



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Adriano Luiz Antunes

Denis Antoniazzi

**INFLUÊNCIA DE VARIÁVEIS SOCIODEMOGRÁFICAS SOBRE O
EMPREENDEDORISMO NO BRASIL**

CURITIBA

2014

Adriano Luiz Antunes
Denis Antoniazzi

**INFLUÊNCIA DE VARIÁVEIS SOCIODEMOGRÁFICAS SOBRE O
EMPREENDEDORISMO NO BRASIL**

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório de Estatística do Curso de Graduação em Estatística da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Profa. Dra. Suely Ruiz Giolo

CURITIBA
2014

***Aos nossos pais, que sempre nos deram
apoio e incentivo para que nossos
sonhos se tornassem realidade.***

A vocês nossa eterna gratidão!

AGRADECIMENTOS

A Deus, por nos manter sempre fortes e confiantes, mesmo nos momentos difíceis.

As nossas amadas, Priscila Del Pilar Arriagada Gajardo e Larissa Andressa Souza, pela compreensão de nossa ausência e que, com carinho, sempre estiveram do nosso lado e de forma especial nos deram todo o apoio durante todo o curso, assim como pelas inúmeras conversas e questões que certamente ajudaram a melhorar este trabalho.

À Professora Doutora Suely Ruiz Giolo, nossos sinceros agradecimentos pelas reuniões, pelo incentivo na construção dos modelos e revisões, por todas as discussões, pelo inestimável ensino e por ter feito o melhor possível para nos transmitir sua experiência.

À Professora Doutora Nívea da Silva Matuda, pela disponibilidade de participar da banca deste trabalho e pelo valioso trabalho que tem realizado frente à Coordenação do Curso de Estatística, em conjunto com o Professor Doutor César Augusto Taconelli.

Aos nossos amigos de curso, pelo companheirismo durante todo curso.

Ao Instituto Brasileiro de Qualidade e Produtividade (IBQP), em especial à Simara Maria de Souza Silveira Greco, coordenadora nacional do projeto GEM, por ter possibilitado o acesso aos dados do GEM nacional utilizados neste trabalho.

“É hora de todos nós ficarmos em pé e aplaudirmos
as pessoas empreendedoras, o realizador –
aquele que reconhece o desafio e faz
alguma coisa em relação à ele”.

Vince Lombardi

RESUMO

Influência de Variáveis Sociodemográficas Sobre o Empreendedorismo no Brasil

Este trabalho apresenta um estudo sobre o empreendedorismo no Brasil. Os dados analisados referem-se ao ano de 2012 e são oriundos da pesquisa *Global Entrepreneurship Monitor* (GEM) realizada pelo Instituto Brasileiro de Qualidade e Produtividade (IBQP). Para a análise dos dados foi, primeiramente, ajustado um modelo de logitos generalizados a fim de estudar a influência de diversas variáveis sociodemográficas nos estágios do empreendedorismo (não empreendedores, empreendedores iniciais e empreendedores estabelecidos) dos brasileiros. Em seguida, foi ajustado um modelo para dados binários (modelo de regressão logística), considerando somente empreendedores iniciais, com a finalidade de estudar a motivação destes para empreender (oportunidade ou necessidade) segundo as mesmas variáveis sociodemográficas. Ambos os modelos ajustados mostraram que as variáveis região geográfica, gênero, faixa etária, renda salarial e escolaridade dos brasileiros estão associadas tanto ao estágio do empreendedorismo quanto à motivação de empreender. Alguns dos resultados mostraram que as mulheres têm apresentado maior propensão para iniciar um novo negócio motivadas, na maioria das vezes, pela necessidade, ao contrário dos homens que são mais propensos a iniciar um novo negócio por oportunidade. Em geral, a maior chance de iniciar um novo empreendimento foi observada entre os mais jovens (18-33 anos) com escolaridade acima do segundo grau. Ainda, predominância de empreendedores estabelecidos foi observada entre os homens na faixa etária de 34 a 48 anos. Destaque quanto ao empreendedorismo pôde também ser observado em particular para a região Centro-Oeste do Brasil que, em relação às demais regiões, apresentou percentuais mais elevados de empreendedores iniciais e estabelecidos.

Palavras-chave: Dados binários. Dados categóricos. Empreendedores. Modelo de logitos generalizados. Motivação para empreender. Regressão logística.

LISTA DE FIGURAS

Figura 1 – Exemplo gráfico das diferenças entre os valores observados e valores preditos pelo modelo de logitos generalizados	11
Figura 2 – Percentuais da população adulta segundo o estágio do empreendedorismo e a motivação observados nas regiões brasileiras.....	16
Figura 3 – Percentuais da população adulta segundo o estágio do empreendedorismo.....	17
Figura 4 – Percentuais dos empreendedores iniciais segundo a motivação por gênero.....	17
Figura 5 – Dispersão da população adulta segundo o estágio por idade.....	18
Figura 6 – Dispersão dos empreendedores iniciais segundo a motivação por idade.....	19
Figura 7 – Percentuais da população adulta segundo o estágio de empreendedorismo por renda familiar em salários mínimos.....	19
Figura 8 – Percentuais dos empreendedores iniciais segundo a motivação por renda familiar em salários mínimos.....	20
Figura 9 – Percentuais da população adulta segundo o estágio por escolaridade.....	20
Figura 10 – Percentuais dos empreendedores iniciais segundo a motivação por escolaridade.....	21

Figura 11 – Diferenças entre os valores observados e os valores preditos pelo modelo de logitos generalizados para cada estágio do empreendedorismo.....	23
Figura 12 – Distribuição dos resíduos brutos associados ao modelo final e representados para cada logito estudado.....	24
Figura 13 – Distribuição dos resíduos de Pearson associados ao modelo final e representados para cada logito estudado	24
Figura 14 – Diferenças entre os valores observados e os valores preditos pelo modelo de regressão logística.....	31
Figura 15 – Distribuição dos resíduos de Pearson e <i>deviance</i> associados ao modelo final.....	32
Figura 16 – Curva ROC associada ao modelo de regressão logística ajustado....	33

LISTA DE QUADROS E TABELAS

Quadro 1 – Descrição das variáveis pesquisadas no GEM 2012 – Brasil	06
Tabela 1 – Representação das probabilidades associadas a um estudo ilustrativo com duas variáveis explicativas X_1 e X_2 categóricas e uma variável resposta politômica.....	09
Tabela 2 – Resultados de alguns dos modelos ajustados aos dados de empreendedorismo.....	22
Tabela 3 – Estimativas, por logito, dos parâmetros do modelo final ajustado.....	22
Tabela 4 – <i>Odds ratios</i> entre empreendedores iniciais e não empreendedores....	25
Tabela 5 – <i>Odds ratios</i> entre empreendedores estabelecidos e não empreendedores.....	27
Tabela 6 – <i>Odds ratios</i> entre empreendedores iniciais e estabelecidos	29
Tabela 7 – Resultados de alguns dos modelos de regressão logística ajustados aos dados.....	30
Tabela 8 – Estimativas dos parâmetros do modelo de regressão logística final ajustado	31
Tabela 9 – <i>Odds ratios</i> entre empreendedores iniciais motivados por oportunidade e os motivados por necessidade.....	34

SUMÁRIO

1 INTRODUÇÃO	01
2 REVISÃO DE LITERATURA	03
3 MATERIAL E MÉTODOS	05
3.1 MATERIAL	05
3.1.1 Conjunto de Dados	05
3.1.2 Recursos Computacionais	07
3.2 MÉTODOS	07
3.2.1 Modelos de Logitos Generalizado	08
3.2.1.1 Estimação dos parâmetros para o modelo de logitos generalizados.....	10
3.2.1.2 Avaliação da qualidade do ajuste.....	10
3.2.1.3 Razão de chances.....	12
3.2.2 Modelos para Dados com Resposta Binária.....	13
3.2.2.1 Estimação dos parâmetros para o modelo com resposta binária.....	14
4 RESULTADOS E DISCUSSÃO	16
4.1 ANÁLISES DESCRITIVAS.....	16
4.1.1 Empreendedorismo por regiões geográficas.....	16
4.1.2 Empreendedorismo por gênero.....	17
4.1.3 Empreendedorismo por idade.....	18
4.1.4 Empreendedorismo por renda familiar.....	19
4.1.5 Empreendedorismo por escolaridade.....	20
4.2 RESULTADOS DO MODELO DE LOGITOS GENERALIZADOS.....	21
4.3 RESULTADOS DO MODELO PARA DADOS BINÁRIOS.....	30
5 CONCLUSÃO	35
REFERÊNCIAS	38
APÊNDICES	41

1 INTRODUÇÃO

Empreendedorismo é um dos temas que está sendo discutido atualmente de forma mais intensa por organizações públicas e privadas. Uma das principais razões abordadas para difundir o pensamento empreendedor é que a atividade empreendedora influencia diretamente no desenvolvimento econômico do país. Segundo o economista francês Jean-Batiste Say, é a atividade que transfere recursos econômicos de um setor de produtividade mais baixo para um setor mais elevado e de maior rendimento (BRITO; WEBER, 2003). Nesse mesmo sentido, Joseph Schumpeter enfatiza que o empreendedorismo é a máquina propulsora do desenvolvimento da economia (SCHUMPETER, 1934).

