópicos seguintes do referencial, serão expostas as principais técnicas de Análise de Dados Categorizados e ajuste de Modelo de Regressão Logística.

3.1.A Universidade de Brasília

No Brasil existem 2.365 Instituições de Ensino Superior (IES), segundo o levantamento do Censo da Educação Superior realizado em 2011 divulgado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Deste total, 4,4% são Federais, 7,7% Estaduais e Municipais e 88% são da rede privada, como ilustra o gráfico abaixo.

Gráfico 1 - Distribuição das Instituições de Ensino Superior por Categoria Administrativa, Brasil - 2011



Fonte: MEC/INEP. Censo do Ensino Superior - 2011.

O Distrito Federal conta com apenas 3 IES pública, enquanto que pela rede privada são 56 IES. Das três Instituições Públicas no DF, uma é Estadual (Escola Superior de Ciências da Saúde - ESCS) e duas são Federais, Universidade de Brasília (UnB) e Instituto Federal de Educação, Ciência e Tecnologia de Brasília

(IFB), sendo que o IFB foi criado a pouco tempo, dezembro de 2008, ou seja, a UnB foi por muitos anos a única IES federal do Distrito Federal e é a única Universidade Federal do DF num total de 59 IES. Dado tal característica da UnB, este trabalho procura conhecer melhor seus alunos analisando além do perfil socioeconômico, a trajetória pré-universitária e perspectivas futuras.

O primeiro *campus* da UnB, localizado na Asa Norte, foi inaugurado em abril de 1962, quando Brasília tinha apenas dois anos, e contava com nove prédios e 13 mil metros quadrados de área construída. Com o passar dos anos a Universidade acompanhou também o crescimento da capital federal e hoje o *campus* da Asa Norte já conta com 552.171,40 metros quadrados de área construída.

Nesses 51 anos, a UnB foi palco de fatos históricos no cenário político e educacional brasileiro. Sofreu invasões no período da ditadura Militar e inovou com a criação do Programa de Avaliação Seriada (PAS), Sistema de Cotas para Negros e vagas para indígenas.

3.1.1. Os Campi

A UnB surgiu inicialmente com o *campus* Darcy Ribeiro – Plano Piloto. Em 2006, foi inaugurado o segundo *campus* da UnB, a Faculdade UnB Planaltina, fruto do planejamento estratégico de expansão da UnB, visando a ampliação da oferta de vagas para a população do DF e entorno com a criação de 4 novos cursos voltados para o desenvolvimento regional.

No ano seguinte, através do Decreto nº 6.096, de 24 de abril de 2007, do Ministério da Educação – MEC, foi instituído o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (Reuni). Com a colaboração do Reuni, foram criados, em 2008, dois *campi* da Universidade de Brasília nas Regiões Administrativas Ceilândia e Gama.

A Faculdade de Ceilândia – UnB, conta com 5 cursos da área da Saúde e a Faculdade do Gama – UnB oferece aos alunos 5 especialidades de Engenharia.

O Reuni tem como objetivos criar condições para a ampliação do acesso e permanência na educação superior, em nível de graduação, aumentar a qualidade

dos cursos e melhorar o aproveitamento da estrutura física e de recursos humanos existentes nas Universidades federais, respeitadas as características particulares de cada instituição e estimulada a diversidade do sistema de ensino superior.

3.1.2. Formas de admissão

As formas de admissão mais tradicionais da UnB são o vestibular e o PAS. As provas são organizadas pelo Centro de Seleção e Promoção de Eventos - CESPE. O vestibular é realizado duas vezes ao ano, onde as vagas ofertadas no primeiro semestre do ano são divididas com o PAS. As provas do PAS são aplicadas para alunos do Ensino Médio ao final de cada série. No final do Ensino Médio, a média das três notas resulta na classificação do candidato.

Do total das vagas oferecidas pela UnB para o vestibular, 20% são reservadas para sistema de cotas para negros, para alunos de cor preta e parda. Os candidatos que se inscrevem por cotas devem se autodeclarar negros e passar por uma entrevista para atestar que tem traços fenotípicos e se autodeclara negro.

As vagas remanescentes do vestibular e do PAS são ofertadas para candidatos avaliados pelo Exame Nacional de Ensino Médio – ENEM.

As vagas resultantes de mudança de curso e evasão de alunos são oferecidas por Portador de Diploma de Ensino Superior (DSC) e Transferência Facultativa. O DSC é oferecido para estudantes que já concluíram o ensino superior e possuem entre 20% e 75% da carga horária do curso que se pretende entrar. O ingresso por transferência facultativa é voltada para alunos regulares de outras Instituições de Ensino Superior nacionais ou estrangeiras que desejam ingressar na UnB. Para ingressar por DCS ou transferência facultativa o candidato deve possuir entre 20% e 75% da carga horária do curso que se pretende entrar. A carga horária mínima é exigida, pois as turmas iniciais dos cursos costumam não ter vaga. Os solicitantes precisam fazer uma prova dissertativa, aplicada pelo CESPE. Existem ainda outras formas de ingressar na UnB, como transferência obrigatória, acordos culturais, etc.

Segundo, Francis e Tannuri-Pianto (2010), o perfil socioeconômico dos indivíduos que se candidatam a uma vaga na UnB se difere do perfil dos indivíduos que de fato ingressam na instituição. De forma que os alunos vêm de famílias de renda familiar e nível educacional dos pais superior aos dos candidatos. Há desigualdade entre os alunos quanto à raça, os alunos pretos são os que apresentam menor nível de renda e educação dos pais. Porém, neste trabalho, será pesquisado apenas o perfil dos alunos da Universidade.

3.2. Análise de Dados Categorizados

A maioria das respostas do questionário aplicado aos ingressantes da UnB em 2012 se classifica como variáveis categóricas ou variáveis quantitativas categorizadas. Portanto, faz-se essencial a realização de um levantamento sobre técnicas de Dados Categorizados.

Uma variável categorizada é aquela na qual a escala de medida consiste em um conjunto de categorias, (AGRESTI, 2002). Temos os seguintes exemplos de variáveis categóricas: forma de ingresso na UnB, as respostas possíveis são as seguintes categorias: Vestibular, PAS, DSC ou Transferência. A variável renda é um exemplo de variável quantitativa e quando dividida em faixas salariais ou número de salários mínimos se torna uma variável categorizada.

Outra definição importante no estudo de Dados Categorizados é a distinção entre as variáveis, elas podem ser:

- Variável resposta (ou dependente);
- Variável explicativa (ou independente).

Tomando como exemplo as variáveis curso e sexo, a escolha do curso não influencia no sexo do aluno, mas o sexo do aluno pode explicar a escolha do curso, sendo possível identificar cursos que ingressam mais homens ou mais mulheres. Portanto, neste caso, o sexo seria a variável explicativa e curso a variável resposta. Em alguns casos, a escolha da variável resposta pode não ser trivial.

3.2.1. Comparação entre categorias

Quando se tem uma tabela de contingência 2x2, ou seja, variável resposta e explicativa tem duas categorias, pode-se utilizar as seguintes medidas para comparar as proporções de sucesso, p_1 e p_2 , de grupos independentes: diferença de proporções, risco relativo e razão de chances.

