

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

DEPARTAMENTO DE ESTATÍSTICA

THALES JANUZZI FEITAL

REGRESSÃO LOGÍSTICA APLICADA A DADOS AMOSTRAIS  
COMPLEXOS: UM ESTUDO SOBRE A PERCEPÇÃO DOS JOVENS EM  
RELAÇÃO À IMPORTÂNCIA DA EDUCAÇÃO

THALES JANUZZI FEITAL

REGRESSÃO LOGÍSTICA APLICADA A DADOS AMOSTRAIS  
COMPLEXOS: UM ESTUDO SOBRE A PERCEPÇÃO DOS JOVENS EM  
RELAÇÃO À IMPORTÂNCIA DA EDUCAÇÃO

Monografia apresentada à Universidade  
Federal de Juiz de Fora como requisito  
parcial à obtenção do título de Bacharel  
em Estatística.

Orientador: Marcel de Toledo Vieira

Co-orientador: Augusto C. Souza

JUIZ DE FORA

28/11/11

THALES JANUZZI FEITAL

REGRESSÃO LOGÍSTICA APLICADA A DADOS AMOSTRAIS  
COMPLEXOS: UM ESTUDO SOBRE A PERCEPÇÃO DOS JOVENS EM  
RELAÇÃO À IMPORTÂNCIA DA EDUCAÇÃO

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
DEPARTAMENTO DE ESTATÍSTICA

Thales Januzzi Feital

Regressão logística aplicada a dados amostrais complexos: um estudo sobre a percepção  
dos jovens em relação à importância da educação

Monografia aprovada em \_\_\_\_/\_\_\_\_/\_\_\_\_ para obtenção do título de Bacharel  
em Estatística.

Banca Examinadora:

-----  
Prof. Marcel de Todelo Vieira

-----  
Prof.(a) Camila Borelli Zeller

-----  
Prof. Ronaldo Rocha Bastos

Dedico esta monografia a todos  
que estiveram ao meu lado  
durante estes anos.

## Agradecimentos:

Agradeço a meus pais Maurício e Siléia e aos meus irmãos Rhenan e Joyce pelo apoio dado durante todo este processo de minha vida, a minha namorada Monique pela compreensão e colaboração principalmente nestes momentos finais, ao meu amigo Patrick por sempre estar à disposição para ajudar sempre que possível e a todos os meus familiares que sempre estiveram ao meu lado.

Também gostaria de agradecer aos meus colegas de faculdade pelo companheirismo e ajudas prestadas sempre que necessário. Em especial, gostaria de agradecer ao Vitor, Thiago, Samuel, Jarbas, Marcos e Lu.

Sou grato a todos os professores do Departamento de Estatística da UFJF pelo tempo dedicado e disposição em ajudar sempre que necessário. Em especial ao Professor Clécio por sua dedicação como coordenador do curso, ao Professor Luiz Cláudio, Professor Mário Sérgio e a Mestra Maria Evangelista pela atenção e ajuda durante a realização de um projeto de pesquisa que tanto colaborou para meu crescimento profissional e acadêmico, ao Professor Marcel tanto por sua dedicação ao departamento quanto pela paciência e tempo gastos em minha orientação para esta monografia e neste caso também ao Professor Augusto por sua grande ajuda.

“O conhecimento nos torna sábios,  
A fé e o amor nos tornam humanos.”

Autor desconhecido.

## RESUMO

A educação é de grande importância para o desenvolvimento dos jovens. Atualmente, todas as esferas de governo vem demonstrando preocupação no que diz respeito à qualidade e à universalização do acesso à educação, que vem motivando uma série de políticas públicas. Segundo resultados da pesquisa *Motivos da Evasão Escolar*, realizada com dados levantados nas principais capitais do País com o apoio da Fundação Getulio Vargas e de outras instituições, em 2006, 2,6% dos jovens com idade entre 10 e 14 anos estavam fora da escola, sendo que dos 15 aos 17 anos essa frequência é de 17,8% e a partir dos 18 anos quase 50% dos alunos abandonam seus estudos. Além disso, segundo o mesmo estudo, a proporção de jovens que trabalham segue um caminho oposto, sendo 13% de jovens com 10 anos, tendo um aumento para 37% aos 17 anos e 54% dos jovens com 18 anos estão presentes no mercado de trabalho. Um resultado que chama atenção são os motivos para a evasão escolar, dentre os quais se destacam problemas como baixa renda, pouca oferta de oportunidades em instituições de ensino e outros motivos, e o simples fator “Falta de interesse” que obteve mais de 40% das respostas. Levando-se em conta que os jovens estão saindo da escola por falta de interesse, em sua maioria, e que estão ingressando no mercado de trabalho, surge nossa questão de interesse: Será que os jovens hoje em dia, não vêem a educação como algo de importância para sua vida profissional?

Esta monografia tem por objetivo avaliar este aspecto analisando a população de jovens entre 14 e 24 anos do estado de Minas Gerais, por meio dos dados da Pesquisa por Amostra de Domicilio de Minas Gerais de 2009 (PAD-MG). A PAD-MG 2009 busca tem como objetivo estudar a população de Minas Gerais incluindo suas características sócio-econômicas, com o objetivo de auxiliar a implementação e a manutenção de políticas publicas. Alguns indicadores produzidos pela PAD-MG não são fornecidos por outras pesquisas do gênero existentes, como por exemplo, a percepção do jovem quanto à importância da educação. A amostra foi selecionada através de um plano amostral complexo incluindo estratificação e seleção em múltiplos estágios de amostragem estratificada. Adotamos o método de Análise de Regressão Logística, levando-se em consideração os pesos amostrais, com o objetivo de identificar o perfil dos jovens quanto a sua percepção da importância da educação para a sua carreira profissional.

Ao final desta monografia foi possível observar que possível observar que o desenho amostral causa um efeito nos resultados da análise

**Palavras-chave:** *análise de regressão logística, amostragem complexa, percepção sobre a importância da educação*



## Sumário

Índice de tabelas e quadros.....	11
Capítulo 1.....	12
1.1-Introdução .....	12
1.2-Objetivos .....	13
1.3-Descrição dos capítulos .....	14
Capítulo 2.....	15
2.1-Pesquisa por Amostragem de Domicílios de Minas Gerais 2009 .....	15
2.1.1-Descrição das variáveis .....	17
2.1.2-Plano amostral da PAD-MG 2009 .....	21
2.1.3-Pesos Amostrais da PAD-MG 2009 .....	22
2.1.4-Variáveis utilizadas.....	23
2.1.5-Registros excluídos do banco de dados e problemas encontrados .....	44
Capítulo 3.....	46
3.1-Metodologia .....	46
3.1.1-Regressão logística.....	46
3.1.2-Regressão logística para amostragem complexa .....	50
Capítulo 4.....	53
4.1-Aplicações e Resultados .....	53
4.1.1-Análise de Regressão Logística Simples.....	53

4.1.2-Primeiro modelo de regressão .....	60
4.1.3-Segundo modelo de regressão.....	68
Capítulo 5.....	74
5.1-Conclusões .....	74
Bibliografia .....	76

## Índice de tabelas e quadros

Tabela 1 - Análise de regressão logística considerando e não considerando o plano amostral.....	54
Tabela 2 - Modelo final considerando e não considerando o plano amostral .....	62
Tabela 3 - Análise de Regressão Logística com e sem o plano amostral com as mesmas variáveis.....	69
Quadro 1 - Seções e Público Alvo da PAD-MG 2009 .....	16
Quadro 2 - Variáveis Utilizadas.....	25
Quadro 3 - Variáveis retiradas por apresentarem frequência alta de dados faltantes.....	59

## Capítulo 1

### 1.1-Introdução

A educação é considerada atualmente como sendo de grande importância para o desenvolvimento humano. Além disso, a sociedade vem cada vez mais demonstrando grande preocupação quanto à qualidade da educação, o que vem mobilizando políticas públicas com o intuito de aperfeiçoar o sistema público de educação e também disponibilizá-lo de forma universal a toda a população.

Segundo dados publicados pelo estudo *Motivos da Evasão Escolar* (NERI, et al., 2009), realizado a partir de dados coletados nas principais capitais do País com o apoio da Fundação Getúlio Vargas e de outras instituições, no ano de 2006, 2,6% dos jovens com idade entre 10 e 14 anos estavam fora da escola, sendo que dos 15 aos 17 anos essa proporção era de 17,8% e a partir dos 18 anos quase 50% dos jovens que deveriam completar seus estudos estavam fora da escola. Em comparação, segundo o mesmo estudo, a proporção de jovens que trabalham também possui a mesma tendência de aumento para faixas etárias mais velhas, sendo 13% de jovens com 10 anos, tendo um aumento para 37% aos 17 anos e 54% dos jovens com 18 anos estavam presentes no mercado de trabalho.