No Brasil, o empreendedorismo ganhou popularidade no final da década de 90 e tem recebido grande apoio de organizações públicas e privadas. A preocupação com a criação e necessidade da durabilidade das empresas são os principais motivos pela busca de informações (DORNELAS, 2008).

Desde o ano de 1999, o empreendedorismo está sendo monitorado pelo projeto *Global Monitor Entrepreneurship* (GEM). Este projeto é conduzido em âmbito internacional pelo *Global Entrepreneurship Research Association* (GERA)¹, entidade formada pela *London Business School*², *Babson College*³, *Universidad Del Desarrollo*⁴ e representantes dos países participantes da pesquisa. No Brasil, desde 2000, é conduzida pelo Instituto Brasileiro da Qualidade e Produtividade (IBQP) com o apoio do Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE). O GEM tem por objetivo apresentar informações do empreendimento e do perfil do empreendedor, monitorando a dinâmica do empreendedorismo no mundo.

Os resultados do GEM incluem comparações globais, relatórios nacionais e tópicos especiais baseado no ciclo de coleta anual, bem como a presença ativa de mais de 300 acadêmicos e pesquisadores. De acordo com o GEM, o empreendedorismo é qualquer tentativa de criação de um novo negócio ou novo

¹ *Global Entrepreneurship Research Association* (GERA) – Organização sem fins lucrativos geridos por representantes das equipes nacionais, das duas instituições fundadoras do GEM e de instituições patrocinadoras da iniciativa.

² *London Business School* (LBS) é uma escola de negócios e uma faculdade integrante da Universidade de Londres, localizada no centro de Londres, Inglaterra, Reino Unido.

³ *Babson College* é uma escola de negócios privada localizada em Wellesley, Massachusetts.

⁴ *Universidad del Desarrollo* está localizada no Chile tendo sede em *Concepción* e dois campi em Santiago.

empreendimento como, por exemplo, uma atividade autônoma, uma nova empresa, ou a expansão de um empreendimento existente por um indivíduo, grupos de indivíduos ou por empresas já estabelecidas (GEM, 2012).

Alguns dos principais resultados apresentados pelo GEM são: o estágio do empreendimento e a motivação que levou o empreendedor a iniciar um novo negócio; a primeira tem o intuito de verificar o perfil e a dinâmica entre os tipos de empreendedores e a segunda procura entender a motivação que está por trás do empreendedorismo, seja em decorrência da necessidade, seja da oportunidade. Percebe-se que nos países com maior desenvolvimento econômico, a razão entre oportunidade e necessidade é superior à dos demais.

O objetivo principal deste trabalho é o de estudar a associação do estágio do empreendedorismo e da motivação dos empreendedores iniciais segundo algumas características sociodemográficas. Os dados utilizados referem-se à pesquisa realizada no ano de 2012 e foram fornecidos pelo Instituto Brasileiro de Qualidade e Produtividade (IBQP), localizado em Curitiba, Paraná.

Este estudo está estruturado da seguinte forma: informações sobre empreendedorismo e estudos relacionados são encontradas no Capítulo 2; no Capítulo 3 são apresentadas as metodologias utilizadas na análise dos dados e, no Capítulo 4, têm-se as análises descritivas dos dados, bem como comentários sobre a distribuição das características sociodemográficas em relação ao estágio do empreendedorismo e a motivação dos empreendedores iniciais. Neste capítulo, ainda é ajustado o modelo de logitos generalizados com o intuito de verificar as influências das características sociodemográficas no estágio do empreendedorismo, assim como o modelo de regressão logística para verificar a influência das mesmas características sobre a motivação de empreender. Conclusões e discussões são apresentadas no Capítulo 5.

2 REVISÃO DE LITERATURA

Empreendedorismo é um dos temas que têm sido foco de vários estudos acadêmicos e, nos últimos anos, ganhou um destaque imensurável por ser, segundo estudiosos, um dos principais pilares para o desenvolvimento econômico dos países (THURIK et al., 2008, VAN STEL; CARREE; THURIK, 2005).

Filion (1999) afirma que aqueles que pesquisam sobre empreendedorismo concordam em dizer que a origem do termo está nas obras de Richard Cantillon. No século XVIII, Cantillon definiu os empreendedores como indivíduos que compravam a matéria prima por um preço certo e a vendiam a terceiros por um preço incerto depois de processá-las, pois identificavam uma oportunidade de negócio e assumiam o risco.

Jean-Baptiste Say, em 1803, considerava o desenvolvimento econômico como resultado da criação de novos empreendimentos; associou os empreendedores à inovação e identificou-os como agentes da mudança (SAY, 1971). Em 1934, Schumpeter argumentou que os empreendedores são a força motriz do crescimento econômico, ao introduzir no mercado inovações que tornam obsoletos os produtos e as tecnologias existentes.

O empreendedorismo é hoje um fenômeno global e a riqueza gerada pelos empreendedores contribui para a melhoria da qualidade de vida de toda a população e, em muitos casos, percebe-se que há um investimento maior em novos empreendimentos que beneficiam suas próprias comunidades (BRITO; WEVER, 2003). De acordo com a ENDEAVOR (2013), o Brasil está entre os países com maior proporção de pessoas que aspiram constituir um negócio próprio.

Hart e O'Reilly (2006) em um estudo realizado sobre o empreendedorismo utilizando dados do GEM em seus respectivos países de origem, Reino Unido e Irlanda do Norte, apresentaram, quanto à atividade dos empreendedores no Reino Unido, que um indivíduo na faixa etária entre 35-44 anos tem chance 1,4 vezes de ser empreendedor inicial quando comparado à faixa etária 45-54 anos. Já na Irlanda do Norte, esta chance foi de 1,16.

Com o intuito de descrever e analisar os perfis de empreendedores homens e mulheres, Gimenez et al. (2000) conduziram no Norte do Paraná uma pesquisa cujos resultados indicaram uma grande semelhança no nível de atividade

empreendedora entre os gêneros. Homens e mulheres parecem seguir um caminho semelhante no processo do empreendedorismo.

Bhola et al. (2006) estudaram os empreendedores motivados por oportunidade e necessidade segundo diversas características, entre elas, o gênero e o nível de escolaridade. Utilizaram dados da pesquisa *Flash Eurobarometer on Entrepreneurship*⁵ aplicada em 25 países do continente europeu e também nos Estados Unidos. De acordo com os resultados obtidos, os homens apresentaram maior propensão do que as mulheres para iniciar um novo negócio quando a motivação é a oportunidade. Além disso, pessoas com baixo nível de escolaridade apresentaram tendência a se lançar em um novo empreendimento motivadas pela necessidade.

Com o objetivo de contribuir para este tema, este estudo visa identificar características sociodemográficas (variáveis explicativas) que influenciam o empreendedorismo no Brasil e delinear semelhanças com os outros estudos descritos nesta seção.

⁵ *Eurobarometer* é uma série de pesquisas de opinião pública realizada regularmente em nome da Comissão Europeia desde 1973. Essas pesquisas abordam uma grande variedade de temas da atualidade relacionados com a União Europeia.

3 MATERIAL E MÉTODOS

3.1 MATERIAL

Para o desenvolvimento do presente trabalho foi feito uso de um banco de dados fornecido pelo Instituto Brasileiro de Qualidade e Produtividade (IBQP), bem como de recursos computacionais para sua análise. Ambos são descritos a seguir.

3.1.1 Conjunto de Dados

O projeto de pesquisa de âmbito mundial *Global Entrepreneurship Monitor* (GEM) analisa os níveis de empreendedorismo nos diversos países em que atua. Iniciado em 1999, como uma parceria entre a *London Business School* e o *Babson College*, o primeiro estudo abrangeu 10 países e, desde então, cerca de 100 países já participaram do GEM. Os objetivos principais deste projeto são:

- Medir as diferenças dos níveis de atividades empreendedoras entre os países;
- Descobrir os fatores favoráveis ao empreendedorismo nos países estudados;
- Sugerir políticas que possam aumentar o nível nacional de atividade empreendedora.

No Brasil, esta pesquisa vem sendo conduzida pelo Instituto Brasileiro de Qualidade e Produtividade (IBQP) desde o ano de 2000, com o apoio técnico e financeiro do SEBRAE. Para avaliar o nível da atividade empreendedora de cada país participante, são entrevistados membros da população adulta (18 a 64 anos) selecionados por meio de amostra probabilística. Os empreendedores identificados nas entrevistas são classificados conforme o desenvolvimento do empreendimento, sua motivação para empreender e suas características sociodemográficas. A metodologia pode ser consultada em detalhes no relatório da pesquisa *Empreendedorismo no Brasil 2012* (GEM, 2012).

O banco de dados utilizado neste trabalho foi fornecido pelo IBQP e contempla 10.000 indivíduos entrevistados na pesquisa realizada no ano de 2012. Para estes indivíduos foram registrados: a) o estágio do empreendedorismo de acordo com a seguinte classificação: não empreendedores, empreendedores iniciais e empreendedores estabelecidos e; b) informações sociodemográficas: região geográfica do Brasil em que reside, gênero, faixa etária, nível de escolaridade e renda salarial. Para os 1.549 indivíduos classificados como empreendedores iniciais foi também registrada a motivação (oportunidade ou necessidade) que os levou a abrir um novo negócio. O Quadro 1 a seguir apresenta uma descrição das variáveis disponíveis.