➤ *Diferença de Proporções*

A diferença de proporções é dada por $p_1 - p_2$. Se a diferença está próxima ou é igual à zero a variável resposta não depende da variável explicativa. Tem-se que a estimativa do desvio padrão de $p_1 - p_2$ é dado por:

$$\hat{\sigma}(p_1 - p_2) = \sqrt{\frac{p_1(1 - p_1)}{n_{1+}} + \frac{p_2(1 - p_2)}{n_{2+}}} \quad (1)$$

Para amostras suficientemente grandes o intervalo de $p_1 - p_2$ é dado por:

$$IC(p_1 - p_2) = (p_1 - p_2) \pm Z_{\frac{\alpha}{2}} \hat{\sigma}(p_1 - p_2) \quad (2)$$

➤ *Risco Relativo*

O Risco Relativo (RR) é a razão das proporções de sucesso de dois grupos. Pode ser calculado para estudo de coorte ou transversal. Calculado o risco relativo, interpreta-se da seguinte forma:

- Quando $0 < RR < 1$, o grupo do denominador tem $1/RR$ vezes mais risco de apresentar a característica em estudo.
- Quando o $RR = 1$ a característica em estudo não depende dos grupos da variável explicativa.
- Quando $RR > 1$ então o grupo do numerador tem RR vezes mais risco de apresentar a característica em estudo.

Para a distribuição de Risco Relativo ficar simétrica e se aproximar de uma distribuição Normal, trabalha-se com o logaritmo do Risco Relativo, cujo desvio padrão e intervalo de confiança é dado por:

$$\hat{\sigma}(\log(RR)) = \sqrt{\frac{(1 - p_1)}{p_1 n_{1+}} + \frac{(1 - p_2)}{p_2 n_{2+}}} \quad (3)$$

$$IC(\log(RR)) = \log(RR) \pm Z_{\frac{\alpha}{2}} \hat{\sigma}(\log(RR)) \quad (4)$$

O intervalo do Risco Relativo é encontrado aplicando a função exponencial nos valores do intervalo de $\log(RR)$.

➤ *Razão de Chances*

Antes de falar sobre a razão de chance é importante ter em mente a diferença entre probabilidade e chance. A probabilidade é definida como número de sucesso dividido pelo número total de observações e chance é igual ao número de sucesso dividido pelo número de fracasso. Portanto a chance (*odds*) pode ser escrita da seguinte maneira:

$$Odds = \frac{p}{1 - p} \quad (5)$$

Onde:

p : proporção de sucesso.

A Razão de Chance (*Odds Ratio*), como o nome sugere, é a razão de duas chances. Esta medida é indicada para estudos observacionais, mas pode ser aplicada para demais delineamentos. Em uma tabela de contingência 2x2 a Razão de Chance pode ser calculada pelo produto cruzado dos termos da tabela:

$$OR = \frac{\frac{p_1}{1 - p_1}}{\frac{p_2}{1 - p_2}} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad (6)$$

Os valores da Razão de Chance variam entre zero e infinito e são interpretados de acordo com os critérios abaixo:

- Quando $0 < OR < 1$, a chance de sucesso da categoria da segunda linha é maior.
- Quando $OR = 1$, não há associação entre as variáveis.
- Quando $OR > 1$, a chance de sucesso da categoria da primeira linha é maior.

Para grandes amostras é possível provar que o logaritmo da Razão de Chance tem distribuição de probabilidade aproximadamente Normal com média igual a zero e desvio padrão:

$$\hat{\sigma}(\log(OR)) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \quad (7)$$

Com isso, podemos calcular o intervalo de confiança do $\log(OR)$:

$$IC(\log(OR)) = \log(OR) \pm Z_{\frac{\alpha}{2}} \hat{\sigma}(\log(OR)) \quad (8)$$

O intervalo da Razão de Chances é encontrado aplicando a função exponencial nos valores do intervalo de $\log(OR)$.

Utilizando as variáveis do exemplo anterior, foi criada a seguinte tabela de contingência:

Sexo	Área do Curso		Total
	Humanas	Exatas	
Mulheres	435 (0,75)	147 (0,25)	582 (1,0)
Homens	175 (0,41)	257 (0,59)	432 (1,0)
Total	610	404	1014

Supondo que as informações foram obtidas de um estudo transversal, podemos calcular o risco relativo e a razão de chance. Considerara-se sucesso o indivíduo ser da área de Humanas, logo, $p_1 = 0,75$ e $p_2 = 0,41$. Com essas

probabilidades podemos calcular o Risco Relativo, onde $RR = 0,75 / 0,41 = 1,83$. Portanto, o risco de uma mulher escolher um curso de humanas é 83% maior que um homem.

Calculando a Razão de Chance do exemplo encontra-se o valor 4,35. Logo, a chance de uma mulher ser de um curso de humanas é 4,35 vezes maior que um homem.

3.3. Regressão Logística

A regressão logística pode ser usada em estudo que a variável resposta é qualitativa e dicotômica ou binária, assumindo os valores 0 (zero) e 1 (um). Uma generalização do modelo de regressão logística pode ser ajustado para variáveis qualitativas com mais de 2 níveis. Existem diversos softwares estatísticos que oferecem técnica de Regressão Logística, o SAS oferece os procedimentos *Catmod*, *Reg logistic* e *Probit* para o ajuste.

3.3.1. Regressão Logística Simples

A regressão logística é um caso particular dos Modelos Lineares Generalizados (*GLM*, do inglês *Generalized Linear Models*), cujo componente aleatório Y_i tem distribuição Binomial e a função link é a transformação *logit*. O modelo de regressão logística tem forma linear para o *logit* da probabilidade de sucesso dada por:

$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \beta_1 x \quad (9)$$

Suponha uma variável explicativa categórica X . Para uma variável resposta Y , tem-se que $\pi(x)$ representa a probabilidade de sucesso na i -ésima categoria de X . Aplicando a função exponencial na função *logit* da regressão logística chega-se a esta probabilidade:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \quad (10)$$

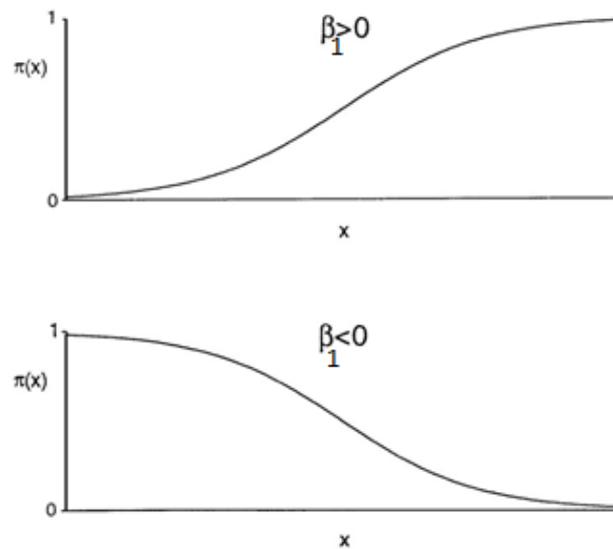
Uma das vantagens da função *logit* é que ela pode assumir qualquer valor real, enquanto que o valor de $\pi(\mathbf{x})$ é restrito ao intervalo (0,1).

3.3.2. Interpretação do Modelo

Os parâmetros de um modelo determinam o comportamento da curva de regressão e sentido de associação de cada variável explicativa com a variável resposta.

A curva do modelo de regressão logística tem formato *sigmoidal* ou forma de S. O parâmetro β_1 determina se a curva é crescente ou decrescente para os níveis de x e $|\beta_1|$ define a magnitude do crescimento/decrescimento da curva. Se $\beta_1 = 0$ a curva S se torna uma reta horizontal onde os valores de $\pi(x)$ se tornam constantes para os níveis de x . A figura a seguir ilustra o comportamento da curva de regressão para valores de $\beta_1 > 0$ e $\beta_1 < 0$, é possível observar que a curva tende assintoticamente para 0 e 1.