Um resultado que chama atenção se refere aos motivos para a evasão escolar, que incluem problemas tais como renda, pouca oferta de oportunidades em instituições de ensino e outros motivos. O fator “Falta de interesse”, por exemplo, obteve mais de 40% das respostas (NERI, et al., 2009). Levando-se em conta que estes jovens que estão saindo da escola por falta de interesse, em sua maioria, estão ingressando no mercado de trabalho, surge uma questão

interessante: os jovens não veem, hoje em dia, a educação como algo importante para sua vida profissional?

## 1.2-Objetivos

Este estudo tem por objetivo:

- avaliar a questão levantada acima a partir da população de jovens entre 14 e 24 anos do estado de Minas Gerais
- avaliar empiricamente a importância de considerar o desenho amostral nas análises empreendidas, salientando que a não consideração do desenho amostral poderia acarretar em conclusões equivocadas.

Para isso foi utilizado o banco de dados da Pesquisa por Amostra de Domicílio de Minas Gerais de 2009 (PAD-MG) que busca conhecer a população mineira bem como suas características, ações, posição social, etc, com o objetivo de auxiliar a implementação e a manutenção de políticas públicas (Fundação João Pinheiro, 2009). A PAD-MG produz informações socioeconômicas, sendo que alguns de seus indicadores não são fornecidos por outras pesquisas do gênero existentes, como por exemplo, a percepção do jovem quanto à importância da educação (Fundação João Pinheiro, 2009).

A amostra foi selecionada a partir do método de amostragem estratificada e conglomerada em dois estágios, tendo como níveis de estratificação as Regiões de Planejamento e Mesorregiões do estado (SUYAMA, 2008). Esta pesquisa começou a ser elaborada no ano de 2007 pela equipe do programa Estado para Resultados (EpR) com a colaboração da CEDEPLAR/UFMG, Fundação Science e Fundação João Pinheiro. O instrumento de coleta foi baseado na

Pesquisa Nacional por Amostra de Domicílios do Instituto Brasileiro de Geografia e Estatística (IBGE), no questionário de Avaliação de Impacto do Bolsa Família (2005), na Pesquisa de Orçamento Familiar (2002/2003), na Munic (2005), no Inquérito Domiciliar sobre Comportamentos de Risco e Morbidade Referida de Doenças e Agravos não Transmissíveis (2003) e na Pesquisa de Públicos do Porto (2001) (Fundação João Pinheiro, 2009).

Neste trabalho, utilizamos o método de regressão logística com a finalidade de identificar variáveis que possam apresentar alguma relação com a variável resposta dicotômica “Você acha que o que você aprendeu ou viveu na escola foi muito importante, pouco importante ou nada importante para o seu futuro profissional” onde 1 é muito importante e 0 é nada ou pouco importante e também avaliar os efeitos do plano amostral da PAD-MG.

### **1.3-Descrição dos capítulos**

Esta monografia está organizada da seguinte maneira. No capítulo 2, será apresentada uma descrição sobre a PAD-MG 2009, sua população alvo, seu objetivo, o plano amostral utilizado e os pesos amostrais calculados. No capítulo 3, será abordado o método utilizado, regressão logística binária, bem como os métodos de estimação capazes de considerar o desenho amostral nas análises. No capítulo 4, serão apresentados os resultados das análises realizadas no presente estudo. No capítulo 5, serão apresentadas as considerações finais e as conclusões desta monografia.

## Capítulo 2

### 2.1-Pesquisa por Amostragem de Domicílios de Minas Gerais 2009

Na PAD-MG 2009 foi entrevistada uma amostra de residentes no estado de Minas Gerais, selecionada com o objetivo de permitir inferências sobre características socioeconômicas da população mineira e apresentar indicadores importantes para políticas públicas no estado (Fundação João Pinheiro, 2009).

A coleta dos dados foi realizada entre junho e novembro de 2009 para uma amostra de 18 mil domicílios distribuídos em 308 municípios de Minas Gerais. A amostra coletada é representativa para os estratos Urbano/rural; Região Metropolitana de Belo Horizonte/ Não metropolitano; Regiões de planejamento (Noroeste, Norte, Rio Doce, Mata, Sul, Triângulo, Alto Paranaíba, Centro-Oeste, Jequitinhonha/Mucuri e Central; Belo Horizonte/Demais municípios de Minas Gerais e Mesorregiões de Minas Gerais (Fundação João Pinheiro, 2009).

Os pesquisadores de campo possuíam um computador portátil com o questionário programado em um pacote de domínio público desenvolvido pelo Bureau de Censo norte americano denominado CSpro. Para responder ao questionário, buscou-se preferencialmente o responsável pelo domicílio ou seu cônjuge (Fundação João Pinheiro, 2009).

Os questionários foram divididos em diferentes seções e aplicados a públicos específicos como informado no Quadro 1.

## Quadro 1 - Seções e Público Alvo da PAD-MG 2009

Características do domicílio	Todos os moradores
Características dos moradores	Todos os moradores
Assistência social	Todos os moradores
Gastos coletivos	Todos os moradores
Gastos coletivos – Alimentos e Bebidas	Todos os moradores
Educação	Todos os moradores
Gastos individuais	Todos os moradores
Renda exceto trabalho e benefícios	Todos os moradores
Trabalho	A partir de 5 anos de idade
Saúde	Todos os moradores
Saúde – condições de gravidez	Mulheres entre 10 e 49 anos
Saúde – condições de saúde infantil	Menores de 7 anos
Saúde – hábitos de vida saudável	Maiores de 14 anos
Saúde – acesso aos serviços preventivos de saúde feminina	Mulheres maiores de 25 anos
Juventude	Entre 14 e 24 anos

Fonte: (Fundação João Pinheiro, 2009)

O banco de dados foi dividido em oito arquivos equivalentes às seções indicadas:

Seção A\_domicilio; Seção B\_caracteristicas\_moradores; Seção C\_educacao;



Seção D\_saude; Seção E\_trabalho; Seção F\_rendimentos; Seção J\_gastos individuais e Seção K\_juventude (Fundação João Pinheiro, 2010).

### 2.1.1-Descrição das variáveis

Foram observadas por meio deste estudo muitas características dos domicílios e dos moradores.

Foi definido como domicílio o local de moradia independente (sem acesso por outro domicílio) possuindo um ou mais cômodos separados fisicamente por algum material. Os residentes dos domicílios eram moradores que o habitavam na data da pesquisa ou estavam ausentes por no máximo 12 meses (Fundação João Pinheiro, 2009).

Foram excluídas pessoas residentes em embaixadas, consulados e aquelas institucionalizadas, ou seja, residentes em domicílios coletivos de estabelecimentos institucionais (militares em instalações militares, presos em penitenciárias, internos em escolas, etc.) (Fundação João Pinheiro, 2009).

Os domicílios foram classificados como Urbano ou Rural tendo como base a legislação vigente segundo o Censo demográfico de 2000, e como Casa, Apartamento, Quarto ou Cômodo. Perguntou-se se o domicílio era Próprio, Alugado ou outra situação. Observou-se também, as características das calçadas e das ruas, bem como o local onde o terreno se encontrava; o material presente na construção externa das paredes do domicílio, do piso e o telhado; o número de cômodos existentes e existência de água, luz, esgoto e coleta de lixo (seletiva ou não); e se o domicílio recebeu visita de um agente sanitário de saúde e se

recebeu algum benefício social e seu valor; além da presença de itens como fogão, geladeira, etc (Fundação João Pinheiro, 2009).

Quanto aos residentes, foram pesquisadas as características das famílias e dos membros incluídos em cada família; perguntou-se idade, Estado Civil, Cor ou Raça, Nacionalidade e Naturalidade; posse de documentos como CPF, RG e Título de eleitor; se aplicável, perguntou-se ao responsável pela família o motivo de outra família presente e se alguém tinha a intenção de se mudar; e para mulheres acima de 10 anos, foi observada a fecundidade (Fundação João Pinheiro, 2009).

Também foram observadas características quanto à educação, classificando os indivíduos como alfabetizados ou não; qual a frequência a uma instituição de ensino e se já frequentou alguma no passado; pesquisou-se o nível de grau de ensino concluído por cada indivíduo e, se aplicável, a duração do ensino fundamental; se a rede de ensino era pública ou privada; para os que estavam frequentando um curso superior no momento da pesquisa, perguntou-se a área do curso, a duração, o período que se encontrava e o turno; existência de algum programa social do governo para alunos ou famílias dos alunos; quanto à existência de merendas gratuitas e da frequência em que os alunos se beneficiam; tipo de transporte utilizado e o tempo gasto; também se deixou de comparecer a sua instituição de ensino nos últimos 30 dias e o motivo; a idade dos moradores quando iniciaram a primeira série (primeiro ano do ensino fundamental); existência de um responsável que acompanha as atividades escolares em geral; o curso mais elevado que frequentou e, se aplicável, o motivo

de parar de frequentar a escola; realização de cursos profissionalizantes, área do curso e aplicação do curso no trabalho (para maiores de 10 anos); e os gastos individuais com a educação (Fundação João Pinheiro, 2009).