Quadro 1 - Descrição das variáveis pesquisadas no GEM 2012 - Brasil

Item	Descrição	Categorias
Variáveis Respostas		
Estágio do Empreendedorismo	Classificação do indivíduo entrevistado quanto ao empreendedorismo	- Não Empreendedores
		- Empreendedores Iniciais (Negócio aberto até 42 meses)
		- Empreendedores Estabelecidos (Negócio aberto a mais de 42 meses)
Motivação para o empreendedorismo	Motivo pelo qual o empreendedor inicial optou por abrir um novo negócio	- Oportunidade
		- Necessidade
Variáveis Explicativas		
Escolaridade	Nível escolar do indivíduo entrevistado	- Abaixo do segundo grau
		- Segundo grau ou Superior
Faixa de Renda	Renda familiar, em salários mínimos	- Mais que 03 salários
		- De 01 a 03 salários
Faixa Etária	Idade, em anos	- 18-33 anos
		- 34-48 anos
		- 49-64 anos
Gênero	Gênero	- Feminino
		- Masculino
Região	Região demográfica na qual o entrevistado reside	- Centro-Oeste
		- Nordeste
		- Norte
		- Sudeste
		- Sul

Fonte: Os Autores (2014).

A análise dos dados descritos foi realizada em duas etapas. A primeira, considerando toda a amostra (10.000 indivíduos), utilizou as informações sociodemográficas (região geográfica, gênero, faixa etária, nível de escolaridade e renda salarial) e relacionou estas com a classificação do indivíduo referente ao estágio do empreendedorismo (isto é, não empreendedores, empreendedores iniciais e empreendedores estabelecidos). A segunda considerou apenas os empreendedores iniciais (1.549 indivíduos) e, com base nas mesmas informações sociodemográficas, relacionou estas com a motivação (oportunidade ou necessidade) que os levou a abrir um novo negócio.

Alguns ajustes foram realizados no banco de dados para iniciar as análises referentes aos dois estudos. Primeiramente, os indivíduos classificados simultaneamente na pesquisa como empreendedores iniciais e estabelecidos, foram considerados apenas como empreendedores estabelecidos. Ainda, indivíduos que optaram em não responder algumas das questões sobre as características sociodemográficas (abordadas no decorrer do estudo como variáveis explicativas) foram excluídos do banco de dados. Com estes ajustes, a frequência final de indivíduos dentro de cada estudo ficou da seguinte forma: no primeiro estudo 9.354, (6.402 não empreendedores, 1.531 empreendedores iniciais e 1.421 empreendedores estabelecidos) e no segundo estudo 1.531 (todos empreendedores iniciais).

3.1.2 Recursos Computacionais

O *software* R (R DEVELOPMENT CORE TEAM, 2013), versão 3.0.1, foi utilizado na análise estatística dos dados com o auxílio dos pacotes: VGAM, lattice, ggplot2 e GGally.

3.2 MÉTODOS

Para a análise dos dados descritos foram considerados métodos propostos na literatura para a análise de dados categorizados, tendo em vista a natureza categórica das variáveis respostas disponíveis (Quadro 1).

A análise de dados categorizados permite analisar as características ou atributos definidos pela combinação de categorias entre as variáveis de interesse. A

importância desta área na estatística está relacionada ao fato de estar presente nos mais variados campos da ciência (PAULINO; SINGER, 2006). De acordo com estes autores, as variáveis no estudo seguem duas classificações:

- Variáveis explicativas: são aquelas que possibilitam o entendimento das variáveis resposta;
- Variáveis respostas: são as que descrevem as particularidades de cada unidade amostral ou categoria, por meio de um modelo probabilístico.

A seguir são descritos os modelos estatísticos considerados neste trabalho.

3.2.1 Modelo de logitos generalizados

A primeira análise abordada objetivou estudar a relação entre as variáveis sociodemográficas (região geográfica, gênero, faixa etária, nível de escolaridade e renda) dos entrevistados com a variável estágio do empreendedorismo (não empreendedores, empreendedores iniciais e empreendedores estabelecidos). Como o estágio do empreendedorismo é uma variável politômica e de natureza nominal, ou seja, apresenta mais de duas categorias e não possui ordenação dentre elas, o modelo utilizado para esta análise foi o de logitos generalizados.

Como exemplo para a apresentação deste modelo, considere duas variáveis explicativas X_1 e X_2 (uma com três e a outra com duas categorias) e uma variável resposta Y composta por três categorias tal que:

- $h = 1, 2, 3$ indexam as categorias da variável explicativa X_1 ;
- $i = 1, 2$ indexam as categorias da variável explicativa X_2 ;
- $k = 1, 2, 3$ indexam as categorias da variável resposta.

Diante disto, a probabilidade associada a k -ésima categoria da variável resposta Y e as categorias h e i das variáveis explicativas X_1 e X_2 fica denotada por:

$$\pi_{hik}.$$

A Tabela 1 fornece a representação das probabilidades para este exemplo.

Tabela 1 - Representação das probabilidades associadas a um estudo ilustrativo com duas variáveis explicativas X_1 e X_2 categóricas e uma variável resposta politômica

Variável Explicativa 1	Variável Explicativa 2	Variável Resposta		
		Categorias de Y		
Categorias de X_1	Categorias de X_2	1	2	3
1	1	π_{111}	π_{112}	π_{113}
	2	π_{121}	π_{122}	π_{123}
2	1	π_{211}	π_{212}	π_{213}
	2	π_{221}	π_{222}	π_{223}
3	1	π_{311}	π_{312}	π_{313}
	2	π_{321}	π_{322}	π_{323}

Fonte: Os autores (2014).

No modelo de logitos generalizados, cada logito é definido a partir da probabilidade de cada categoria de resposta sobre a probabilidade da última categoria, ficando expresso por:

$$\text{logit}_{hik} = \ln\left(\frac{\pi_{hik}}{\pi_{hir}}\right),$$

para $k = 1, \dots, (r - 1)$, sendo r o número de categorias da variável resposta ($r = 3$).

Desse modo, para o exemplo representado na Tabela 1 tem-se os seguintes logitos generalizados associados às categorias da variável resposta:

$$\text{logit}_{hi1} = \ln\left(\frac{\pi_{hi1}}{\pi_{hi3}}\right) \text{ e } \text{logit}_{hi2} = \ln\left(\frac{\pi_{hi2}}{\pi_{hi3}}\right).$$

Note que o logito correspondente à categoria de resposta $k=1$ em relação à categoria $k=2$ resulta da razão dos dois logitos apresentados, ou seja:

$$\text{logit}_{hi3} = \frac{\ln\left(\frac{\pi_{hi1}}{\pi_{hi3}}\right)}{\ln\left(\frac{\pi_{hi2}}{\pi_{hi3}}\right)} = \ln\left(\frac{\pi_{hi1}}{\pi_{hi2}}\right).$$

Considerando tais notações e conceitos, tem-se que a equação geral do modelo de logitos generalizados para este exemplo fica expressa por:

$$\text{logit}_{hik} = \beta_{0k} + \beta'_k x_{hi},$$

para $h = 1, 2, i = 1, 2, 3$ e $k = 1, 2, 3$.

Percebe-se que para cada logito do modelo existem diferentes interceptos

(β_{0k}) e diferentes vetores de parâmetros de regressão (β), ou seja, vários parâmetros devem ser estimados para cada subpopulação. Desta maneira, precisa-se de um tamanho amostral razoável para alocar tais logitos e seus graus de liberdade (GIOLO, 2013).

3.2.1.1 Estimação dos parâmetros para o modelo de logitos generalizados

Para estimar o vetor de parâmetros de regressão (β) do modelo de logitos generalizados, utilizou-se o método de máxima verossimilhança. Sendo assim, para o exemplo ilustrado na Tabela 1, o logaritmo da função de verossimilhança pode ser escrito da seguinte forma:

$$l(\beta) = \sum_{k=1}^3 [y_i \beta'_k x_{hi} - \ln(1 + \exp\{\beta'_k x_{hi}\})],$$

para $h = 1, 2$, $i = 1, 2, 3$ e $k = 1, 2, 3$.

Os valores de β que maximizam a função $l(\beta)$ são obtidos resolvendo-se o sistema de equações definido por $U(\beta) = \mathbf{0}$, em que $U(\beta)$ é o vetor escore de derivadas de primeira ordem da função $l(\beta)$. Tal sistema é dado por:

$$U(\beta) = \sum_{k=1}^3 \left[y_i x_{hi} - \frac{\exp\{\beta'_k x_{hi}\}}{1 + \exp\{\beta'_k x_{hi}\}} x_{hi} \right] = 0.$$

A solução deste sistema de equações requer a execução de métodos iterativos como, por exemplo, o método de Newton-Raphson.

3.2.1.2 Avaliação da qualidade do ajuste

O método *stepwise do tipo* seleção regressiva (*backward selection*) é utilizado para seleção de variáveis explicativas dentro de cada modelo. Tal método consiste na retirada sequencial de termos do modelo até não obter uma melhora significativa no aperfeiçoamento do ajuste, ou seja, uma variável explicativa é selecionada para permanecer no modelo, caso produza mudanças significativas na função $l(\beta)$ associada ao modelo que não contém a mesma conforme o nível de significância adotado (PAULINO; SINGER, 2006). Neste trabalho o nível de significância, para inclusão ou remoção de variáveis, foi de 5%.