Figura 1 – Representação da curva $\pi(x)$ para $\beta_1 > 0$ e $\beta_1 < 0$



A principal interpretação da Regressão Logística é realizada através da utilização da chance (odds) e da função *odds ratio* (OR) ou razão de chance, em português. Para chegar à função razão de chance, primeiro define-se a chance de sucesso do modelo sendo:

$$odds1 = \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = \exp(\beta_0 + \beta_1 x) \quad (11)$$

E seja odds2 igual a

$$odds2 = \frac{\pi(\mathbf{x} + \mathbf{1})}{1 - \pi(\mathbf{x} + \mathbf{1})} = \exp(\beta_0 + \beta_1(x + 1)) \quad (12)$$

Tem-se que a razão de chance é dada pela razão das expressões (11) e (12), chega-se a seguinte relação:

$$OR = \frac{odds2}{odds1} = \exp(\beta_1) \quad (13)$$

Se $OR = 1$, ou seja, $\beta_1 = 0$, então a razão de chances não se altera para mudanças nos níveis de X e a variável resposta é independente da variável explicativa, logo, as variáveis X e Y não estão associadas.

3.3.3. Inferência para Modelos de Regressão Logística

Assim como nas regressões lineares, na Regressão Logística busca-se métodos para a estimação dos coeficientes, construção de testes hipóteses e intervalo de confiança. A melhor técnica para se chegar aos coeficientes do modelo é por Máxima Verossimilhança por suas propriedades assintóticas.

Vimos que Y_i tem distribuição de probabilidade Binomial com parâmetros m_i e proporção de sucesso igual a $\pi(x_i)$. Como os Y_i 's são independentes, a função de verossimilhança de (Y_1, \dots, Y_n) é dada por:

$$l(\beta_0, \beta_1) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{m_i - y_i} \quad (14)$$

Onde $\pi(x_i)$ é dada pela expressão em (10).

Aplicando logaritmo na função $l(\beta_0, \beta_1)$ e derivando parcialmente, obtemos as equações normais:

$$\frac{\partial}{\partial \beta_0} L(\beta_0, \beta_1) = \sum_{i=1}^n y_i - \sum_{i=1}^n m_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (15)$$

$$\frac{\partial}{\partial \beta_1} L(\beta_0, \beta_1) = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n m_i x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (16)$$

A solução deste sistema pode ser obtida por métodos computacionais iterativos, como Newton-Rapson, uma vez que as equações são não lineares. Com isso, o estimador de máxima verossimilhança de $\pi(\mathbf{x})$ será:

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x)} \quad (16)$$

3.3.4. Intervalo de confiança

Para grandes amostras, o intervalo de confiança para o parâmetro β_1 no modelo de regressão logística será:

$$\beta_1 \in \widetilde{\beta}_1 \pm z_{\alpha/2} ASE \quad (17)$$

Onde ASE, do inglês *Asymptotic Standard Error*, é o erro padrão assintótico de β_1 .

3.3.5. Seleção de variáveis do modelo

O Teste F, na análise de Regressão Linear, testa se um modelo de regressão específico está bem ajustado aos dados. Na Regressão Logística o teste utilizado com esta finalidade é a *Deviance*.

A *Deviance* é dada por:

$$Deviance = -2 [\ln(L_r) - \ln(L_s)] \quad (18)$$

Onde:

L_r = valor máximo da função de máxima verossimilhança do modelo interessado ou modelo reduzido.

L_s = valor máximo da função de máxima verossimilhança do modelo saturado (mais complexo).

A *Deviance* é um teste de razão de verossimilhança e tem distribuição de probabilidade Qui-quadrado com $c - p$ graus de liberdade, onde c é o número de parâmetros do modelo reduzido e p é o número de parâmetros do modelo saturado.

Nos métodos de seleção automática de variáveis de modelos Logísticos são adotados critérios para adição ou remoção de variáveis, entre estes critérios estão o p-valor de significância de cada variável do modelo e AIC. As técnicas mais conhecidas são *Backward*, *Forward* e *Stepwise*.

O método *Backward* inicia com o modelo mais complexo (saturado) e sucessivamente vai retirando as variáveis do modelo. Em cada etapa elimina-se a variável que possui o maior p-valor obtido quando testado as significâncias dos variáveis. A seleção termina quando não existir nenhuma variável com p-valor maior que um nível de significância α fixado inicialmente.

O método *Forward* segue o caminho oposto do *Backward* e inicia com o modelo mais simples (contendo somente o intercepto) e sucessivamente vai incluindo as variáveis explicativas no modelo. Em cada etapa inclui-se a variável que possui o menor p-valor obtido quando se testa as significâncias das variáveis. A seleção termina quando não existir nenhuma variável com p-valor menor que um nível de significância α fixado inicialmente.

O método *Stepwise* inicia também com o modelo mais simples (contendo somente o intercepto), mas sucessivamente por etapas ele inclui e exclui variáveis do modelo, sendo o método mais iterativo. Na primeira etapa calcula-se o p-valor de cada variável fora do modelo e inclui-se no modelo o que possuir o menor p-valor obtido. A partir da segunda etapa ele realiza o mesmo procedimento para analisar a inclusão de uma nova variável, mas também realiza o teste para a exclusão de variáveis da mesma forma que o método *Backward*. A seleção de variáveis termina quando não existir mais nenhuma variável para se adicionar ou remover do modelo.

É importante ressaltar que nos métodos de seleção automática de variáveis não é testado todos os modelos possíveis.

3.3.6. Construção do Modelo de Regressão

Para se ajustar um modelo de regressão existem etapas essenciais que devem ser cumpridas. O livro *Applied Linear Statistical Models*, Kutner et. al. propõe uma estratégia envolvendo quatro passos: coleta e preparação dos dados, redução do número de variáveis explicativas, refinamento e seleção do modelo e validação do modelo.

Neste trabalho a coleta dos dados foi realizada através da aplicação de questionários. Uma vez que os dados foram coletados iniciou-se a preparação dos dados, que consiste principalmente na análise de consistência com o propósito de verificar se as respostas estão coerentes.

Em seguida, realizou-se a análise descritiva dos dados. Nesta fase é possível detectar as variáveis explicativas que estão mais associadas com a variável resposta do estudo e, portanto terão presença importante no modelo. Por meio de critérios computacionais podemos reduzir o número de variáveis explicativas, detectando quais delas são correlacionadas entre si.

O próximo passo é a construção do modelo de regressão. Existe uma variedade de possíveis bons modelos para os dados. Os gráficos de resíduos ajudam na decisão sobre qual modelo é mais preferível que outro. O diagnóstico da regressão é eficaz na identificação de pontos influentes, *outliers* e colinearidade ou multicolinearidade.

E para finalizar a construção do modelo é realizado o processo de validação do modelo. Usualmente, a validação envolve a avaliação dos coeficientes da regressão, verificando se o comportamento deles se porta no mesmo sentido em que está associado com a variável resposta e se a função é plausível.

4. METODOLOGIA

Os dados utilizados nesse trabalho são provenientes de duas bases de dados geradas a partir do questionário “Perfil dos Estudantes da Universidade de Brasília – Etapa Registro”, elaborado e aplicado pelo grupo de pesquisadores do Observatório da Vida Estudantil da Universidade de Brasília, a todos os estudantes que ingressaram na Universidade de Brasília no primeiro e segundo semestres de 2012, respectivamente, independentemente da forma de ingresso.