Quanto à saúde dos moradores, foi pesquisada a percepção quanto ao estado de saúde; existência de pessoas com problemas crônicos de saúde que exigem acompanhamento constante (Hipertensão, diabetes, AIDS, etc.); presença de algum sintoma diferente nos últimos 30 dias; procura de atendimento de saúde nos últimos 30 dias, o tipo de profissional que realizou o atendimento, o principal motivo que levou ao morador a buscar atendimento, local do atendimento e se foi realizado um pagamento pelo atendimento e seu valor; se recebeu atendimento pelo SUS nos últimos 30 dias; se foi necessário deslocar para outro município a fim de receber atendimento, tipo de transporte utilizado e tempo gasto; motivo de não ter procurado atendimento nos últimos 30 dias; se houve restrição das atividades habituais por motivos de saúde; necessidade de internação para algum tipo de tratamento ou procedimento de saúde; possui plano de saúde; gastos com saúde; para mulheres de 10 a 49 anos, existência de gravidez, o mês de gestação e o número de consultas pré-natal; para menores de 7 anos, peso ao nascer, ocorrência de problemas ao nascimento, acompanhamento de equipe de saúde, tempo de aleitamento, vacinação e responsável pelos cuidados da criança na ausência da mãe; prática de atividades físicas; para maiores de 14 anos, consumo de bebidas alcoólicas, cigarros, idade que começou a fumar ou, se aplicável, parou de fumar; foi perguntado sobre exames para medir pressão arterial, colesterol, glicose, Papanicolau, mama, mamografia e o

tempo desde o último exame realizado; presença de doença coronariana; recebimento de orientação nutricional; consumo de frutas, legumes, verduras, carne vermelha, frango e leite (Fundação João Pinheiro, 2009).

No que diz respeito ao trabalho, a pesquisa buscou conhecer o tipo de trabalho exercido pelos moradores, sendo de carteira assinada ou outras posições possíveis e o cargo/função exercido; tendo a pessoa mais de um trabalho, classificou-se qual seria o principal e qual a atividade do estabelecimento em que trabalha; observou-se o tempo dos contratos de trabalho com prazo determinado; se os autônomos estavam formalizados com seus registros (CNPJ); se contribuía para previdência; qual a jornada semanal de trabalho; experiência no trabalho; se era sindicalizado; as instalações da empresa onde trabalha; transporte utilizado pra ir ao trabalho, gastos e tempo; se procurou/procura trabalho de alguma forma; e se toma tempo com afazeres domésticos (Fundação João Pinheiro, 2009).

Quanto ao rendimento, procurou-se saber o rendimento proveniente de trabalho e de outras fontes (Fundação João Pinheiro, 2009).

Foi observado se algum morador recebe ou recebeu algum serviço de assistência social gratuito (Fundação João Pinheiro, 2009).

Com respeito aos gastos, a pesquisa buscou conhecer os gastos coletivos do domicílio sendo realizados com pagamentos monetários ou não, na aquisição de bens de qualquer gênero. Também se pesquisou a respeito dos gastos individuais (Fundação João Pinheiro, 2009).

Para a seção juventude, buscaram-se conhecer a frequência das práticas culturais domésticas, de sociabilidade e passeios; se participa em programas ou projetos sociais; sua percepção quanto à importância da escola e da visão do jovem com respeito ao trabalho; se tem ou teve alguma vez participação em atividades políticas; e se votou nas últimas eleições e sobre a lembrança dos candidatos em quem votou. (Fundação João Pinheiro, 2009)

### 2.1.2-Plano amostral da PAD-MG 2009

O estado de Minas Gerais é dividido em 10 Regiões de Planejamento ou 12 Mesorregiões. Sendo a população de interesse os residentes em domicílios (unidade amostral) mineiros, o cadastro para a seleção da amostra foi composto pelos setores censitários do Censo Demográfico de 2000, sendo eliminados setores que continham apenas por domicílios coletivos (quartéis, bases militares, aldeias indígenas, etc.) (SUYAMA, 2008).

A fim de realizar uma amostragem estratificada e conglomerada em dois estágios, o estado foi dividido inicialmente em 12 estratos, sendo eles, Belo Horizonte, Resto da Metropolitana, Central Mineira, Oeste de Minas, Campo das Vertentes, Zona da Mata, Sul/Sudoeste de Minas, Triângulo, Alto Paranaíba, Noroeste de Minas, Norte de Minas, Jequitinhonha, Vale do Mucuri e Vale do Rio Doce (SUYAMA, 2008).

Foi observado que a agregação de dois ou mais estratos, melhoraria as estimativas, de forma que foram realizados os seguintes agrupamentos: Jequitinhonha/Mucuri, Centro-Oeste (Oeste de Minas e Central Mineira), Sul

(Sul/Sudeste e Campo das Vertentes), Central (Metropolitana, Campos das Vertentes, Central Mineira e Jequitinhonha (SUYAMA, 2008).

A amostra de 18000 domicílios foi distribuída por meio do seguinte método de alocação :

( 1 )

$$n_h = n \frac{\frac{S_h Y_h^q}{\bar{Y}_h}}{\sum_h \frac{S_h Y_h^q}{\bar{Y}_h}}$$

onde,  $S_h$  é o desvio padrão populacional no estrato  $h$ ,  $Y_h$  é a renda total no estrato  $h$ ,  $q$  é uma potencia que varia entre 0 e 1. Neste caso este valor foi obtido para que fosse encontrado um  $CV$  que se mantivesse abaixo de 0,15:  $q = 0,2$  (SUYAMA, 2008).

Dentro de cada estrato existem as microrregiões (sub estratos), onde foram distribuídos o tamanho da amostra para cada sub estrato pela alocação ótima de Neyman (SUYAMA, 2008).

### 2.1.3-Pesos Amostrais da PAD-MG 2009

Para o cálculo dos pesos amostrais, buscou-se respeitar ao máximo as características do plano amostral adotado. Além disso, foram realizados ajustes para a correção de não resposta e ajustes através de métodos de calibração com o objetivo de garantir que totais amostrais calculados para algumas variáveis coincidisse com os respectivos totais populacionais conhecidos (SILVA & VIEIRA, 2010) .

Sendo o plano amostral em dois estágios, os pesos amostrais também foram calculados em duas etapas. Levou-se em consideração, por exemplo, o fato de

que foram selecionados para a pesquisa 1200 setores dos quais somente 1157 foram entrevistados, onde os demais setores foram perdidos por diversas razões (SILVA & VIEIRA, 2010).

Na primeira etapa foram calculados pesos amostrais para cada setor selecionado e posteriormente foi realizada uma correção dos pesos para os setores que de fato foram entrevistados. Os pesos básicos de domicílios foram calculados através do produto dos pesos ajustados dos setores respondentes e pesos calculados como o inverso da probabilidade de seleção dos domicílios em cada setor. Foi realizado um ajuste para não respostas para os pesos de domicílios (SILVA & VIEIRA, 2010).

Na segunda etapa, foi aplicado o método de calibração. Este processo busca que os pesos dos domicílios ajustados quando multiplicados por fatores de correção, produzam novos pesos que reproduzam exatamente a contagem populacional considerada na calibração, para as seguintes categorias: classes de tamanhos dos municípios (pequenos e grandes); população por sexo, faixa etária e população por situação urbana/rural em municípios pequenos (SILVA & VIEIRA, 2010).

Buscou-se produzir pesos calibrados que fossem o mais próximos o possível dos pesos corrigidos com ajustes para não respostas, e que os mesmos quando aplicados aos dados que produzem estimativas de contagens populacionais próximas aos valores verdadeiros (SILVA & VIEIRA, 2010).

#### **2.1.4-Variáveis utilizadas**

Para o presente estudo foram selecionadas variáveis pertencentes às seções Juventude, Educação, Moradores, Características do domicílio e Saúde para

indivíduos de faixa etária entre 14 e 24 anos (NERI, et al., 2009) sendo as variáveis estudadas descritas no Quadro 2 abaixo.



Quadro 2 - Variáveis Utilizadas

Quantos dias por semana você pratica as seguintes atividades	Ouvir música (CD, DVD, MP3 etc.)	
	Ver televisão	
	Ouvir rádio	
	Ler jornais e/ou revistas	
	Ler livros (excluindo didáticos e profissionais)	
	Praticar esportes	
	Dançar	
	Utilizar suportes de multimídia (CDRom, pen drive, jogos etc.) (excluindo internet)	
	Utilizar internet	
Quantos dias por mês você pratica as seguintes atividades	Ir ao cinema	
	Receber / visitar amigos ou parentes	

(exceto para o trabalho)	Passear em espaços ao ar livre	
	Ir ao shopping	
	Ir à missa ou a cerimônias religiosas	
	Ir a discotecas / bailes e ou bares / restaurantes	
	Assistir eventos esportivos em campos ou quadras	
	Frequentar associações recreativas ou clubes	
E quantas vezes por ano você pratica as seguintes atividades	Ver exposições	
	Ir a bibliotecas	
	Ir a museus	
	Ir ao teatro	
	Ir a concertos de música erudita / clássica	
	Ir a shows de música popular / moderna	
	Ir a espetáculos de dança	

	Ver espetáculos de rua	
	Ir a festas populares	
Programa social	Excetuandose programas sociais que transferem dinheiro, você participa ou é beneficiário de algum programa ou projeto social?	<ol style="list-style-type: none"> <li>1. Sim</li> <li>2. Não</li> </ol>
Você acha que o que você aprendeu ou viveu na escola foi muito importante, pouco importante ou nada importante	Para o seu futuro profissional	<ol style="list-style-type: none"> <li>1. Nada Importante</li> <li>2. Pouco Importante</li> <li>3. Muito Importante</li> </ol>
Educação	Quantos anos tinha quando iniciou a primeira série do ensino fundamental (primeiro ano primário ou 1º grau)?	