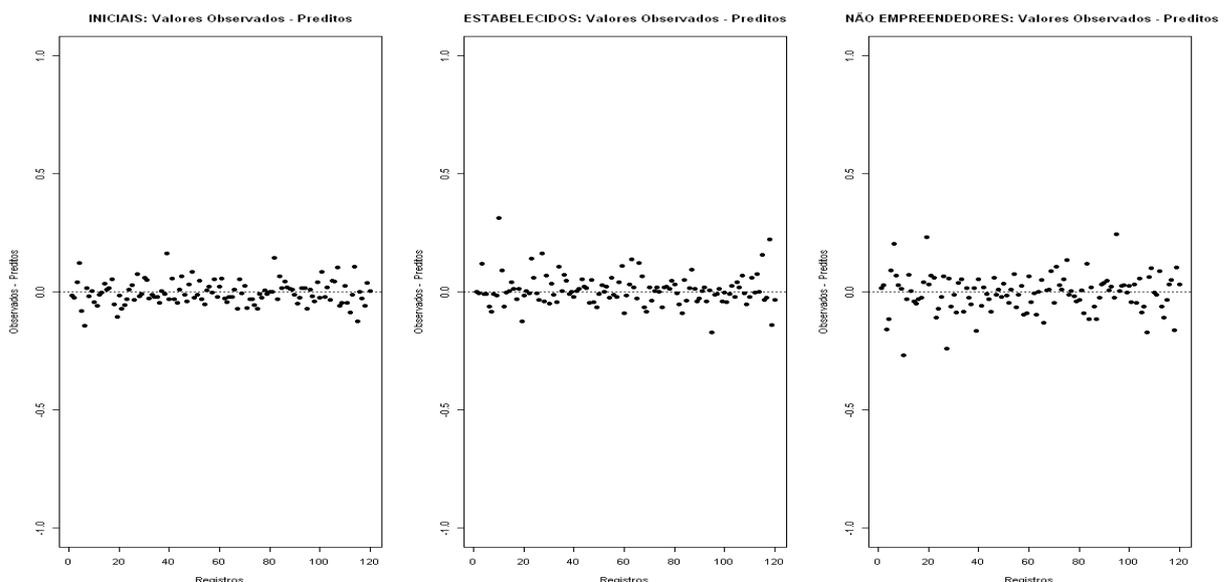
O modelo que melhor se ajustou aos dados foi determinado por meio da comparação com outros modelos com diversas conformações de variáveis explicativas. O critério de comparabilidade utilizado foi o critério de informação de AKAIKE (AIC), o qual é definido como:

$$AIC = 2l - 2p,$$

em que l é o logaritmo neperiano do máximo da função de verossimilhança e p é o número de parâmetros associados ao modelo considerado. Conforme o critério AIC, para que um determinado modelo seja considerado melhor que outro modelo, o mesmo deve apresentar um menor valor de AIC (SAKAMOTO et al, 1986).

Os gráficos confrontando os valores observados em relação aos valores preditos em termos de suas respectivas diferenças foram considerados para cada modelo. A Figura 1 ilustra o comportamento gráfico esperado para as diferenças entre estes valores (pequenas e em torno do zero). Na sequência para complementar análise de qualidade do modelo, avaliamos os gráficos dos resíduos Brutos e de Pearson, para verificar o comportamento dos resíduos e a existência de pontos discrepantes (*outliers*).

Figura 1 – Exemplo gráfico das diferenças entre os valores observados e valores preditos pelo modelo de logitos generalizados



Fonte: Os autores (2014).

3.2.1.3 Razão de chances

Segundo Bland e Altman (2000), nos últimos anos a razão de chances (do inglês *odds ratio*) se tornou amplamente utilizada em relatórios médicos devido à facilidade de sua obtenção e interpretação em uma variedade de estudos.

A razão de chances é uma medida de associação utilizada em muitas áreas e é de grande interesse no modelo de regressão logística. Como mencionado na seção 3.2.2, chance (ou possibilidade) é a probabilidade de ocorrência deste evento dividida pela probabilidade da não ocorrência do mesmo evento (PAULA, 2004).

Desse modo, define-se razão de chances, denotada por OR , como a razão entre a chance de um evento ocorrer em um grupo e a chance deste mesmo evento ocorrer em outro grupo. Assim, se as probabilidades de um evento em cada um dos grupos forem, respectivamente, p e q , então a razão de chances é dada por:

$$\widehat{OR} = \frac{p/(1-p)}{q/(1-q)} = \frac{p(1-q)}{q(1-p)}.$$

Por exemplo, se a variável resposta é a presença ou ausência de câncer de pulmão e se a variável explicativa representa se a pessoa é ou não fumante, uma razão de chances com valor estimado igual a $\widehat{OR} = 2$ pode ser interpretada como: a chance de uma pessoa que fuma desenvolver câncer de pulmão é duas vezes a de uma pessoa que não fuma.

A razão de chances nos modelos de regressão logística, incluindo somente os efeitos principais, é definida pela exponenciação dos parâmetros e fica expressa por:

$$OR = \exp(\beta_k),$$

em que β_k ($k = 1, \dots, p$) são os p parâmetros de regressão. Esta definição tornou o modelo de regressão logística bastante popular.

O intervalo de confiança (IC) para a OR quando determinada categoria de uma variável explicativa é comparada com a categoria de referência desta variável pode ser obtido da seguinte forma:

$$IC(OR) = \left(\exp \left\{ \hat{\beta}_k - z_{\frac{\alpha}{2}} SE(\hat{\beta}_k) \right\}; \exp \left\{ \hat{\beta}_k + z_{\frac{\alpha}{2}} SE(\hat{\beta}_k) \right\} \right),$$

em que $SE(\hat{\beta}_k)$ denota o erro padrão de $\hat{\beta}_k$ e $z_{\frac{\alpha}{2}}$ o percentil $\alpha/2$ da distribuição normal padrão.

Já para a comparação entre duas categorias (digamos k e j) de uma variável explicativa, com ambas não sendo a categoria de referência, o correspondente intervalo de confiança para OR pode ser obtido por:

$$IC(OR) = \left(\exp \left\{ \hat{\beta}_k - \hat{\beta}_j \pm z_{\frac{\alpha}{2}} \sqrt{(var(\hat{\beta}_k) + var(\hat{\beta}_j)) - 2 \cdot cov(\hat{\beta}_k, \hat{\beta}_j)} \right\} \right),$$

em que $\hat{\beta}_k$ e $\hat{\beta}_j$ denotam os estimadores dos k -ésimo e j -ésimo parâmetros de regressão, $j \neq k$, $var(\hat{\beta}_k)$ e $var(\hat{\beta}_j)$ são as variâncias de $\hat{\beta}_k$ e $\hat{\beta}_j$ e $cov(\hat{\beta}_k, \hat{\beta}_j)$ a covariância entre $\hat{\beta}_k$ e $\hat{\beta}_j$.

Nos dois casos, o valor 1 pertencendo ao $IC(OR)$ implicará que as chances (*odds*) não diferem entre as categorias estudadas.

3.2.2 Modelos para dados com resposta binária

Para analisar a relação entre as mesmas variáveis sociodemográficas utilizadas no modelo anterior, considerando a variável resposta Y = motivação dos empreendedores ao abrir um negócio (oportunidade ou necessidade) como dicotômica, assumindo o valor $Y = 1$ se uma das categorias desta variável ocorreu (oportunidade) e $Y = 0$, caso contrário (necessidade). Desse modo, modelos de regressão para dados binários, os quais são úteis para descrever a associação entre uma variável resposta dicotômica e um conjunto de variáveis explicativas, foram considerados para proceder à análise mencionada.

Assim como nos modelos de regressão linear, o interesse nos modelos para dados binários está em modelar o valor médio da variável resposta condicionada aos valores $\mathbf{x} = (x_1, \dots, x_p)$ de p variáveis explicativas, isto é, $E(Y|\mathbf{x})$. Contudo, diferente dos modelos de regressão linear em que $E(Y|\mathbf{x})$ pertence ao intervalo $(-\infty, \infty)$, nos modelos para dados binários $E(Y|\mathbf{x})$ pertence ao intervalo $[0,1]$ (HOSMER; LEMESHOW, 1989).

Dentre os modelos propostos na literatura para modelar dados com resposta binária, optou-se neste trabalho pelo modelo de regressão logística, pois é extremamente flexível do ponto de vista matemático e conduz à interpretações simples dos parâmetros (GIOLO, 2013).

Considerando um conjunto de variáveis explicativas $\mathbf{X} = (X_1, \dots, X_p)$ e

denotando $E(Y|\mathbf{x}) = P(Y=1|\mathbf{x})$ por $\theta(\mathbf{x})$, segue que o modelo de regressão logística fica expresso por:

$$\theta(\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}} = \frac{\exp\{\beta_0 + \sum_{k=1}^p \beta_k x_k\}}{1 + \exp\{\beta_0 + \sum_{k=1}^p \beta_k x_k\}}$$

em que $\mathbf{x} = (x_1, \dots, x_p)$ denota o vetor de valores observados das variáveis explanatórias \mathbf{X} , β_0 é uma constante e β_k ($k = 1, \dots, p$) são os p parâmetros de regressão. O modelo fornece, portanto, a probabilidade de um indivíduo com vetor de valores observados \mathbf{x} , apresentar a resposta de interesse, isto é, $Y = 1$. Conseqüentemente,

$$1 - \theta(\mathbf{x}) = \frac{1}{1 + \exp\{\beta_0 + \sum_{k=1}^p \beta_k x_k\}}$$

fornece a probabilidade deste indivíduo não apresentar a referida resposta.