O questionário foi preenchido via online e vinculado à etapa de registro. Ele é composto por 40 questões no primeiro semestre e 51 questões no segundo semestre, distribuídas em cinco blocos temáticos: 1) Identificação, 2) Perfil Socioeconômico e Demográfico, 3) Trajetória Pré-universitária, 4) Inserção Universitária, e 5) Trabalho e Perspectivas Futuras.

O questionário aplicado sofreu alterações na segunda aplicação. As principais foram: alteração da nomenclatura da descrição da renda individual, adição da opção de residência com companheiro (a) - no primeiro questionário não existia esta opção, mas pela especificação no item ‘outros’, foi possível criar a opção - e foram incluídas questões sobre atendimento diferenciado na realização da prova.

Finalizada consolidação da base de dados, iniciou-se o trabalho a análise descritiva dos dados com a elaboração de tabelas e gráficos, e a criação de uma variável que permitisse a classificação dos estudantes em classes econômicas.

Fundamentado no Critério de Classificação Econômica Brasil (CCEB) elaborado pela Associação Brasileira de Empresa de Pesquisa (ABEP) foi criada a classificação econômica dos alunos da UnB com algumas modificações semelhantes as do relatório do Perfil Socioeconômico e Cultural dos Estudantes de Graduação das Universidades Federais Brasileiras, produzido pelo Fórum Nacional de Pró-Reitores de Assuntos Comunitários e Estudantis (FONAPRACE)¹ em 2011. Lembrando que esta classificação “ênfatiza sua função de estimar o poder de compra das pessoas e famílias urbanas, abandonando a pretensão de classificar a

¹ O relatório do Perfil Socioeconômico e Cultural dos Estudantes de Graduação das Universidades Federais Brasileiras da Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (ANDIFES) foi encomendado ao FONAPRACE com o objetivo de mapear a vida social, econômica e cultural dos estudantes de graduação presencial das Universidades Federais brasileiras.

população em termos de ‘classes sociais’. A divisão de mercado definida abaixo é de classes sociais” (ABEP, 2013). As tabelas a seguir apresentam a pontuação atribuída pela ABEP a bens, facilidades e serviços (Tabela 1) e ao grau de instrução do chefe de família (Tabela 2).

Tabela 1 - Pontuação atribuída pela ABEP a bens, facilidades e serviços para classificação econômica.

Posse de itens	Quantidade de itens				
	0	1	2	3	4 ou +
Televisão em cores	0	1	2	3	4
Rádio	0	1	2	3	4
Banheiro	0	4	5	6	7
Automóvel	0	4	7	9	9
Empregada mensalista	0	3	4	4	4
Máquina de lavar	0	2	2	2	2
Videocassete e/ou DVD	0	2	2	2	2
Geladeira	0	4	4	4	4
Freezer (aparelho independente ou parte da geladeira duplex)	0	2	2	2	2

Tabela 2 - Pontuação atribuída pela ABEP ao grau de instrução do chefe da família para classificação econômica.

Nomenclatura Antiga	Nomenclatura Atual	Pontos
Analfabeto/ Primário incompleto	Analfabeto/ Fundamental 1 Incompleto	0
Primário completo/ Ginásio incompleto	Fundamental 1 Completo / Fundamental 2 Inc.	1
Ginasial completo/ Colegial incompleto	Fundamental 2 Completo/ Médio Incompleto	2
Colegial completo/ Superior incompleto	Médio Completo/ Superior Incompleto	4
Superior completo	Superior Completo	8

No questionário aplicado aos estudantes, não constava as variáveis banheiro e empregada mensalista. Estes foram substituídos pelos itens: acesso a internet, computador e TV por assinatura. O item grau de instrução do chefe de família foi escolhido o grau de instrução mais alto entre o do pai e da mãe. No sistema de pontuação foi adicionado Pós-graduação.

As tabelas 3 e 4 apresentam as adaptações utilizadas na pontuação atribuídas a bens, facilidades e serviços e grau de instrução do chefe da família.

Tabela 3 - Pontuação atribuída a bens, facilidades e serviços para classificação econômica.

Posse de itens	Quantidade de itens				
	0	1	2	3	4 ou +
Televisão em cores	0	1	2	3	4
Rádio	0	1	2	3	4
Automóvel	0	4	7	9	9
Máquina de lavar	0	2	2	2	2
DVD	0	2	2	2	2
Geladeira	0	4	4	4	4
Freezer	0	2	2	2	2
Computador	0	2	2	3	3
Internet	0	2	2	2	2
Celular	0	2	2	2	2
Televisão por assinatura	0	2	2	2	2

Tabela 4- Pontuação atribuída ao grau de instrução do chefe da família para Classificação econômica.

Grau de instrução do chefe de família	Pontos
Não sabe ler nem escrever	0
Fundamental incompleto	1
Fundamental completo / Médio incompleto	2
Médio completo / Superior incompleto	4
Ensino superior completo	8
Pós-graduação	10

A distribuição dos pontos atribuída ao indicador de classe adaptado foi ajustada de tal forma que a soma máxima de pontos seja a mesma da ABEP, 46 pontos, e desta forma adotar as divisões de faixas econômicas conforme a proposta pela ABEP. As faixas de pontuação que permitem identificar a classe socioeconômica do estudante são apresentadas na tabela a seguir:

Tabela 5 - Faixas de pontuação estabelecidas pela ABEP para classificação econômica

Classe	Pontos
A	35 - 46
B	23 - 34
C	14 - 22
D	8 - 13
E	0 - 7

Nas análises das tabelas, foram utilizadas medidas de associação, testes de hipóteses entre outras técnicas empregadas em estudos de Análise de Dados

Categorizados. A Regressão Logística foi utilizada na identificação das características associadas à origem institucional do ensino médio do estudante que ingressou na UnB.

Foram utilizados para auxiliar a criação de tabelas, gráficos, testes estatísticos e modelagem os softwares SAS e R Project.

6. MODELAGEM

Visando complementar a análise do perfil dos estudantes que ingressaram na universidade nos dois semestres de 2012 deseja-se identificar quais características estão associadas aos estudantes que ingressaram na Universidade de Brasília vindos de instituição de ensino médio pública. Para alcançar esse objetivo ajustou-se um modelo regressão logística binária, onde a variável resposta é a natureza da instituição em que o aluno estudou no ensino médio (pública ou particular).

A partir das análises descritivas da seção 5.5 foram eleitas as prováveis variáveis explicativas para o modelo de regressão. Os estudantes de escola particular com bolsa foram agregados ao nível escola particular, uma vez que o perfil destes alunos serem mais próximos. A variável resposta foi classificada como:

$Y_i = 0$, aluno i cursou ensino médio em escola particular

$Y_i = 1$, aluno i cursou ensino médio em escola pública

Devido a mudanças no perfil dos alunos que ingressam na instituição conforme o semestre de ingresso, como se observa na seção 5.1, na análise descritiva por semestre, adotou-se ajustar uma regressão logística para cada semestre.