	<p>Qual foi o curso mais elevado que frequentou no ano passado?</p>	<p>01. Ensino fundamental ou 1º grau regular seriado</p> <p>02. Ensino médio ou 2º grau regular seriado</p> <p>03. Supletivo/EJA (ensino fundamental ou 1º grau)</p> <p>04. Supletivo/EJA (ensino médio ou 2º grau)</p> <p>05. Superior graduação</p> <p>06. Creche</p> <p>07. Ensino fundamental ou 1º grau regular não seriado</p> <p>08. Ensino médio ou 2º grau regular não seriado</p> <p>09. Alfabetização de adultos (AJA)</p> <p>10. Classe de Alfabetização</p>
--	---	--

		<p>11. Maternal, jardim de infância</p> <p>12. Pré vestibular</p> <p>13. Mestrado, doutorado ou especialização</p>
	<p>Qual é o curso que frequenta atualmente?</p>	<p>01. Ensino fundamental ou 1º grau regular seriado</p> <p>02. Ensino médio ou 2º grau regular seriado</p> <p>03. Supletivo/EJA (ensino fundamental ou 1º grau)</p> <p>04. Supletivo/EJA (ensino médio ou 2º grau)</p> <p>05. Superior graduação</p> <p>06. Creche</p> <p>07. Ensino fundamental ou 1º grau regular não seriado</p> <p>08. Ensino médio ou 2º grau regular não</p>

		<p>seriado</p> <p>09. Alfabetização de adultos (AJA)</p> <p>10. Classe de Alfabetização</p> <p>11. Maternal, jardim de infância</p> <p>12. Pré vestibular</p> <p>13. Mestrado, doutorado ou especialização</p>
	<p>Na escola em que estuda, funciona algum programa social do governo estadual para os alunos ou para as famílias dos alunos?</p>	<p>1. Sim</p> <p>2. Não</p>
	<p>Com qual idade deixou de frequentar escola?</p>	
	<p>Teve algum gasto com educação com em 2009, como mensalidades escolares, transporte para ir à escola, material escolar, uniformes, etc.</p>	<p>1. Sim</p> <p>2. Não</p>

	<p>Quem acompanha as atividades escolares?</p>	<ol style="list-style-type: none"> <li>1. Pai</li> <li>2. Mãe</li> <li>3. Pai e mãe</li> <li>4. Irmão/Irmã</li> <li>5. Avô ou avó</li> <li>6. Outro parente</li> <li>7. Outra pessoa</li> <li>8. Ninguém</li> </ol>
	<p>A escola em que estuda oferece merenda gratuita?</p>	<ol style="list-style-type: none"> <li>1. Sim</li> <li>2. Não</li> </ol>
	<p>Qual o principal motivo pelo qual deixou de ir à escola, pelo menos um dia, nos últimos 30 dias?</p>	<ol style="list-style-type: none"> <li>01. Problemas de saúde/doença</li> <li>02. Precisou trabalhar</li> <li>03. Precisou ajudar o pai/mãe no trabalho</li> <li>04. Precisou ficar em casa para olhar os irmãos</li> </ol>

		<p>05. Precisou ficar em casa para cuidar de pessoa doente</p> <p>06. Não gosta da escola</p> <p>07. Não gosta do professor</p> <p>08. Não quer estudar/Não se sente motivado</p> <p>09. Não teve quem o(a) levasse</p> <p>10. Problema de violência na vizinhança da escola ou do domicílio</p> <p>11. Violência dentro da escola</p> <p>12. Outro motivo</p>
	Sabe ler e escrever um bilhete simples no idioma que conhece?	<p>1. Sim</p> <p>2. Não</p>
	Alguma vez você já parou de estudar e depois voltou?	<p>1. Sim</p> <p>2. Não</p>



	<p>Por que não frequenta escola?</p>	<ol style="list-style-type: none"> <li>01. Precisa trabalhar</li> <li>02. Precisa ajudar pai/mãe no trabalho</li> <li>03. Precisa tomar conta de irmãos</li> <li>04. Não tem escola perto</li> <li>05. É caro</li> <li>06. Já formou/terminou os estudos</li> <li>07. Engravidou/casou</li> <li>08. Doença</li> <li>09. Não gosta de estudar</li> <li>10. Atividades domésticas</li> <li>11. Não quer</li> <li>12. Idade avançada / idade muito jovem</li> <li>13. Outro Motivo</li> </ol>
	<p>Curso de educação profissional permitiu melhorar</p>	<ol style="list-style-type: none"> <li>1. Sim</li> </ol>

	sua posição no seu trabalho?	2. Não
	Estuda neste município?	1. Sim 2. Não, em outro município 3. Não, em país estrangeiro 4. Não estuda
	Qual é o curso ou grau mais elevado que frequentou, no qual concluiu pelo menos uma série? (Grau mais elevado na moradia)	01. Ensino fundamental ou 1º grau PASSE AO 36 02. 1º Ciclo do fundamental (antigo primário) 03. 2º Ciclo do fundamental (antigo ginásio) 04. Ensino médio ou 2º grau 05. Supletivo/EJA (ensino fundamental ou 1º grau) 06. Supletivo/EJA (ensino médio ou 2º grau) 07. Superior graduação

		<p>08. Maternal, jardim de infância, etc.</p> <p>09. Classe de Alfabetização / Pré-primário</p> <p>10. Alfabetização de adultos (AJA)</p> <p>11. Pré-vestibular</p> <p>12. Mestrado, doutorado ou especialização</p> <p>00. Nunca concluiu uma série</p>
Transporte	<p>Qual o principal meio de transporte habitualmente utilizado por para ir até a creche ou escola?</p>	<p>01. Ônibus público ou van de linha</p> <p>02. Trem ou metrô</p> <p>03. Integração metrô/ônibus</p> <p>04. Carro ou moto particular</p> <p>05. Ônibus ou van fretados</p> <p>06. Transporte não motorizado (bicicleta, cavalo, canoa, etc.)</p> <p>07. Transporte fluvial motorizado</p>

		08. Nenhum (a pé) 09. Outro
	Quanto tempo, em horas e minutos, leva no percurso de ida à escola?	
Percepção do trabalho e sociedade	A palavra que mais se aproxima do que você pensa sobre trabalho é:	1. Necessidade 2. Auto realização 3. Independência 4. Crescimento 5. Exploração
	A palavra que fica em 2º lugar do que você pensa sobre trabalho é:	1. Necessidade 2. Auto realização 3. Independência 4. Crescimento 5. Exploração

	<p>Vou ler duas afirmações. Após ouvi-las, por favor, me diga com qual dessas afirmações você mais concorda, se com a primeira ou com a segunda.</p>	<ol style="list-style-type: none"> <li>1. Primeira: “Devemos priorizar o crescimento da economia e a criação de empregos mesmo se o meio ambiente for prejudicado”</li> <li>2. Segunda: "Devemos proteger o meio ambiente, mesmo se isso diminuir o crescimento da economia”</li> </ol>				
<p>Gostaria que você me dissesse se você participa, se já participou ou se nunca participou de</p>	<table border="1"> <tr> <td data-bbox="638 710 1375 847">Associação comunitária ou de moradores</td> <td data-bbox="1375 710 2063 1307" rowspan="4"> <ol style="list-style-type: none"> <li>1. Participa</li> <li>2. Já participou</li> <li>3. Nunca participou</li> </ol> </td> </tr> <tr> <td data-bbox="638 847 1375 984">Partido político</td> </tr> <tr> <td data-bbox="638 984 1375 1144">Grupo de promoção e proteção dos direitos de grupos minoritários</td> </tr> <tr> <td data-bbox="638 1144 1375 1307">Sindicato de trabalhadores ou associação profissional</td> </tr> </table>	Associação comunitária ou de moradores	<ol style="list-style-type: none"> <li>1. Participa</li> <li>2. Já participou</li> <li>3. Nunca participou</li> </ol>	Partido político	Grupo de promoção e proteção dos direitos de grupos minoritários	Sindicato de trabalhadores ou associação profissional
Associação comunitária ou de moradores	<ol style="list-style-type: none"> <li>1. Participa</li> <li>2. Já participou</li> <li>3. Nunca participou</li> </ol>					
Partido político						
Grupo de promoção e proteção dos direitos de grupos minoritários						
Sindicato de trabalhadores ou associação profissional						