Observe, ainda, que a transformação em $\theta(\mathbf{x})$ definida pelo logaritmo neperiano da razão entre $\theta(\mathbf{x})$ e $1 - \theta(\mathbf{x})$ fornece um modelo linear, isto é,

$$\ln\left(\frac{\theta(\mathbf{x})}{1 - \theta(\mathbf{x})}\right) = \beta_0 + \sum_{k=1}^p \beta_k x_k = \boldsymbol{\beta}'\mathbf{x}.$$

Tal transformação é denominada logito. Ainda, a razão entre $\theta(\mathbf{x})$ e $1 - \theta(\mathbf{x})$ é denominada chance (do inglês *odds*), ou seja:

$$odds = chance = \frac{\theta(\mathbf{x})}{1 - \theta(\mathbf{x})} = \exp(\boldsymbol{\beta}'\mathbf{x}).$$

Em relação à seleção das covariáveis, avaliação do modelo e interpretação dos coeficientes, foram utilizados os mesmos métodos citados na seção 3.2.1. Contudo, no modelo para dados binários foi adicionada à avaliação do modelo os resíduos deviance e a curva ROC (*receiver operating characteristic*). Para a construção da curva ROC, pares de pontos $(x, y) = (1 - \text{especificidade}, \text{sensibilidade})$ são representados graficamente para diferentes pontos de corte. O modelo com discriminação perfeita seria aquele com sensibilidade e especificidade igual a 1, o que corresponderia a $(x, y) = (0, 1)$. Assim, quanto mais próxima de 1 for a área abaixo da curva ROC, melhor o poder preditivo do modelo.

3.2.2.1 Estimação dos parâmetros para o modelo com resposta binária

Suponha (y_i, x_i) , $i = 1, \dots, n$, uma amostra independente de n pares de observações com y_i representando o valor da variável resposta dicotômica e x_i o vetor de valores das variáveis explicativas associadas ao indivíduo i . Para o ajuste do modelo é necessário estimar o vetor de parâmetros desconhecidos β , o que é usualmente feito pelo método de máxima verossimilhança.

Como $Y_i \sim \text{Bernoulli}(\pi_i)$, segue que sua respectiva função de probabilidade fica dada por:

$$f(y_i, \pi_i) = \theta_i^{y_i}(1 - \theta_i)^{1-y_i},$$

assim como a função de probabilidade conjunta, devido às observações serem independentes, por:

$$\prod_{i=1}^n f(y_i, \pi_i) = \prod_{i=1}^n \theta_i^{y_i}(1 - \theta_i)^{1-y_i}, y_i \in [0,1].$$

Desse modo, a função de verossimilhança fica expressa por:

$$L(\beta) = \prod_{i=1}^n \theta_i^{y_i}(1 - \theta_i)^{1-y_i}, \quad \beta \in \mathbb{R}.$$

Por fim, as estimativas dos parâmetros β serão os valores que maximizam a função de verossimilhança, ou o logaritmo desta função, ou seja:

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^n y_i \ln(\theta(x_i)) + (1 - y_i) \ln(1 - \theta(x_i)). \quad (3.2)$$

Para maximizar (3.2) deve-se diferenciá-la com respeito a cada parâmetro β_j ($j = 0, 1, \dots, p$) obtendo-se o sistema de $p + 1$ equações de máxima verossimilhança a seguir.

$$\begin{cases} \sum_{i=1}^n (y_i - \theta(x_i)) = 0 \\ \sum_{i=1}^n x_{ij} (y_i - \theta(x_i)) = 0 \end{cases} \quad j = 0, 1, \dots, p.$$

Como este sistema de equação não é linear nos parâmetros β_j , $j = 0, 1, \dots, p$,

métodos iterativos são necessários para sua solução. Os métodos iterativos de Newton-Raphson e o escore de Fisher são algoritmos comumente utilizados para esta finalidade (McCULLAGH e NELDER, 1983).

4 RESULTADOS E DISCUSSÃO

4.1 ANÁLISES DESCRITIVAS

Inicialmente, com o auxílio de ferramentas gráficas, foram realizadas análises exploratórias, as quais são descritas a seguir.

4.1.1 Empreendedorismo por regiões geográficas

A partir da Figura 2, que apresenta a distribuição do empreendedorismo segundo o estágio do empreendimento e a motivação, observa-se a maior proporção de empreendedores iniciais na região Centro-Oeste (17,3%). O mesmo ocorre quando a motivação de abrir o negócio é por oportunidade (84,2%). Por outro lado, a região com a menor proporção de empreendedores iniciais é a Sudeste, que não está entre as regiões com menores proporções de empreendedores por oportunidade. Em contrapartida, as regiões Norte e Nordeste tiveram as maiores proporções de empreendedores motivados por necessidade, 43,71% e 39,58%, respectivamente.

Figura 2 – Percentuais da população adulta segundo o estágio do empreendedorismo e a motivação observados nas regiões brasileiras

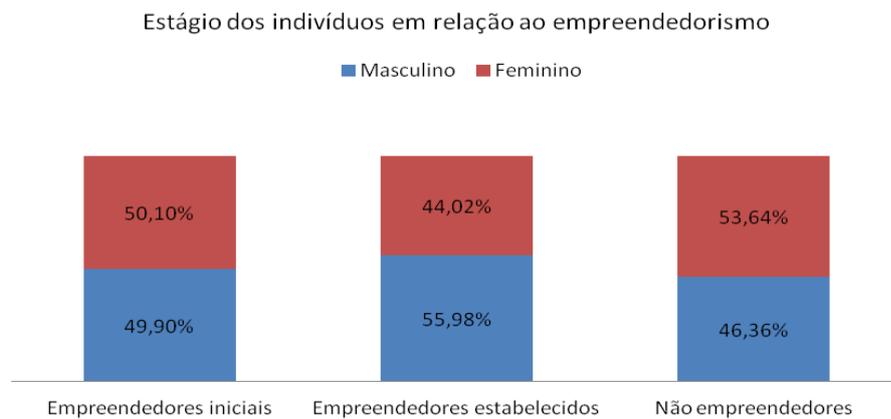


Fonte: Os autores (2014).

4.1.2 Empreendedorismo por gênero

A Figura 3 apresenta a distribuição da população adulta brasileira segundo o empreendedorismo por gênero. Destaca-se a semelhança entre os gêneros nos empreendedores iniciais. Já entre os empreendedores estabelecidos, nota-se uma prevalência maior do gênero masculino (55,98%). Quando se observa os não empreendedores, a proporção é maior no gênero feminino (53,64%).

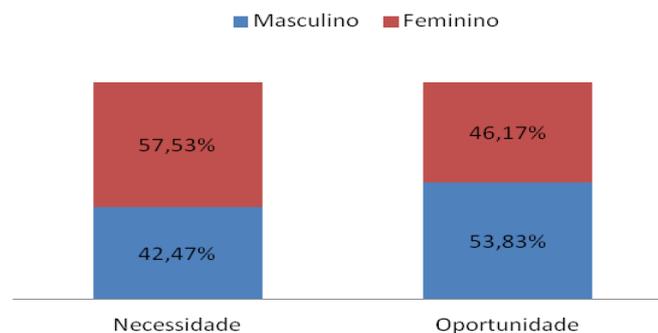
Figura 3 – Percentuais da população adulta segundo o estágio do empreendedorismo



Fonte: Os autores (2014).

Quanto aos empreendedores iniciais, pode-se observar na Figura 4 que as mulheres são mais propensas a iniciar um empreendimento por necessidade (57,53%), ao contrário dos homens que têm maior frequência de indivíduos que começam o negócio por oportunidade (53,83%).

Figura 4 – Percentuais dos empreendedores iniciais segundo a motivação por gênero

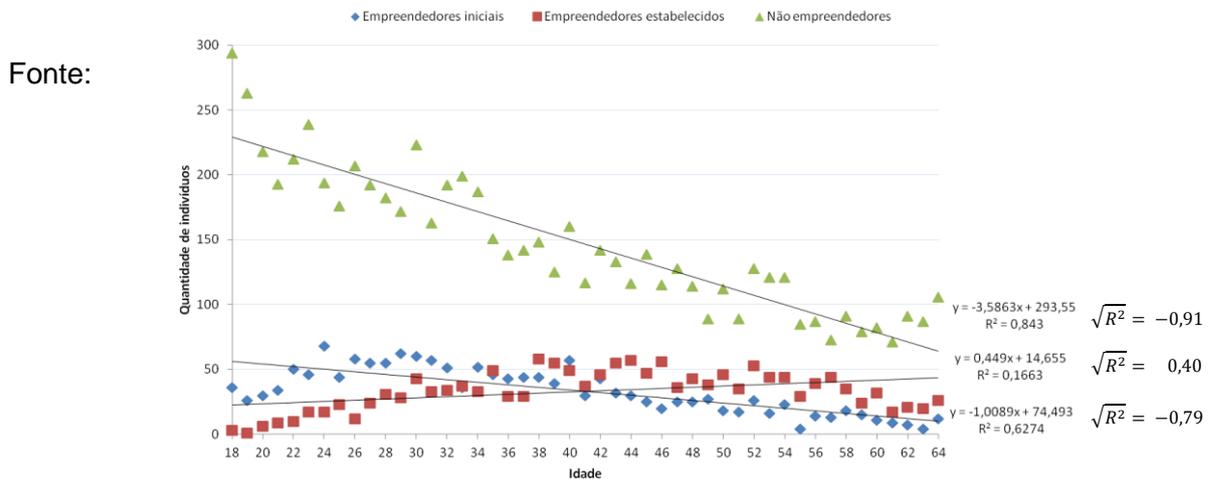


Fonte: Os autores (2014).