Inicialmente, foram eleitas 21 possíveis variáveis explicativas para o modelo, entre elas estão variáveis sociodemográficas (nível de escolaridade da mãe e pai, classe econômica, estado civil, com quem reside, cor, ocupação etc.), trajetória pré-universitária (natureza da instituição que cursou o ensino fundamental, realização de atividade extra-curricular), inserção universitária (forma e sistema de ingresso, campus, turno e área do curso, etc). Utilizando o método de seleção automática *Forward/Stepwise* foram escolhidas 12 variáveis, porém no ajuste do modelo com essas variáveis, o p-valor do ajuste e do efeito principal era significativo, mas intervalo de confiança para a estimativa da razão de chance de todos os níveis indicavam não associação (continha o valor 1 no intervalo) devido ao alto valor do erro padrão. Levantou-se a hipótese da presença de multicolinearidade no modelo.

Foi realiza um estudo quanto ao grau de associação entre as variáveis explicativas utilizando o teste qui-quadrado de independência e identificou-se a variáveis explicativas que estavam fortemente associados.

Com base na natureza das variáveis e considerando o contexto do estudo foram selecionas a variáveis que seriam incluídas no processo de modelagem, entre as que estavam altamente relacionadas.

Sendo assim, deve-se, por exemplo, escolher entre classe econômica e nível escolaridade da mãe. Dando continuidade ao processo de modelagem obteve-se finalmente, para o primeiro modelo (1/2012) foram selecionadas cinco variáveis explicativas: nível de escolaridade da mãe, ocupação do aluno, *campus*, cor e forma de ingresso. As categorias de referência para o calculo da razão de chances das variáveis foram ensino fundamental, nunca trabalhou, *campus* Darcy Ribeiro, cor branca e vestibular, respectivamente. No segundo modelo (2/2012), contém as mesmas variáveis do semestre anterior, acrescida da variável troca de curso se tivesse oportunidade.

6.1. Primeiro semestre (1/2012)

A probabilidade $\pi(x_i)$ de o aluno ser classificado como estudante de escola pública em 1/2012 na UnB é dada por:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_i^M x_{i1} + \beta_j^O x_{j2} + \beta_k^C x_{k3} + \beta_l^R x_{l4} + \beta_m^I x_{m5})}{1 + \exp(\beta_0 + \beta_i^M x_{i1} + \beta_j^O x_{j2} + \beta_k^C x_{k3} + \beta_l^R x_{l4} + \beta_m^I x_{m5})}$$

Onde $\beta_i^M, \beta_j^O, \beta_k^C, \beta_l^R$ e β_m^I são as classes observadas para os níveis das variáveis explicativas: nível de escolaridade da mãe, ocupação do aluno, *campus*, cor e forma de ingresso, respectivamente.

O teste da razão de verossimilhança com a hipótese nula de que todos os parâmetros do modelo são iguais à zero foi rejeitado com p-valor menor que 0,0001,

ou seja, pelo menos um dos parâmetros é diferente de zero. Porém, alguns níveis dessas variáveis não foram significativos e, por isso, seus efeitos não foram estimados. A Tabela 23 mostra os valores das estatísticas dos testes de Wald e seus respectivos p-valor:

Tabela 23 - Estimativas e Estatísticas de Wald do Modelo Logístico (1/2012)

Parâmetro	GL	Estimativa	Estatística de Wald	P-valor	
Intercepto	1	1,314	16,76	<,0001	
Esc. Mãe	Não lê nem escreve	1	0,581	0,300	0,5837
	Ens. Médio	1	-1,329	80,03	<,0001
	Pós-graduação	1	-2,947	268,74	<,0001
	Superior Comp.	1	-2,865	286,03	<,0001
Ocupação	Não trabalha no momento	1	0,631	35,86	<,0001
	Trabalha	1	0,699	24,75	<,0001
Campus	Planaltina	1	0,930	10,71	0,0011
	Ceilândia	1	0,414	5,74	0,0166
	Gama	1	-0,176	1,00	0,3181
Ingresso	PAS	1	0,000	0,00	0,9998
	ENEM	1	0,918	20,06	<,0001
	Vagas remanesc.	1	0,901	16,16	<,0001
	Outros	1	-0,184	0,49	0,4861
Cor	Parda	1	0,657	44,37	<,0001
	Preta	1	0,631	15,42	<,0001
	Outros	1	0,104	0,20	0,6576

O teste de razão de verossimilhança encontrou-se uma estatística igual a 1017,8 com 16 graus de liberdade e p-valor < 0,0001, então, a um nível de significância de 5% existem evidências para rejeitar a hipótese nula do teste e, portanto pelo menos um dos betas estimados é diferente de zero.

No diagnóstico do modelo foram utilizados os gráficos com resíduos de Pearson e a Deviance para identificar observações discrepantes ou que não foram bem explicadas pelo modelo. Nos dois gráficos, as observações estão dentro do intervalo desejado e nenhuma observação se sobressai.

Gráfico 11 – Resíduos de Pearson para o modelo ajustado em 1/2012

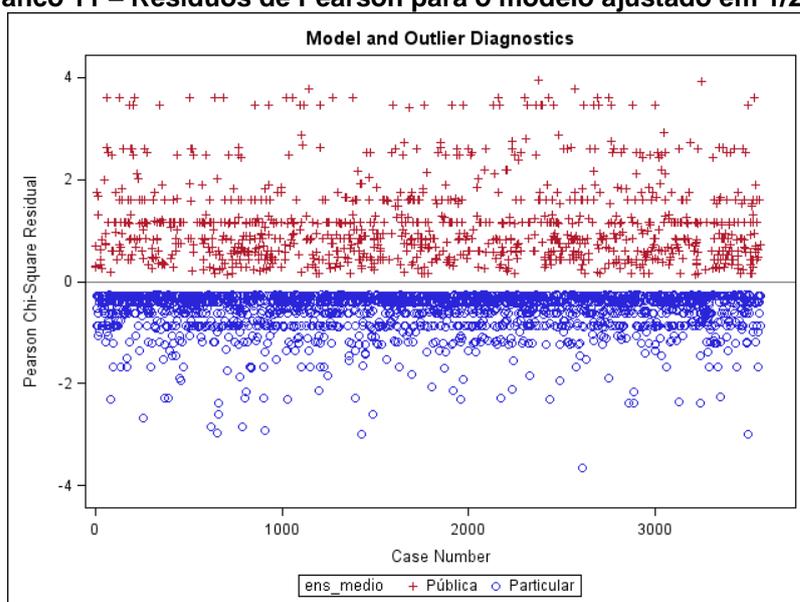
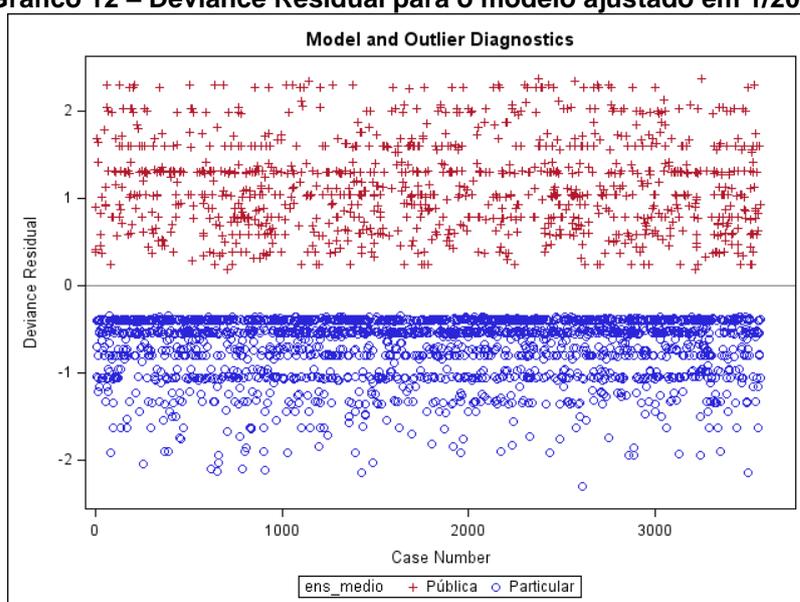
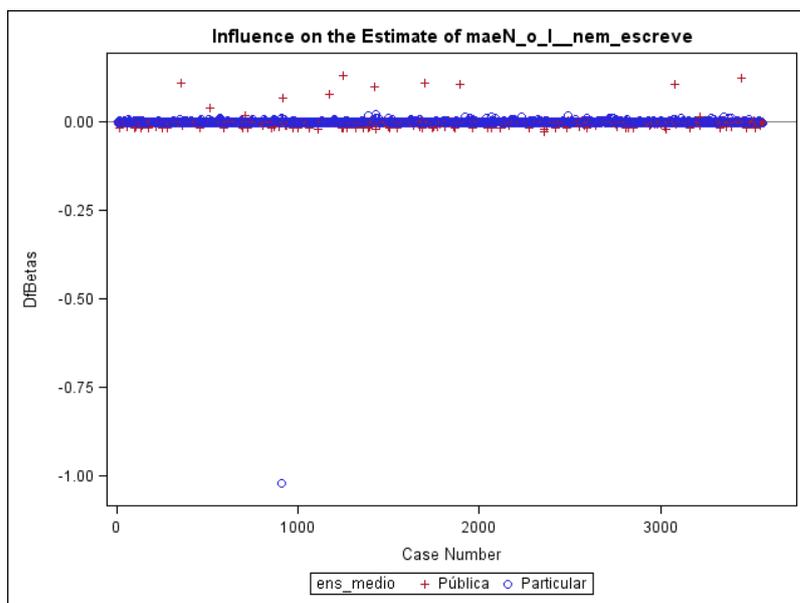