	Associação estudantil, grêmio, centro acadêmico ou união de estudantes	
	Clube ou associação esportiva, recreativa ou de lazer	
	Grupo religioso	
	ONG (Organização Não Governamental)	
	Grupos artístico-culturais	
	Conselho de política pública	
	Grupo temporário para resolver problema local imediatamente	
Características da moradia	Quantos cômodos estão servindo	

	permanente de dormitório para os moradores deste domicílio?	
	Número de moradores.	
Saúde	Tem algum problema de saúde que exige acompanhamento constante?	01. Sim, hipertensão (pressão alta) 02. Sim, diabetes 03. Sim, obesidade 04. Sim, tuberculose 05. Sim, hanseníase (lepra) 06. Sim, doença neurológica (Afasias, Alzheimer, Doença de Huntigton, Esclerose múltipla, Parkinson, Síndromes) 07. Sim, AIDS 08. Sim, câncer

		<p>09. Sim, atraso no desenvolvimento (para crianças até 12 anos)</p> <p>10. Sim, depressão</p> <p>11. Sim, insuficiência renal</p> <p>12. Sim, outro</p> <p>13. Não</p>
Moradores	Até 31 de maio de 2009, teve quantos filhos nascidos vivos (ou seja, que apresentaram algum sinal de vida ao nascer)?	
	Cor ou raça	<p>1. Branca</p> <p>2. Preta</p> <p>3. Parda</p> <p>4. Amarela</p> <p>5. Indígena</p>



	Qual é o estado civil atualmente?	<ol style="list-style-type: none"> <li>1. Casado/a</li> <li>2. Desquitado/a ou separado/a judicialmente</li> <li>3. Divorciado/a</li> <li>4. Viúvo/a</li> <li>5. Solteiro/a</li> </ol>
	Qual é a idade em anos completos ou a idade presumida?	
	Sexo	<ol style="list-style-type: none"> <li>1. Masculino</li> <li>2. Feminino</li> </ol>
	Qual a relação de convivência que tem com o responsável pelo domicílio?	<ol style="list-style-type: none"> <li>01. Pessoa responsável</li> <li>02. Cônjuge/ companheiro(a)</li> <li>03. Filho(a), enteado(a)</li> <li>04. Pai, mãe, sogro(a)</li> <li>05. Neto(a), bisneto(a)</li> </ol>

		<p>06. Irmão, irmã</p> <p>07. Nora, genro</p> <p>08. Outro parente</p> <p>09. Agregado</p> <p>10. Pensionista</p> <p>11. Empregada/o doméstica/o</p> <p>12. Parente de empregada/o doméstica/o</p>
Eleições	Você votou nas últimas eleições para prefeito e vereador em 2008?	<p>1. Sim</p> <p>2. Não</p> <p>3. Ainda não tinha tirado o título de eleitor</p>
	Sem citar nomes, você se lembra em quem votou nas últimas eleições em 2008?	<p>1. Sim, lembra todos</p> <p>2. Sim, lembra parcialmente</p> <p>3. Não</p> <p>4. Voto branco ou nulo</p>

---

Fonte: O autor

Sendo o interesse do presente estudo observar os fatores que influenciam a percepção dos jovens quanto à importância da educação para sua formação profissional, a variável resposta utilizada foi “Para o seu futuro profissional”. Para efeito de análise de Regressão Logística, esta variável foi categorizada como “Nada ou pouco importante” e “Muito importante”.

A variável “Quanto tempo, em horas e minutos, leva no percurso de ida à escola?” foram reduzidas a uma variável convertida para minutos denominada “Tempo total gasto para ir a escola”.

As variáveis intituladas “Gostaria que você me dissesse se você participa, se já participou ou se nunca participou de” foram reduzidas a uma única variável denominada “Participação de alguma atividade filantrópica” sendo as opções de resposta “Sim” caso o indivíduo tenha declarado que participa ou já participou ao menos de um grupo de atividades e “Não” caso não participe de nenhum grupo de atividades.

Com respeito às “Características da moradia” as variáveis foram utilizadas para gerar uma nova variável que representa o número de pessoas por cômodo utilizado como dormitório no domicílio.

#### **2.1.5-Registros excluídos do banco de dados e problemas encontrados**

Sendo o interesse do presente estudo indivíduos entre 14 e 24 anos, foram excluídos das análises indivíduos que não informaram a idade e que apresentavam diferentes idades nas seções, totalizando 51 casos.

Foram encontrados indivíduos identificados com diferente sexo em duas seções estudadas, desta forma estes também foram retirados da análise totalizando 16 casos.

Também foi observado, duplicação de alguns indivíduos, identificada analisando os números de identificação presente no banco de dados. Para o presente estudo optou-se por excluir tais registros resultando em 115 exclusões.

A variável “Utilizar suportes de multimídia (CDRom, pen drive, jogos etc.) (excluindo internet)” apresentou descrição incorreta no banco de dados, fato que não interfere nas análises, a descrição foi corrigida segundo a pergunta da mesma presente no questionário e no Quadro 2 acima.

## Capítulo 3

### 3.1-Metodologia

As análises apresentadas nesta monografia foram realizadas a partir do Software SPSS for Windows 14.0 (número de série 9656438).

#### 3.1.1-Regressão logística

Métodos de regressão, de um modo geral, buscam estabelecer relações entre uma variável resposta e variáveis explicativas. O modelo de regressão logística binária, em particular, difere do modelo de regressão linear por apresentar em sua variável resposta apenas valores binários ou dicotômicos, como por exemplo,  $Y = 0$  e  $Y = 1$  onde o valor 1 assume o que pode ser denominado evento de interesse (HOSMER & LEMESHOW, 2000; Estatcamp, 2011).

Em um modelo linear, considerando  $p$  variáveis independentes em um vetor  $X(x_1, x_2, \dots, x_p)$ , temos (HOSMER & LEMESHOW, 2000; Estatcamp, 2011):

( 2 )

$$E(Y|X) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

onde  $E(\varepsilon) = 0$ ,  $Y$  é a variável aleatória que representa o valor da variável resposta,  $x$  é o valor da variável explicativa,  $\beta$  são os parâmetros do modelo que definem a reta de regressão,  $\varepsilon$  é o erro (Estatcamp, 2011).

No caso em que  $Y$  apresenta uma distribuição *Bernoulli* ( $\pi$ ), temos (Estatcamp, 2011):

( 3 )

$$P(Y = 1) = \pi(X)$$

sendo esta a probabilidade de sucesso. Desta forma é possível afirmar que (HOSMER & LEMESHOW, 2000; Estatcamp, 2011):

( 4 )

$$\pi(X) = E(Y)$$

Desta forma, as suposições do modelo linear são violadas: o erro não assume a normalidade, a variância não é homogênea e apresenta restrição na resposta média em que  $0 \leq E(Y) \leq 1$  que é inapropriada para um modelo linear (Estatcamp, 2011).

Existem muitas funções que foram propostas com a finalidade de analisar tais variáveis, sendo que o mais simples é o modelo logístico. A utilização da regressão logística apresenta vantagens do ponto de vista estatístico e interpretações bem simples (HOSMER & LEMESHOW, 2000; Estatcamp, 2011).

Supondo uma amostra de  $n$  observações independentes  $(x_i, y_i)$  e assumindo que  $Y \sim \text{Bernoulli}(\pi(X))$ , temos que (Estatcamp, 2011; HOSMER & LEMESHOW, 2000):

( 5 )

$$P[Y = y_i] = \pi(X)^{y_i}(1 - \pi(X))^{1-y_i}$$

onde  $i = 1, 2, \dots, n$ ,  $X$  é o vetor das variáveis explicativas,  $y_i$  é o número de ocorrência de um evento e  $n$  é o tamanho da amostra (Estatcamp, 2011).

Para que seja realizado um ajuste para a resposta média ao modelo linear é utilizado uma função de ligação (HOSMER & LEMESHOW, 2000; Estatcamp, 2011):

( 6 )

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

Realizando uma transformação para o modelo *logit*, temos (HOSMER & LEMESHOW, 2000):

( 7 )

$$g(X) = \ln \left[ \frac{\pi(X)}{1 - \pi(X)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

onde  $g(X)$  apresenta parâmetros lineares, que podem ser contínuos e podem variar entre  $+\infty$  a  $-\infty$  e  $\beta$  determina a taxa de crescimento ou decréscimo da função *logit* (HOSMER & LEMESHOW, 2000).

Pelo método de estimação por máxima verossimilhança (MV) os parâmetros para  $\beta$  podem ser estimados. Este método busca fornecer valores para os parâmetros para maximizar a probabilidade de se obter determinado conjunto de dados (Estatcamp, 2011).

Assumindo a independência de  $(x_0, y_0), \dots, (x_n, y_n)$ , a função de verossimilhança é dada por (Estatcamp, 2011; HOSMER & LEMESHOW, 2000):

( 8 )

$$L(\beta_0, \beta_1, \dots, \beta_p | (x_i, y_i)) = \sum_{i=1}^n \{y_i g(X) - \ln[1 + g(X)]\}$$

Estimando os parâmetros  $\beta$  é possível ajustar um modelo. Estes estimadores de máxima verossimilhança maximizam o logaritmo da função de verossimilhança que possui um máximo, dado que a função logaritmo é estritamente crescente (Estatcamp, 2011).