4.1.3 Empreendedorismo por idade

A distribuição dos indivíduos de acordo com a idade mostrada na Figura 5 revela diferenças no comportamento segundo o estágio do empreendimento. Entre os não empreendedores e os empreendedores iniciais há uma frequência maior de indivíduos jovens e esta relação diminui com o aumento da idade, indicando fortes correlações negativas, $\sqrt{R^2} = -0,91$ para os não empreendedores e $\sqrt{R^2} = -0,79$ para os empreendedores iniciais. Entretanto, nos empreendedores estabelecidos existe um indicativo de que a maior frequência está nas idades intermediárias e mais altas.

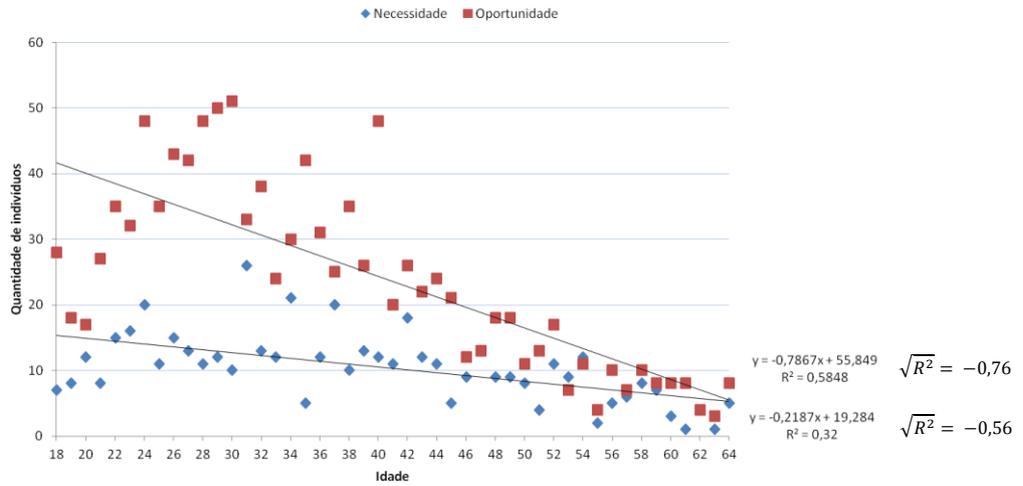
Figura 5 – Dispersão da população adulta segundo o estágio por idade



Os autores (2014).

A dispersão dos empreendedores iniciais segundo a motivação, mostrada na Figura 6, apresenta uma forte correlação negativa entre os empreendedores motivados por oportunidade, $\sqrt{R^2} = -0,76$, ou seja, os indivíduos mais jovens têm maior predominância ao iniciar um novo negócio motivados por oportunidade. No entanto, esta mesma relação não acontece quando se observa os resultados dos empreendedores iniciais cuja motivação é a necessidade. Neste caso não ocorre um padrão claro na distribuição dos indivíduos conforme a idade.

Figura 6 – Dispersão dos empreendedores iniciais segundo a motivação por idade

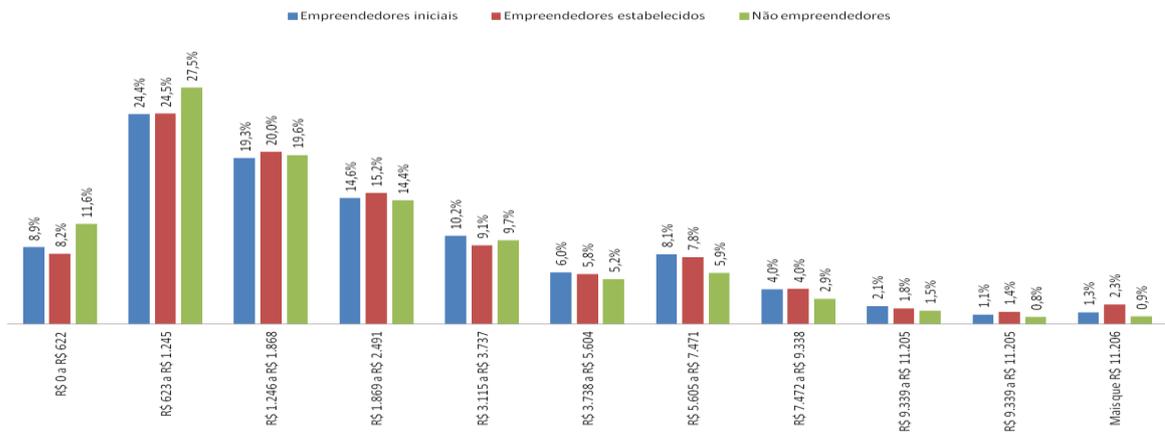


Fonte: Os autores (2014).

4.1.4 Empreendedorismo por renda familiar

O levantamento sobre a renda familiar no GEM 2012 foi realizado utilizando como referência o salário mínimo (R\$622,00). A Figura 7 apresenta os resultados relativos aos estágios de empreendedorismo, no qual se observa nas duas primeiras faixas salariais uma leve predominância dos indivíduos classificados como não empreendedores. Por outro lado, tal predominância não permanece nas demais faixas, havendo alternância entre os empreendedores iniciais e empreendedores estabelecidos.

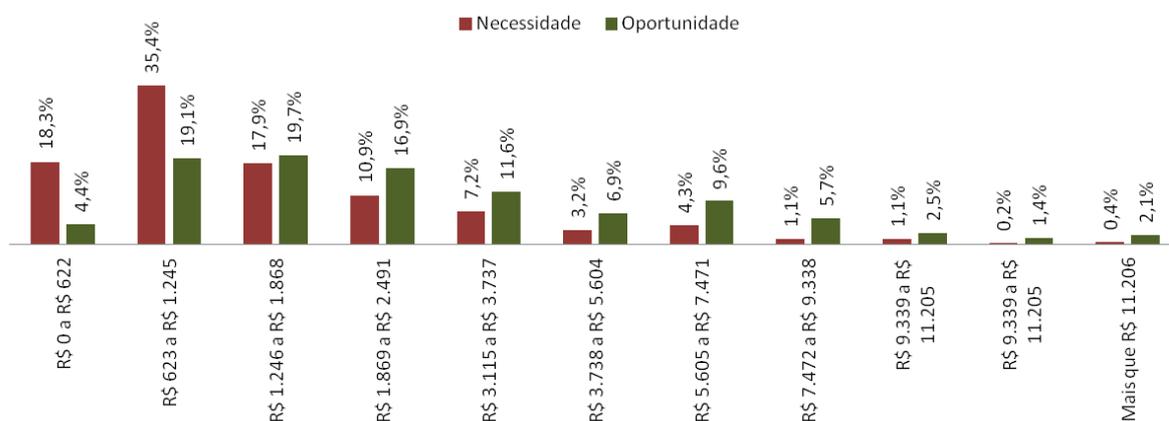
Figura 7 – Percentuais da população adulta segundo o estágio de empreendedorismo por renda familiar em salários mínimos



Fonte: Os autores (2014).

Em relação à motivação dos empreendedores iniciais segundo a renda familiar, observa-se na Figura 8 comportamentos diferentes entre os motivados por oportunidade e os motivados por necessidade. Os motivados por necessidade estão concentrados nas faixas salariais mais baixas, cerca de 71% estão nas faixas de até 3 salários mínimos. Em contrapartida, aqueles motivados por oportunidade têm as maiores frequências em faixa salariais mais altas, cerca de 57% estão acima de 3 salários mínimos.

Figura 8 – Percentuais dos empreendedores iniciais segundo a motivação por renda familiar em salários mínimos

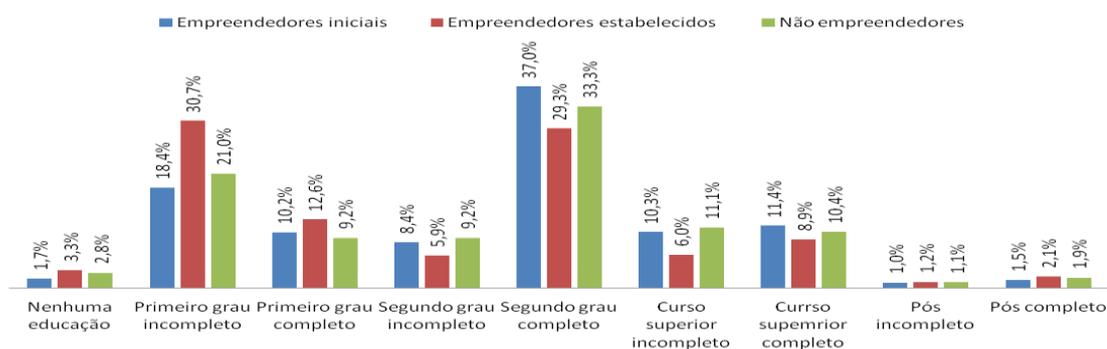


Fonte: Os autores (2014).