Gráfico 12 – Deviance Residual para o modelo ajustado em 1/2012



Na análise dos DfBetas, apenas uma observação, 911, representa grande influência na estimação do parâmetro escolaridade da mãe ao nível, não lê nem escreve, para um estudante de escola particular. Nos outros gráficos este indivíduo está dentro do intervalo esperado, portanto, no geral, o modelo é adequado aos dados.

Gráfico 13 – DfBetas: Influência na estimativa do parâmetro nível de escolaridade da mãe: analfabeta



A tabela a seguir traz as estimativas das razões de chances e seu respectivo intervalo de confiança. Os intervalos que contem o valor 1 então a razão de chance não se altera para mudanças nos níveis do efeito, mantendo as outras variáveis constantes, ao nível de significância de 5%.

Tabela 24 - Estimativas das Razões de Chances para as Variáveis Significativas do Modelo Logístico (1/2012).

Variáveis	Efeitos	Razão de chance	Intervalo de confiança (95%)	
Esc. Mãe	Não lê nem escreve vs Superior Comp.	31,387	3,945	249,697
	Ens. Fund. vs Superior Comp.	17,551	12,592	24,463
	Ens. Médio vs Superior Comp.	4,647	3,675	5,876
Ocupação	Trabalha vs Nunca trabalhou	2,011	1,527	2,648
	Não trabalha no momento vs Nunca trabalhou	1,879	1,529	2,311
Campus	Planaltina vs Darcy Ribeiro	2,534	1,452	4,421
	Ceilândia vs Darcy Ribeiro	1,513	1,078	2,124
	Gama vs Darcy Ribeiro	0,839	0,594	1,184
Ingresso	ENEM vs Vestibular	2,504	1,676	3,742
	Vagas remanesc. vs Vestibular	2,461	1,587	3,818
	PAS vs Vestibular	1,000	0,820	1,220
	Outros vs Vestibular	0,832	0,496	1,395
Cor	Parda vs Branca	1,929	1,590	2,340
	Preta vs Branca	1,880	1,372	2,576
	Outros vs Branca	1,110	0,701	1,758

A chance de um ingressante da UnB ter concluído o ensino médio em escola pública cuja mãe não lê nem escreve é 31,4 vezes maior que a mãe ter ensino superior completo, quando a mãe tem ensino fundamental ou médio completo é 17,5 e 4 vezes maior do que ter ensino superior, respectivamente. Mantida constantes as demais variáveis, a chance de o ingressante da UnB ter cursado o ensino médio em escola pública quando exerce uma atividade remunerada é 2 vezes maior do que nunca ter trabalhado, o mesmo ocorre quando os alunos não estavam trabalhando mas já exerceu alguma atividade do que nunca ter trabalhado.

A chance de o aluno da UnB ser de escola pública quando ele estuda nos *campi* Ceilândia e Planaltina é 51,3% e 153,4% maior do que quando ele estuda no Darcy Ribeiro, respectivamente.

No primeiro semestre do ano, mantida as outras variáveis constantes, a chance de o estudante da UnB ser de escola pública quando a forma de ingresso na Universidade foi pelo ENEM ou vagas remanescentes é aproximadamente 2,5 vezes maior do que ingressar pelo vestibular.

E por fim, a chance do aluno da UnB ser egresso de escola pública quando se é cor de parda é 93% maior do que quando se é de cor branca e quando se é de cor preta, 88% maior do que quando se é de cor branca.

6.2. Segundo semestre (2/2012)

No modelo para explicar o perfil do aluno que ingressa na UnB no segundo semestre tem a adição de uma variável, troca de curso, e a forma de ingresso passa a ser pelo vestibular, ENEM e transferência. A probabilidade $\pi(x_i)$ de o aluno ser classificado como estudante de escola pública em 2/2012 na UnB é dada por:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_i^M x_{i1} + \beta_j^O x_{j2} + \beta_k^C x_{k3} + \beta_l^R x_{l4} + \beta_m^I x_{m5} + \beta_1 x_n)}{1 + \exp(\beta_0 + \beta_i^M x_{i1} + \beta_j^O x_{j2} + \beta_k^C x_{k3} + \beta_l^R x_{l4} + \beta_m^I x_{m5} + \beta_1 x_n)}$$

Onde $\beta_i^M, \beta_j^O, \beta_k^C, \beta_l^R, \beta_m^I$ e β_1 são as classes observadas para os níveis das variáveis explicativas: nível de escolaridade da mãe, ocupação do aluno,

campus, cor, forma de ingresso e troca de curso se tivesse oportunidade, respectivamente.

O teste da razão de verossimilhança com a hipótese nula de que todos os parâmetros do modelo são iguais à zero foi rejeitado com p-valor menor que 0,0001, ou seja, pelo menos um dos parâmetros é diferente de zero. Porém, alguns níveis dessas variáveis não foram significativos e, por isso, seus efeitos não foram estimados. A tabela 20 mostra os valores das estatísticas dos testes de Wald e seus respectivos p-valor:

Tabela 25 - Estimativas e Estatísticas de Wald do Modelo Logístico (2/2012)

Parâmetro	GL	Estimativa	Estatística de		
			Wald	P-valor	
Intercepto	1	2,50	28,06	<.0001	
Esc. Mãe	Não lê nem escreve	1	0,08	0,01	0.9069
	Ens. Médio	1	-1,18	50,87	<.0001
	Superior Comp.	1	-2,26	159,39	<.0001
	Pós-graduação	1	-2,46	167,65	<.0001
Ocupação	Não trabalha no momento	1	0,35	9,68	0.0019
	Trabalha	1	0,72	29,94	<.0001
Campus	Planaltina	1	1,40	11,20	0.0008
	Ceilândia	1	0,35	2,63	0.1049
	Gama	1	0,23	1,50	0.2213
Ingresso	ENEM	1	0,63	15,89	<.0001
	Outros	1	-0,76	4,22	0.0400
Troca de curso	Sim	1	0,27	6,54	0.0106
Cor	Parda	1	0,08	0,13	0.7167
	Preta	1	0,50	22,85	<.0001
	Outros	1	0,79	24,46	<.0001

O teste de razão de verossimilhança encontrou-se uma estatística igual a 647,5 com 15 graus de liberdade e p-valor < 0,0001, então, a um nível de significância de 5% existem evidências para rejeitar a hipótese nula do teste e, portanto pelo menos um dos betas estimados é diferente de zero.