Com o objetivo de maximizar a função de verossimilhança devemos derivar em relação aos parâmetros do modelo e igualar o resultado a zero. Assim teremos as equações de Verossimilhança (Estatcamp, 2011):



( 9 )

$$\frac{\partial}{\partial \beta_0} L(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} = 0$$

( 10 )

$$\frac{\partial}{\partial \beta_j} L(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} = 0$$

para  $j = 1, 2, \dots, p$ .

Como estas equações não são lineares nos parâmetros, é necessário métodos numéricos iterativos como, por exemplo, Newton-Raphson e neste caso teremos (Estatcamp, 2011):

( 11 )

$$\beta^{(m+1)} = \beta^{(m)} + [-I(\beta^{(m)})^{-1}]U'(\beta^{(m)})$$

Onde  $m = 0, 1, \dots$ ,  $U'(\beta)$  são as derivadas de segunda ordem do logaritmo da função de verossimilhança em relação aos parâmetros do modelo,  $I(\beta)$  é a matriz de informação de Fisher e  $\beta$  é um vetor  $\beta(\beta_1, \beta_2, \dots, \beta_j)$ ,  $j = 1, 2, \dots, p$  (Estatcamp, 2011).

A interpretação dos parâmetros encontrados no modelo de regressão logística é dada pela comparação da probabilidade de sucesso com a de fracasso utilizando a função *odds ratio* (OR) ou razão de chances. Esta função é dada por (Estatcamp, 2011; HOSMER & LEMESHOW, 2000):

( 12 )

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = \exp(\beta)$$

Na regressão logística a estatística de teste é encontrada pela estatística de Wald, sendo a hipótese nula (HOSMER & LEMESHOW, 2000):

( 13 )

$$H_0: \beta = 0$$

e a estatística de teste é dada por (HOSMER & LEMESHOW, 2000):

$$Z = \frac{\hat{\beta}}{\widehat{EP}}$$

onde  $Z$  apresenta uma distribuição aproximadamente normal padrão quando  $\beta = 0$ ,  $Z^2$  apresenta distribuição aproximadamente Qui-quadrado com 1 G.L. e  $\widehat{EP}$  é o erro padrão de  $\hat{\beta}$  (HOSMER & LEMESHOW, 2000).

O erro padrão da estimativa é uma medida da dispersão dos pontos amostrais da reta de regressão, ou seja, é uma medida das diferenças entre os valores amostrais observados e previstos (TRIOLA, 2008).

Podemos encontrar o erro padrão por (HOSMER & LEMESHOW, 2000)

( 14 )

$$\widehat{EP}(\hat{\beta}) = \sqrt{\widehat{Var}(\hat{\beta})}$$

Porém, os estimadores de máxima verossimilhança apresentam inconsistência quando aplicado a amostras complexas (estratificadas e conglomeradas), como é o caso da PAD-MG. Desta forma, é necessário utilizar o método da máxima pseudo-verossimilhança (MPV) (CHAMBERS & SKINNER, 2003).

### 3.1.2-Regressão logística para amostragem complexa

O método de máxima pseudo-verossimilhança produz estimativas levando em consideração as características do plano amostral. Mesmo que o estimador encontrado na MPV possa ser em algumas situações igual ao de MV, a

estimativa da variância pelo MPV é preferível em relação ao método que não leva em consideração o plano amostral. Este procedimento fornece estimativas mais robustas para a realização de inferência para a população (PESSOA & SILVA, 1998).

As dificuldades na aplicação da MPV é a necessidade de conhecimento das características do plano amostral e o fato de que suas propriedades não são conhecidas para pequenas amostras, o que impossibilita algumas análises estatísticas (PESSOA & SILVA, 1998).

Para situações que os dados são provenientes de amostras complexas, o estimador de MPV é obtido a partir de correções que levam em consideração os pesos amostrais. O estimador de MPV de  $\beta$  no modelo considerado é dado pelas equações descritas em (9) e (10), porém neste caso, é necessária uma ponderação. Assim temos (PESSOA & SILVA, 1998):

( 15 )

$$\sum_{i \in S} w_i \left[ \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \right] = 0$$

( 16 )

$$\sum_{i \in S} w_i \left[ \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \right] = 0$$

onde  $w_i$  é o peso da  $i$ -ésima observação amostral.

Os parâmetros são estimados por métodos numéricos iterativos como já exemplificado na seção anterior.

Para situações em que os dados são coletados por planos amostrais complexos, não existe normalmente uma forma analítica direta para calcularmos o erro

padrão de estimadores como de coeficientes de modelos de regressão logística. Desta forma, são utilizados métodos, tais como a linearização de Taylor (também chamado de método delta) e os métodos de replicação de Jackknife e bootstrap (WOLTER, 1985). Neste trabalho, utilizamos o método da linearização de Taylor. O método da linearização de Taylor vem sendo utilizado com frequência para se obter aproximações de estimadores não-lineares por estimadores lineares. Este método tem como base expansão em series de Taylor do estimador pontual do parâmetro de interesse, em torno do verdadeiro parâmetro e considera apenas as aproximações de primeira ordem.

## Capítulo 4

### 4.1-Aplicações e Resultados

#### 4.1.1-Análise de Regressão Logística Simples

Em um primeiro momento foi realizado uma análise de regressão logística simples (univariada) entre a variável resposta (Para futuro profissional) e cada uma das demais variáveis descritas na seção 2.1.4, de forma que se definiu que as variáveis que apresentassem significância estatística ao nível de 10% seriam utilizadas no modelo de regressão multivariado. Esta análise foi realizada considerando-se o plano amostral e depois sem considera-lo, para que fosse possível observar possíveis diferenças nos resultados.

Os resultados das análises e seus respectivos valores p, estão descritos na Tabela 1:

Tabela 1 - Análise de regressão logística considerando e não considerando o plano amostral

Variável	Valor p considerando desenho amostral	Valor p sem considerar desenho amostral
Ouvir música (CD, DVD, MP3 etc.)	0,000*	0,000*
Ver televisão	0,013*	0,005*
Ouvir rádio	0,327	0,000*
Ler jornais e/ou revistas	0,001*	0,000*
Ler livros (excluindo didáticos e profissionais)	0,000*	0,000*
Praticar esportes	0,009*	0,000*
Dançar	0,697	0,323
Utilizar suportes de multimídia (CDRom, pen drive, jogos etc.) (excluindo internet)	0,019*	0,003*
Utilizar internet	0,000*	0,000*
Receber / visitar amigos ou parentes	0,673	0,885
Passear em espaços ao ar livre	0,222	0,013*
Ir ao shopping	0,121	0,186
Ir a discotecas / bailes e ou bares / restaurantes	0,555	0,041*
Assistir eventos esportivos em campos ou	0,165	0,271

quadras		
Ir ao cinema	0,013*	0,000*
Frequentar associações recreativas ou clubes	0,009*	0,004*
Ir à missa ou a cerimônias religiosas	0,002*	0,000*
Ver exposições	0,086*	0,011*
Ir a bibliotecas	0,001*	0,000*
Ir a museus	0,283	0,099*
Ir ao teatro	0,253	0,015*
Ir a concertos de música erudita / clássica	0,209	0,149
Ir a shows de música popular / moderna	0,276	0,764
Ir a espetáculos de dança	0,063*	0,211
Ver espetáculos de rua	0,553	0,878
Ir a festas populares	0,021*	0,024*
Excetuandose programas sociais que transferem dinheiro, você participa ou é beneficiário de algum programa ou projeto social?	0,270	0,063*
Quantos anos tinha quando iniciou a primeira série do ensino fundamental (primeiro ano primário ou 1º grau)?		
Alguma vez você já parou de estudar e depois voltou?	0,000*	0,000*