4.1.4 Empreendedorismo por escolaridade

Como pode ser observado na Figura 9, a escolaridade que predomina entre os empreendedores estabelecidos é a de primeiro grau incompleto (30,7%). No caso dos empreendedores iniciais, a maior frequência é de indivíduos com o segundo grau completo (37,0%), o que também ocorre com os não empreendedores (33,3%).

Figura 9 – Percentuais da população adulta segundo o estágio por escolaridade



Fonte: Os autores (2014).

A Figura 10 apresenta a distribuição dos empreendedores iniciais segundo a motivação em relação à escolaridade. Observa-se em ambos os casos (empreendedores por oportunidade e por necessidade), que ocorre a predominância de indivíduos com o segundo grau completo. Nota-se, também, que os níveis mais altos de escolaridade apresentam maior proporção de empreendedores motivados pela oportunidade.

Figura 10 – Percentuais dos empreendedores iniciais segundo a motivação por escolaridade



Fonte: Os autores (2014).

4.2 RESULTADOS DO MODELO DE LOGITOS GENERALIZADOS

Considerando o modelo de logitos generalizados para o primeiro estudo, que objetivou avaliar a relação entre cinco variáveis explicativas (Quadro 1) de natureza sociodemográficas com a variável estágio do empreendedorismo (não empreendedores, empreendedores iniciais e empreendedores estabelecidos), foram selecionadas para o modelo final, em decorrência do método de seleção regressiva (*backward selection*), ao nível de 5% de significância, todas as cinco variáveis descritas no Quadro 1. O modelo com estas variáveis apresentou o menor valor de AIC.

A Tabela 2 apresenta alguns dos modelos ajustados e seus respectivos valores AIC. Conforme citado anteriormente, o modelo contendo todas as cinco variáveis explicativas (último da Tabela 2) apresentou o menor valor AIC.

Tabela 2 – Resultados de alguns dos modelos ajustados aos dados de empreendedorismo

Modelos	g.l.	Deviance	Diferença de Deviances	Dif. de g.l.	AIC
Nulo	238	876,4751	-	-	1794,0623
X ₁	230	856,9116	19,5635	8	1790,4988
X ₂ X ₁	228	815,3085	41,6031	2	1752,8957
X ₃ X ₁ , X ₂	224	351,4804	463,8281	4	1297,0676
X ₄ X ₁ , X ₂ , X ₃	222	317,8640	33,6164	2	1267,4512
X ₅ X ₁ , X ₂ , X ₃ , X ₄	220	279,8048	38,0592	2	1233,3920

X₁: Região, X₂: Gênero, X₃: Faixa etária, X₄: Renda e X₅: Escolaridade; g.l. = graus de liberdade.

Fonte: Os autores (2014).

Para o modelo selecionado, as estimativas dos parâmetros encontram-se na Tabela 3. Por se tratarem de variáveis explicativas categóricas, uma categoria de cada uma delas foi utilizada como categoria de referência. São elas: região Centro-Oeste, gênero feminino, faixa etária de 18-33 anos, renda de 0 a 3 salários mínimos e escolaridade abaixo do segundo grau.

Tabela 3 – Estimativas, por logito, dos parâmetros do modelo final ajustado

(continua)

Coeficientes	Estimativas	Erro Padrão	Valor p
Logito 1 (Iniciais / Não Empreendedores)			
Intercepto	-1,4386	0,0898	0,0000
Região Nordeste	-0,1159	0,0873	0,1859
Região Norte	-0,1640	0,0905	0,0713
Região Sudeste	-0,3658	0,0924	0,0001
Região Sul	-0,2280	0,0902	0,0122
Masculino	0,1081	0,0579	0,0633
Faixa etária: 34-48 anos	0,1481	0,0651	0,0239
Faixa etária: 49-64 anos	-0,4085	0,0864	0,0000
Faixa de renda: mais que 3 salários	0,2823	0,0623	0,0000
Escolaridade: segundo grau ou superior	0,0058	0,0687	0,9330
Logito 2 (Estabelecidos / Não Empreendedores)			
Intercepto	-2,2327	0,1001	0,0000
Região Nordeste	-0,2305	0,0936	0,0146
Região Norte	-0,2264	0,0956	0,0188
Região Sudeste	-0,2012	0,0933	0,0320
Região Sul	-0,0808	0,0918	0,3798
Masculino	0,3640	0,0600	0,0000

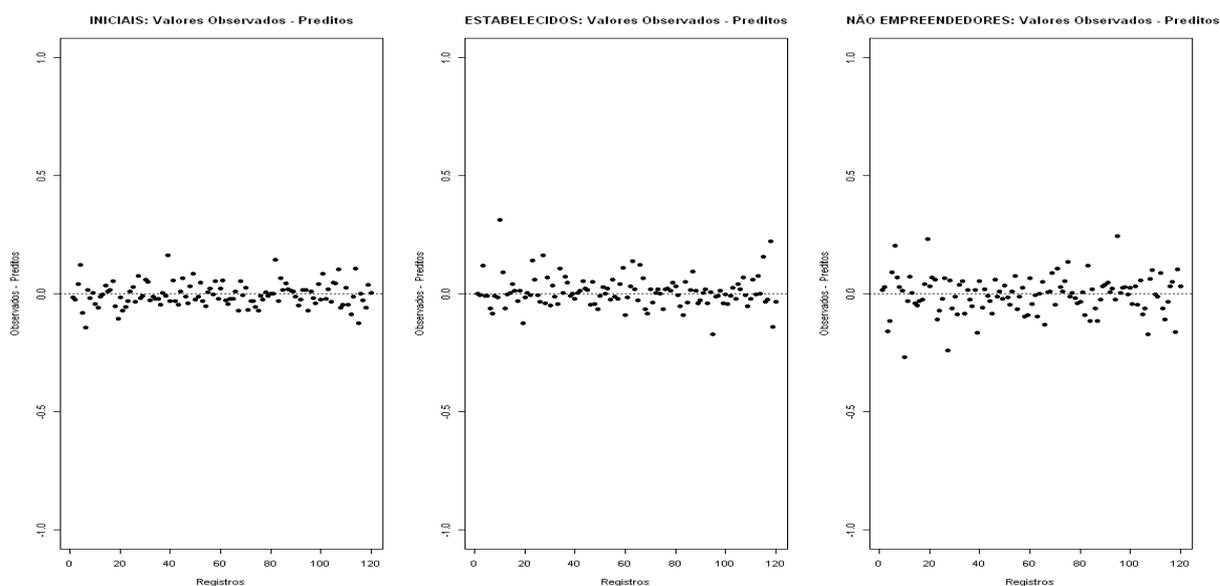
Tabela 3 – Estimativas, por logito, dos parâmetros do modelo final ajustado

Coeficientes	Estimativas	Erro Padrão	Valor p
Faixa etária: 34-48 anos	1,1631	0,0770	0,0000
Faixa etária: 49-64 anos	1,1819	0,0833	0,0000
Faixa de renda: mais que 3 salários	0,3903	0,0656	0,0000
Escolaridade: segundo grau ou superior	-0,4062	0,0672	0,0000
Logito 3 (Iniciais / Estabelecidos)			
Intercepto	0,7941	0,1223	0,0000
Região Nordeste	0,1147	0,1155	0,3219
Região Norte	0,0624	0,1185	0,5989
Região Sudeste	-0,1646	0,1187	0,1668
Região Sul	-0,1472	0,1158	0,2049
Masculino	-0,2559	0,0755	0,0008
Faixa etária: 34-48 anos	-1,0150	0,0916	0,0000
Faixa etária: 49-64 anos	-1,5904	0,1097	0,0000
Faixa de renda: mais que 3 salários	-0,1081	0,0819	0,1881
Escolaridade: segundo grau ou superior	0,4120	0,0864	0,0000

Fonte: Os autores (2014).

Quanto à adequação do modelo ajustado, observa-se na Figura 11, que as diferenças entre os valores das probabilidades observadas e as previstas a partir do modelo de logitos generalizados estão próximas de zero, indicando que o modelo selecionado apresenta ajuste satisfatório aos dados.