Nos gráfico de diagnóstico do modelo, resíduos de Pearson e a Deviance, todas as observações estão dentro do intervalo desejado e nenhuma observação se sobressai ou parece estar mal ajustada.

Gráfico 14- Resíduos de Pearson para o modelo ajustado em 2/2012

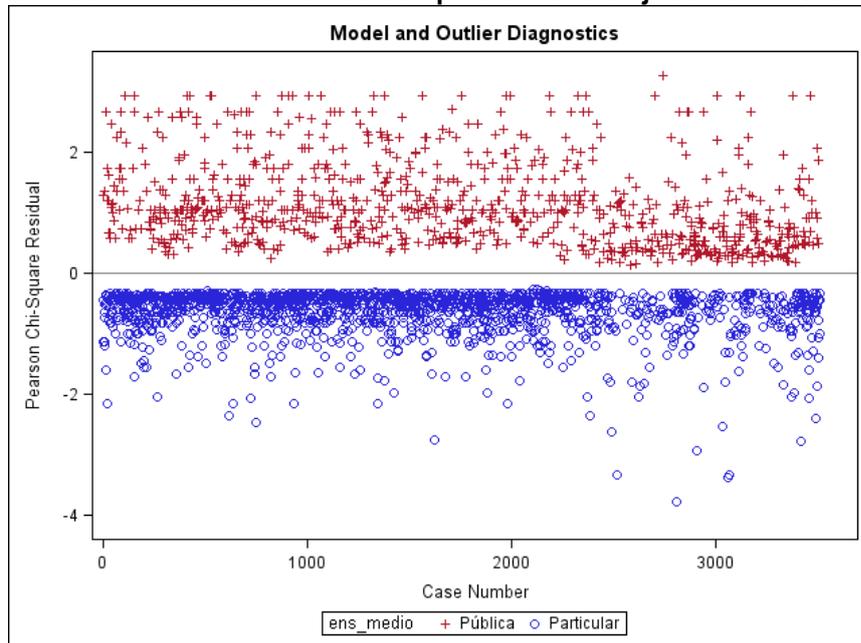
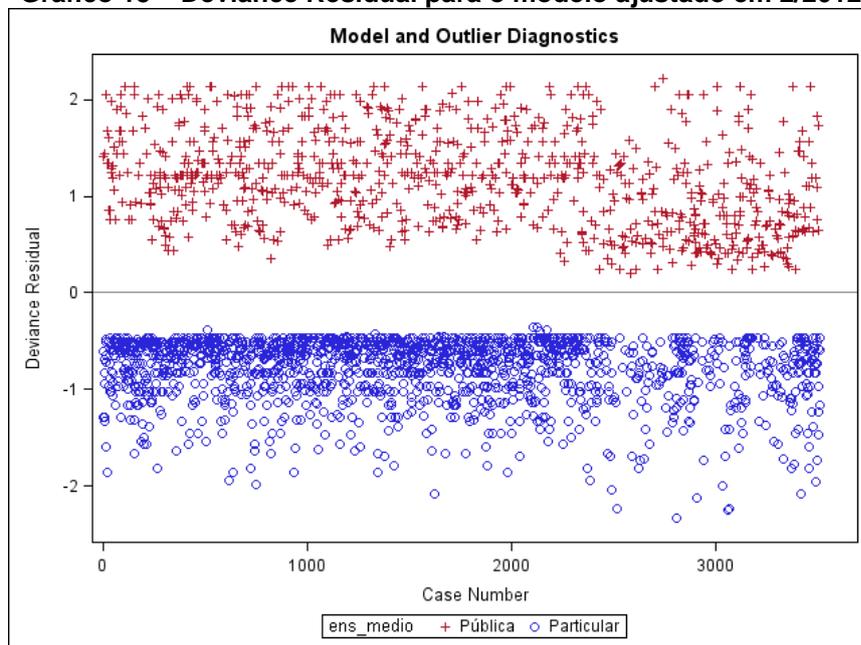


Gráfico 15 – Deviance Residual para o modelo ajustado em 2/2012



Na análise dos DfBetas, nenhuma observação representou grande influência na estimação dos parâmetros, portanto, pode-se afirmar que o modelo é adequado aos dados e partir para a análise das razões de chance.

Tabela 26 - Estimativas das Razões de Chances para as Variáveis Significativas do Modelo Logístico (2/2012).

Variáveis	Efeitos	Razão de chance	Intervalo de confiança (95%)	
Esc. Mãe	Não lê nem escreve vs Superior Comp.	10,403	2,841	38,094
	Ens. Fund. vs Superior Comp.	9,617	6,768	13,667
	Ens. Médio vs Superior Comp.	2,956	2,342	3,730
	Pós-graduação vs Superior Comp.	0,820	0,614	1,096
Ocupação	Não trabalha no momento vs Nunca trabalhou	1,419	1,138	1,769
	Trabalha vs Nunca trabalhou	2,057	1,589	2,663
Campus	Planaltina vs Darcy Ribeiro	4,075	1,790	9,277
	Ceilândia vs Darcy Ribeiro	1,419	0,930	2,165
	Gama vs Darcy Ribeiro	1,254	0,872	1,804
Ingresso	ENEM vs Vestibular	1,873	1,376	2,549
	Outros vs Vestibular	0,469	0,227	0,966
Trocaría de curso	Sim vs Não	1,314	1,066	1,619
Cor	Parda vs Branca	1,655	1,346	2,035
	Preta vs Branca	2,202	1,611	3,012
	Outros vs Branca	1,079	0,714	1,631

Observe que a chance de um ingressante da UnB ter concluído o ensino médio em escola pública cuja mãe tem ensino fundamental ou ser analfabeta é quase 10 vezes maior que a mãe ter ensino superior completo e a chance de o ingressante ser egresso de escola pública quando a mãe tem o ensino médio completo é 2,3 vezes maior do que quando tem 3º grau, mantidas constantes as outras variáveis.

A chance de o estudante da UnB ser egresso de escola pública quando exerce uma atividade remunerada é 106% maior do que nunca ter trabalhado, o mesmo ocorre para alunos que não estavam trabalhando mas já exerceu alguma atividade remunerada, 42% maior do que o aluno que nunca trabalhou, mantendo as demais variáveis constantes.

Em relação ao campus de ingresso, a chance de o aluno da UnB de escola pública quando ele estuda no *campus* de Planaltina é 4 vezes maior do que quando ele estuda no Darcy Ribeiro.

No que diz respeito à forma de ingresso na universidade, mantida as outras variáveis constantes, a chance de o estudante da UnB ser de escola pública quando a forma de ingresso na UnB foi pelo ENEM é 87,3% maior do que ter ingressado pelo vestibular tradicional. A chance de o estudante da UnB ser de

escola pública quando o aluno ingressa por outras formas (transferência, DSC, etc.) é 53% menor do que ter ingressado pelo vestibular. A chance de o aluno da UnB vindo de escola pública quando ele trocava de curso se tivesse oportunidade é 31,4% maior do que ele querer permanecer no curso escolhido.