Com qual idade deixou de frequentar escola?	0,000*	0,000*
Qual foi o curso mais elevado que frequentou no ano passado?	0,000*	0,000*
Qual é o curso que frequenta atualmente?	0,028*	0,104
Sabe ler e escrever um bilhete simples no idioma que conhece?	0,000*	0,000*
Na escola em que estuda, funciona algum programa social do governo estadual para os alunos ou para as famílias dos alunos do governo estadual?	0,519	0,514
A escola em que estuda oferece merenda gratuita?	0,221	0,262
Qual o principal motivo pelo qual deixou de ir à escola, pelo menos um dia, nos últimos 30 dias?	0,000*	0,001*
Quem acompanha as atividades escolares?	0,756	0,506
Por que não frequenta escola?	0,000*	0,000*
Curso de educação profissional permitiu melhorar sua posição no seu trabalho?	0,194	0,022*
Estuda neste município?	0,000*	0,000*
Teve algum gasto com educação com em 2009, como mensalidades escolares,	0,000*	0,000*



transporte para ir à escola, material escolar, uniformes, etc.		
Qual é o curso ou grau mais elevado que frequentou, no qual concluiu pelo menos uma série? (Grau mais elevado na moradia)	0,000*	0,000*
Qual o principal meio de transporte habitualmente utilizado por para ir até a creche ou escola?	0,000*	0,414
Tempo total gasto para ir a escola	0,107	0,122
A palavra que mais se aproxima do que você pensa sobre trabalho é:	0,000*	0,000*
A palavra que fica em 2º lugar do que você pensa sobre trabalho é:	0,001*	0,000*
Vou ler duas afirmações. Após ouvi-las, por favor, me diga com qual dessas afirmações você mais concorda, se com a primeira ou com a segunda.	0,000*	0,000*
Participação de alguma atividade filantrópica	0,002*	0,000*
Pessoas por cômodo servindo de dormitório	0,000*	0,000*
Tem algum problema de saúde que exige acompanhamento constante?	0,000*	0,649
Qual é o estado civil atualmente?	0,000*	0,000*

Cor ou raça	0,008*	0,016*
Sexo	0,000*	0,000*
Qual é a idade em anos completos ou a idade presumida?	0,000*	0,000*
Qual a relação de convivência que tem com o responsável pelo domicílio?	1,000	0,000*
Até 31 de maio de 2009, teve quantos filhos nascidos vivos (ou seja, que apresentaram algum sinal de vida ao nascer)?	0,000*	0,000*
Você votou nas últimas eleições para prefeito e vereador em 2008?	0,156	0,082*
Sem citar nomes, você se lembra em quem votou nas últimas eleições em 2008?	0,001*	0,000*

---

\*valor  $p \leq 0,1$

Fonte: O Autor

Com base nos resultados obtidos acima, as variáveis com significância estatística foram selecionadas para o ajuste em nos modelos de regressão logística. Primeiramente foi criado um modelo considerando as variáveis significantes (Tabela 1) com o peso amostral e um com as variáveis significantes que não considerava o peso amostral, depois foram criados dois modelos de regressão onde somente entraram as variáveis significantes quando considerado o peso amostral (Tabela 1).

Foram excluídas da análise as variáveis que apresentavam mais de 30% de dados faltantes como descritas no Quadro 3.

**Quadro 3 - Variáveis retiradas por apresentarem frequência alta de dados faltantes**

Variável	% de dados faltantes
Você votou nas últimas eleições para prefeito e vereador em 2008?	40,6
Sem citar nomes, você se lembra em quem votou nas últimas eleições em 2008?	60,6
Na escola em que estuda, funciona algum programa social do governo estadual para os alunos ou para as famílias dos alunos do governo estadual?	71,7
A escola em que estuda oferece merenda gratuita?	47,8
Qual o principal meio de transporte habitualmente utilizado por para ir até a creche ou escola?	47,1
Qual o principal motivo pelo qual deixou de ir à escola, pelo menos um dia, nos últimos 30 dias?	83,1

Quantos anos tinha quando iniciou a primeira série do ensino fundamental (primeiro ano primário ou 1º grau)?	48,6
Quem acompanha as atividades escolares?	57,0
Por que não frequenta escola?	53,3
Com qual idade deixou de frequentar escola?	54,7
Qual foi o curso mais elevado que frequentou no ano passado?	53,6
Qual é o curso que frequenta atualmente?	47,3
Curso melhorou posição no trabalho	94,7
Qual é o curso ou grau mais elevado que frequentou, no qual concluiu pelo menos uma série? (Grau mais elevado na moradia)	32,9
Tempo total gasto para ir a escola	47,2

Fonte: O Autor

#### 4.1.2-Primeiro modelo de regressão

A partir das variáveis que apresentaram significância estatística e que não foram excluídas no critério de dados faltantes maiores do que 30%, foi ajustado um modelo de regressão logística, onde primeiramente se levou em consideração o

desenho amostral e depois um modelo onde não se leva em consideração o desenho amostral. No modelo onde se considera o plano amostral, foram incluídas as variáveis significantes na Tabela 1 das análises onde também se considerava o plano amostral, o mesmo ocorreu com o modelo onde não se considera o plano amostral. O modelo foi encontrado do seguinte modo: todas as variáveis foram inicialmente incluídas e foram, uma a uma, retiradas manualmente segundo sua significância. Para esta análise o nível de significância considerado foi de 5% e o resultado final pode ser observado na Tabela 2:

Tabela 2 - Modelo final considerando e não considerando o plano amostral

Variáveis	Categorias	Com plano amostral				Sem plano amostral			
		Valor p	EP	Exp(B)	IC	Valor p	EP	Exp(B)	IC
Alguma vez você já parou de estudar e depois voltou?	Sim					0,000	0,074	0,618	0,535- 0,715
	Não						.	1	.
Estuda neste município?	Sim		0,099	0,455	0,375- 0,552				
	Não	0,000	0,301	0,503	0,279- 0,907				
	Não Estuda		.	1	.				
A palavra que mais se aproxima do que você pensa sobre trabalho é:	Necessidade		0,327	0,617	0,325- 1,172	0,000	0,262	1,682	1,007- 2,808
	Auto realização		0,345	0,387	0,197- 0,762		0,278	3,039	1,763- 5,239

	Independência	0,331	0,469	0,245- 0,897		0,262	2,345	1,404- 3,916
	Crescimento	0,329	0,353	0,186- 0,673		0,264	2,934	1,750- 4,918
	Exploração	.	1	.		.	1	.
	Necessidade	0,253	0,704	0,428- 1,156		,209	1,297	0,861- 1,995
	Auto realização	0,262	0,605	0,362- 1,012		,217	1,975	1,292- 3,019
A palavra que fica em 2º lugar do que você pensa sobre trabalho é:	Independência	0,024	0,255	0,346- 0,941	0,000	,208	1,761	1,171- 2,650
	Crescimento	0,255	0,495	0,300- 0,817		,211	2,020	1,336- 3,055
	Exploração	.	1	.		.	1	.
Sexo	Masculino	0,002	0,094	1,121- 1,620	0,000	0,071	0,767	0,667- 0,881

	Feminino	.	1	.		.	1	.
Ir a espetáculos de dança		0,033	0,005	1,011	1,001-			
					1,021			
Pessoas por cômodo servindo como dormitório		0,001	0,053	1,186	1,068-	0,000	0,040	0,865
					1,317			0,800-
								0,934
Ouvir rádio						0,006	0,011	1,032
								1,009-
								1,056
Ler livros (excluindo didáticos e profissionais)						0,000	0,018	1,077
								1,040-
								1,116
Passear em espaços ao ar livre						0,014	0,004	1,009
								1,002-
								1,016
Sabe ler e escrever um bilhete simples no idioma que conhece?	Sim					0,000	0,264	3,230
								1,925-
	Não						.	1
								5,421
Vou ler duas afirmações. Após ouvi-las, por favor, me diga com qual dessas afirmações você mais concorda, se com a primeira ou com a	Devemos priorizar o crescimento da economia e a					0,000	0,105	0,622
								0,507-
								0,764



segunda.	criação de empregos Devemos proteger o meio ambiente	. 1 .
----------	--	-------

Fonte: O Autor

Legenda: EP-Erro Padrão, Exp(B)-Razão de Chance, IC-Intervalo de Confiança

#### *4.1.2.1-Considerando o Desenho Amostral na análise*

Quando levamos em consideração o desenho amostral temos que a variável “Estuda neste município?” é considerada significativa de modo que, quem estuda no mesmo município ou em outro, diminui a chance de considerar a educação muito importante em relação aqueles que não estudam.

Em relação a variável “A palavra que mais se aproxima do que você pensa sobre trabalho é:” e “A palavra que fica em 2º lugar do que você pensa sobre trabalho é:” quando observamos o modelo que considerou o plano amostral, temos que as categorias “Auto realização”, “Independência” e “Crescimento” diminuem as chances de considerarem a educação muito importante para o futuro profissional em relação aos que responderam “Exploração”.

Observando a variável “Sexo” temos que esta se mostra significativa sendo que a chance de homens considerarem a educação como muito importante aumenta em 34,8% em relação às mulheres.

Com relação a variável “Ir a espetáculos de dança” as chances aumentam em 1,1% de considerarem a educação como muito importante de acordo com as frequências a espetáculos de dança.

Quando observamos a variável “Pessoas por cômodo servindo como dormitório” temos que, de acordo com o numero de pessoas por cômodo a chance aumenta em 18,6% de considerarem a educação como muito importante para o futuro profissional.

#### ***4.1.2.2-Não considerando o Desenho Amostral na análise***

Quanto ao modelo que não leva em consideração o desenho amostral, temos que a variável “Alguma vez você já parou de estudar e depois voltou?” apresenta uma estimativa de redução de quem relatou ter parado e retornado aos estudos na chance de considerar a educação importante.

Quanto a variável “A palavra que mais se aproxima do que você pensa sobre trabalho é:” e “A palavra que fica em 2º lugar do que você pensa sobre trabalho é:” as categorias “Necessidade”, “Auto realização”, “Independência” e “Crescimento” aumentam as chances em relação aos que responderam “Exploração” em considerar a educação como importante para o futuro profissional.

Observando a variável “Sexo” temos que a chance de ser homem e considerar a educação importante diminui em relação a ser mulher.

Quando observamos a variável “Pessoas por cômodo servindo de dormitório” quanto maior o numero de pessoas vemos que as chances de considerarem a educação importante diminuem.

As variáveis “Ouvir rádio”, “Ler livros (excluindo didáticos e profissionais)” e “Passear em espaços ao ar livre” somente se encontram no modelo que não consideram o plano amostral. Assim como a variável “Sabe ler e escrever um bilhete simples no idioma que conhece?” e “Vou ler duas afirmações. Após ouvi-las, por favor, me diga com qual dessas afirmações você mais concorda, se com a primeira ou com a segunda.”.

#### **4.1.2.3-Outras observações**

Outra característica interessante a observar é que os erros padrão sempre se apresentam menores no modelo onde não se leva em consideração o plano amostral, o que é um possível efeito da conglomeração na amostra da PAD-MG e está de acordo com a literatura sobre Amostragem (CHAMBERS & SKINNER, 2003).

#### **4.1.3-Segundo modelo de regressão**

A partir das variáveis que apresentaram significância estatística e que não foram excluídas no critério de dados faltantes maiores do que 30%, foi ajustado um modelo de regressão logística sem considerar o plano amostral com as variáveis estatisticamente significantes do modelo de regressão logística simples com plano amostral (Tabela 1):

Tabela 3 - Análise de Regressão Logística com e sem o plano amostral com as mesmas variáveis

Variáveis	Categorias	Com plano amostral				Sem plano amostral				
		Valor p	EP	Exp(B)	IC	Valor p	EP	Exp(B)	IC	
Estuda neste município?	Sim	0,000	0,099	0,455	0,375- 0,552					
	Não		0,301	0,503	0,279- 0,907					
	Não Estuda		.	1	.					
A palavra que mais se aproxima do que você pensa sobre trabalho é:	Necessidade	0,000	0,327	0,617	0,325- 1,172	0,000		0,257	1,612	0,974- 2,667
	Auto realização		0,345	0,387	0,197- 0,762					1,771- 5,167
	Independência		0,331	0,469	0,245- 0,897					1,361- 3,725
	Crescimento		0,329	0,353	0,186- 0,897					1,746- 3,725

				0,673			4,813	
	Exploração	.	1	.		1	.	
A palavra que fica em 2º lugar do que você pensa sobre trabalho é:	Necessidade	0,253	0,704	0,428- 1,156		,208	1,345 2,021	
	Auto realização	0,262	0,605	0,362- 1,012		,215	2,041 3,108	
	Independência	0,024	0,255	0,346- 0,941	0,000	,207	1,832 2,749	
	Crescimento		0,255	0,495	0,300- 0,817		,209	2,127 3,205
	Exploração		.	1	.		1	.
		Masculino	0,002	0,094	1,348	1,121- 1,620	0,069	0,716 0,821
Sexo	Feminino		.	1	.	.	1	
Ir a espetáculos de dança		0,033	0,005	1,011	1,001- 1,021			

Pessoas por cômodo servindo como dormitório	0,001	0,053	1,186	1,068- 1,317	0,000	0,039	0,845	0,782- 0,913
Ouvir música (CD, DVD, MP3 etc.)					0,009	0,013	1,034	1,008- 1,061
Ir a Bibliotecas					0,001	0,001	1,003	1,001- 1,005
Vou ler duas afirmações. Após ouvi-las, por favor, me diga com qual dessas afirmações você mais concorda, se com a primeira ou com a segunda.				Devemos priorizar o crescimento da economia e a criação de empregos Devemos proteger o meio ambiente	0,000	0,103	0,608	0,496- 0,744
						.	1	.

Fonte: O Autor

Legenda: EP-Erro Padrão, Exp(B)-Razão de Chance, IC-Intervalo de Confiança





Quando realizamos o modelo de regressão logístico binário sem considerar o plano amostral, neste caso comparando com o mesmo modelo considerando o plano amostral, temos que variáveis como “Ouvir música (CD, DVD, MP3 etc.)”, “Ir a Bibliotecas” e “Vou ler duas afirmações. Após ouvi-las, por favor, me diga com qual dessas afirmações você mais concorda, se com a primeira ou com a segunda” apresentaram significância estatística de modo que entrariam no modelo final.

Também observamos que as variáveis “A palavra que mais se aproxima do que você pensa sobre trabalho é:” e “A palavra que fica em 2º lugar do que você pensa sobre trabalho é:” apresentam comportamento distinto em cada modelo, ou seja, no modelo onde se considera o plano amostral, suas categorias mostram uma redução na razão de chance de considerar a educação importante para o futuro profissional em relação aos que responderam “Exploração”, e quando observamos o modelo que não leva em consideração o plano amostral, vemos que as mesmas variáveis aumentam a razão de chance. Comportamento similar ocorre quanto a variável “Sexo”, porém, no modelo onde se considera o plano amostral ser do sexo masculino aumenta a razão de chance em relação a ser do sexo feminino. O que também é observado em relação a variável “Pessoas por cômodo servindo como dormitório”.

Também observamos que variáveis como “Ir a espetáculos de dança” e “Estuda neste município?” somente estão presentes no modelo onde se considera o plano amostral.

## Capítulo 5

### 5.1-Conclusões

Foi possível observar durante o processo de execução desta monografia que a pesquisa e o banco de dados utilizado apresentaram muitos erros e inconsistência, o que, de certa forma prejudicou a análise com as variáveis selecionadas. Assim os resultados apresentados nesta monografia, devem ser vistos apenas como um exercício da aplicação de métodos estatísticos adequados para análise e variáveis que estão relacionadas com a percepção dos jovens mineiros quanto à importância da educação e o modelo final ajustado não deve ser considerado como o melhor modelo possível.

Pelos resultados foi possível observar que o desenho amostral causa um efeito nos resultados da análise. Em geral, variáveis que não apresentaram significância estatística no modelo que leva em consideração o plano amostral, apresentaram significância estatística quando o mesmo não foi levado em consideração.

Quando analisamos as mesmas variáveis em dois modelos – Considerando o plano amostral e não considerando – observamos a existência de dois grupos de diferentes variáveis onde que, uma análise feita de forma incorreta, acarretaria em um modelo distinto do modelo correto o que resultaria em uma análise equivocada e na tomada de decisões erradas.

Em situações em que temos o plano amostral complexo, os dados apresentam efeitos de conglomeração e estratificação que interferem diretamente na estimação pontual, na estimação dos erros padrão e conseqüentemente nos testes

estatísticos e nos intervalos de confiança (PESSOA & SILVA, 1998), o que pode ser ilustrado a partir dos resultados desta monografia.

## Bibliografia

CHAMBERS, R. L., & SKINNER, C. J. (2003). *Analysis of Survey Data*. Chichester: Wiley.

Estatcamp. (2011). *Estatcamp*. Acesso em 13 de Novembro de 2011, disponível em Portal Action: <http://portalaction.com.br/687-3-regressão-logística>

Fundação João Pinheiro. (2009). *Relatório Metodológico da Pesquisa por Amostragem de Minas Gerais 2009*. Belo Horizonte: Fundação João Pinheiro.

Fundação João Pinheiro. (04 de Agosto de 2010). *Fundação João Pinheiro - FJP*. Acesso em 01 de 04 de 2011, disponível em [www.fjp.gov.br](http://www.fjp.gov.br)

HOSMER, D. W., & LEMESHOW, S. (2000). *Applied Logistic Regression* (2<sup>a</sup> ed.). New York: John Wiley e Sons.

JOHNSON, R. A., & WICHERN, D. W. (1988). *Applied Multivariate Statistical Analysis* (4 ed.). New Jersey: Prentice Hall.

LEHTONEN, R., & PAHKINEN, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys* (2<sup>o</sup> ed.). Chichester: TLF<sub>e</sub>BOOK.

NERI, M., MELO, L. C., MONTE, S. d., NERI, A., PONTES, C., ANDARI, A. U., et al. (2009). *Motivos da Evasão Escolar*. Rio de Janeiro: FGV/IBRE,CPS.

PESSOA, D. G., & SILVA, P. d. (1998). *Análise de Dados Amostrais Complexos*.

SILVA, P. d., & VIEIRA, M. d. (2010). *Dados e Métodos Usados para o Cálculo dos Pesos de Domicílios e Moradores da PAD-MG*. Belo Horizonte: Fundação João Pinheiro.

SUYAMA, E. (2008). *Plano Amostral para a Pesquisa por Amostragem de Domicílios em Minas Gerais - PAD-MG*. Belo Horizonte: Fundação João Pinheiro.

TRIOLA, M. F. (2008). *Introdução à Estatística* (10<sup>a</sup> ed.). Rio de Janeiro: LTC.

WOLTER, K. M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.