E por último, a chance do aluno da UnB ser egresso de escola pública quando se é cor de parda é 65,5% maior do que quando se é de cor branca e quando se é de cor preta, 120% maior do que quando se é de cor branca.

7. CONCLUSÃO

O retrato geral do aluno da UnB foi montado através de diversas variáveis, iniciando pelo semestre de ingresso, seguido por *campus*, turno do curso, cotas para negros e escola que cursou o ensino médio.

No primeiro semestre quase 50% dos alunos da universidade ingressou pelo PAS (Programa de Avaliação Seriada), condição que resulta nas diferenças de um semestre para outro. Com a presença de ingressantes pelo PAS, a idade média dos alunos é menor que no segundo semestre, os brancos somam quase 54%, aproximadamente 60% são egressos de escola particular. Em contraste, no segundo semestre 86% dos alunos ingressam pelo vestibular, resultando no aumento de negros (pardos e pretos) e mais trabalhadores.

Cada *campus* da UnB tem características diferentes, o Darcy Ribeiro, primeiro *campus* construído, acumula pouco mais de 80% dos alunos da UnB, único *campus* mais de 50% dos pais e mães com ensino superior, pelo menos 60% dos alunos estudaram em escola particular no ensino médio, ministra aula para todas as ciências, e por isso tem percentual de homens e mulheres equilibrado. O *campus* do Gama é o que mais se aproxima do perfil de aluno do Darcy Ribeiro, sendo que o percentual de alunos egressos de escola particular e de homens chega a ser maior, devido à presença de somente cursos de engenharia. Os *campi* de Ceilândia e Planaltina são que mais apresentam egressos de escola pública e pais e mães com menores níveis de escolaridade, onde Planaltina atinge os extremos, mais de 75% de alunos de escolas públicas e mais de 40% dos pais com até o ensino fundamental.

Os pontos de diferenças entre os turnos estão na menor quantidade indivíduos que nunca exerceram atividade remunerada, menor quantidade pertencente a classe econômica A e idade média, aproximadamente 2 anos mais velhos.

Na análise por sistema de ingresso, entre os cotistas, o percentual de estudantes de escola pública e particular pouco difere e para os demais ingressantes, tem maioria de egressos de escola particular. Quanto a composição racial, mais da metade dos cotistas autodeclaram pretos. Apesar das cotas raciais no vestibular, cerca de 30% dos não cotistas são negros.

O estudo do perfil do aluno por tipo de escola que estudou no ensino médio foi incentivado pela lei 21.711/2012 que determina que em todas as universidades federais vinculadas ao Ministério da Educação (MEC) devem reservar pelo menos 50% das vagas dos cursos de graduação para estudantes que tenham cursado integralmente o ensino médio em escola pública. A lei começa a entrar em vigor este ano e de acordo com os resultados obtidos, os alunos de escolas públicas que ingressaram na UnB antes da reserva de vagas, já apresentam resultados favoráveis a famílias menos favorecidas. A chance de o aluno da UnB ser egresso de escola pública e pertencer a família cuja mãe não tem ensino superior é maior do que ela tenha alguma formação superior, assim como a chance de o aluno da UnB ser egresso de escola pública quando é de cor negra é maior do que ser de cor branca, quando já trabalhou ou está trabalhando é maior do que nunca ter trabalhado e quando a forma de ingresso foi o ENEM é maior do que ter ingressado na UnB pelo vestibular.

De acordo com os modelos ajustados, as variáveis que ajudam a explicar em qual instituição o aluno da UnB estudou no ensino médio no primeiro semestre são: nível de escolaridade da mãe, ocupação do aluno, campus, cor e forma de ingresso. No segundo semestre, foi adicionada a variável troca de curso se tivesse oportunidade.

Então através dos resultados deste trabalho indica-se que uma política pública voltada para o aluno de escola pública poderia proporcionar a entrada de estudantes de famílias menos favorecidas nas universidades públicas, possibilitando a médio e longo prazo o desenvolvimento econômico e social destas famílias.

Como sugestão para trabalhos futuros, seria importante acompanhar o perfil desses alunos vindos de escola pública e usar os resultados obtidos em 2012 para comparar o perfil do aluno antes e depois da lei 21.711/2012 entrar em vigor.

REFERÊNCIAS BIBLIOGRÁFICAS

AGRESTI, Alan. **An Introduction to Categorical Data Analysis**, John Wiley & Sons, New York, 1996.

ASSOCIAÇÃO BRASILEIRA DE EMPRESA DE PESQUISA, **Adoção do CCEB 2008 Critério de Classificação Econômica Brasil**, São Paulo. Disponível em: <<http://www.abep.org/novo/Content.aspx?ContentID=139>> Acesso em: 26 de novembro de 2012.

_____, **Critério de Classificação Econômica Brasil 2012**, São Paulo. Disponível em: <<http://www.abep.org/novo/Content.aspx?ContentID=30>> Acesso em: 26 de novembro de 2012.

BRASIL. Decreto nº 6.096, de 24 de abril de 2007. **Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – Reuni**. Brasília, 24 abr. 2007. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6096.htm>. Acesso em: 10 de fevereiro de 2013.

MINISTÉRIO DA EDUCAÇÃO E CULTURA. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Resumo Técnico – Censo Escolar 2010**. Brasília, 2011. 30 p.

_____. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Sinopses Estatísticas da Educação Superior – Graduação**, Brasília. Disponível em: <<http://portal.inep.gov.br/superior-censosuperior-sinopse>>. Acesso em: 29 de janeiro de 2013.

COLLET, D., **Modelling Binary Data**, Chapman e Hall, 1994.

FÓRUM NACIONAL DE PRÓ-REITORES DE ASSUNTOS COMUNITÁRIOS E ESTUDANTIS, **Perfil Socioeconômico e Cultural dos Estudantes de Graduação das Universidades Brasileiras Federais Brasileiras: Relatório Final da Pesquisa**, Brasília, 2011.

FRANCIS, A. M., TANNURI-PIANTO, M. E., **Using Brazil's Racial Continuum to Examine the Short-Term Effects of Affirmative Action in Higher Education**, *Journal of Human Resources*, 2012. 47(3): 754-784

HOSMER, D. W. e LEMESHOW, S. **Applied Logistic Regression**, John Wiley & Sons, New York, 1989.

KUTNER, M. H., NACHTSHEIM, C. J., NETER, J. e LI, W., **Applied Linear Statistical Models**. McGraw-Hill / Irwin, 2005. Quinta Edição.

STOKES, M. E., DAVIS, C. S. e KOCH, G.G., **Categorical Data Analysis Using the SAS System**, SAS Institute, 1995.

TANNURI-PIANTO, M, E., TORRES, I. O. **Estudo sobre o perfil dos candidatos e alunos da Universidade de Brasília no período pós-cotas raciais: 2004-2011**, 2011. In: apresentado ao 9º Congresso de Iniciação Científica do DF e 18º Congresso de Iniciação Científica da UnB, Brasília, 2012. Anais do 18º Congresso de Iniciação Científica, 2012

UNIVERSIDADE DE BRASÍLIA. Informações sobre a história e números da Universidade de Brasília. Disponível em: <<http://www.unb.br/sobre>> Acesso em: 10 de dezembro de 2012.