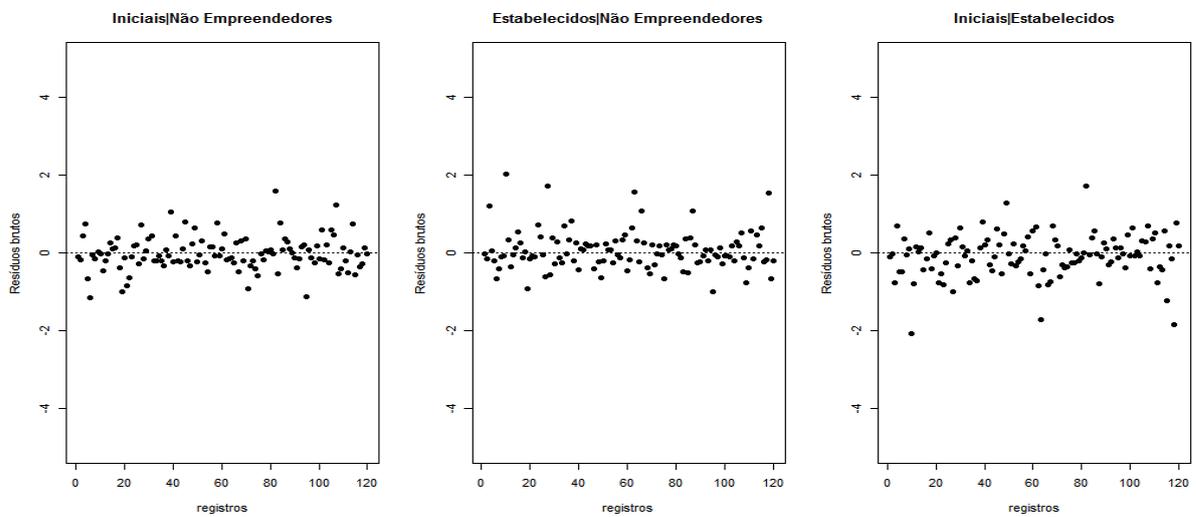
Figura 11 – Diferenças entre os valores observados e os valores previstos pelo modelo de logitos generalizados para cada estágio do empreendedorismo



Fonte: Os autores (2014).

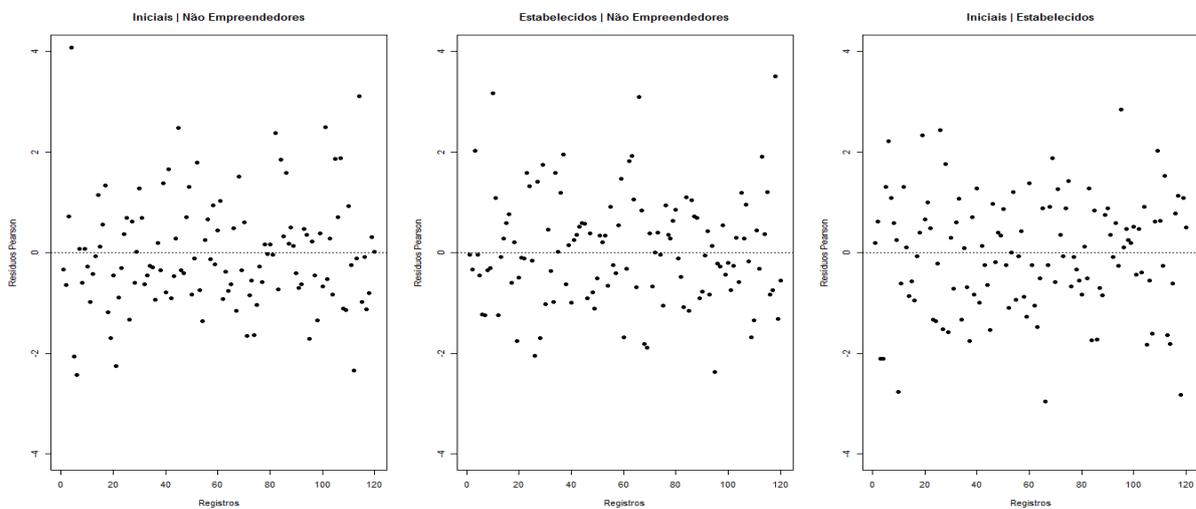
Para complementar o diagnóstico relativo à qualidade de ajuste do modelo, as Figuras 12 e 13 apresentam, respectivamente, os gráficos dos resíduos brutos e de Pearson para o modelo selecionado. Análise visual sugere adequação satisfatória do modelo aos dados, bem como ausência de pontos que possam ser considerados atípicos (*outliers*), pois os mesmos estão distribuídos aleatoriamente em torno de zero e dentro de uma faixa de variação aceitável, entre -3 e 3.

Figura 12 – Distribuição dos resíduos brutos associados ao modelo final e representados para cada logito estudado



Fonte: Os autores (2014).

Figura 13 – Distribuição dos resíduos de Pearson associados ao modelo final e representados para cada logito estudado



Fonte: Os autores (2014).

A partir das razões de chances (*odds ratios*) de empreendedores iniciais em relação aos não empreendedores, apresentadas na Tabela 4, têm-se as conclusões a seguir:

Tabela 4 – *Odds ratios* entre empreendedores iniciais e não empreendedores

Variáveis	Categorias	Logito 1 (Iniciais / Não Empreendedores)			
		<i>Odds ratios</i> : π_{h1}/π_{h3}		I.C. 95%	
		Expressão	Valor	I.i.	I.s.
Região	Centro-Oeste Nordeste	$1 / \exp\{\beta_{11}\}$	1,1229	0,9462	1,3325
	Centro-Oeste Norte	$1 / \exp\{\beta_{21}\}$	1,1782	0,9867	1,4068
	Centro-Oeste Sudeste	$1 / \exp\{\beta_{31}\}$	1,4417	1,2028	1,7281 *
	Centro-Oeste Sul	$1 / \exp\{\beta_{41}\}$	1,2561	1,0525	1,4990 *
	Nordeste Norte	$\exp\{\beta_{11}\} / \exp\{\beta_{21}\}$	1,0493	0,8769	1,2556
	Nordeste Sudeste	$\exp\{\beta_{11}\} / \exp\{\beta_{31}\}$	1,2840	1,0690	1,5422 *
	Nordeste Sul	$\exp\{\beta_{11}\} / \exp\{\beta_{41}\}$	1,1186	0,9353	1,3379
	Norte Sudeste	$\exp\{\beta_{21}\} / \exp\{\beta_{31}\}$	1,2237	1,0162	1,4736 *
	Norte Sul	$\exp\{\beta_{21}\} / \exp\{\beta_{41}\}$	1,0661	0,8884	1,2793
	Sul Sudeste	$\exp\{\beta_{41}\} / \exp\{\beta_{31}\}$	1,1478	0,9534	1,3818
Gênero	Masculino Feminino	$\exp\{\beta_{51}\}$	1,1142	0,9946	1,2481
Faixa Etária	34-48 anos 18-33 anos	$\exp\{\beta_{61}\}$	1,1596	1,0207	1,3176 *
	18-33 anos 49-64 anos	$1 / \exp\{\beta_{71}\}$	1,5045	1,2702	1,7821 *
	34-48 anos 49-64 anos	$\exp\{\beta_{61}\} / \exp\{\beta_{71}\}$	1,7447	1,4682	2,0733 *
Faixa de Renda	Mais que 3 salários Até 3 salários	$\exp\{\beta_{81}\}$	1,3262	1,1736	1,4985 *
Escolaridade	Segundo grau ou Superior Abaixo do Segundo grau	$\exp\{\beta_{91}\}$	1,0058	0,9945	1,0173

Fonte: Os autores (2014).

- Indivíduos que residem na Região Centro-Oeste apresentaram chance de se tornarem empreendedores iniciais, em relação aos da Região Sudeste, de aproximadamente 1,4 vezes e, em relação aos da Região Sul, de 1,3 vezes. Ainda, comparando os indivíduos da Região Nordeste e da Região Norte com os da Região Sudeste, tais chances foram de 1,3 e 1,2, respectivamente.

- Quanto à variável gênero, a comparação entre os empreendedores iniciais e os não empreendedores não mostrou diferenças significativas ao nível de 5%, indicando que ambos os sexos têm comportamentos semelhantes.
- Quanto à idade, ocorreu a predominância de empreendedores iniciais entre os indivíduos da faixa etária de 34-48 anos sobre as demais faixas. A maior razão de chances observada foi de 1,7 vezes, correspondendo à comparação dos indivíduos de 34-48 anos com os de 49-64 anos. Os indivíduos da faixa 49-64 anos em relação aos da faixa etária de 18-33 anos também apresentaram menores chances. Com isto, observou-se entre estes indivíduos que empreendimentos iniciais têm maior chance de serem conduzidos por pessoas na faixa etária intermediária.
- Em se tratando da renda familiar, os indivíduos com tal renda acima de três salários mínimos, apresentaram chance de serem empreendedores iniciais 1,3 vezes a dos indivíduos com renda abaixo de três salários mínimos; esta estimativa variou, com 95% de confiança, entre 1,2 e 1,5.
- Assim como para variável gênero, a variável escolaridade não mostrou diferenças significativas, ao nível de 5%, indicando que ambos os níveis de escolaridade têm comportamentos semelhantes.

A partir das razões de chances (*odds ratios*) de empreendedores estabelecidos em relação aos não empreendedores, apresentadas na Tabela 5, obtêm-se as seguintes conclusões:

- Similar ao observado para o logito 1, os indivíduos que residem na Região Centro-Oeste apresentaram chance maior de se tornarem empreendedores estabelecidos em comparação aos indivíduos da Região Sudeste; esta chance foi de aproximadamente 1,2 vezes. As maiores diferenças foram encontradas na Região Nordeste, seguida da Região Norte, ambas com aproximadamente 1,3 vezes.

Tabela 5 – Odds ratios entre empreendedores estabelecidos e não empreendedores

Variáveis	Categorias	Logito 2 (Estabelecidos / Não Empreendedores)			
		Odds ratios: π_{h2}/π_{h3}		I.C. 95%	
		Expressão	Valor	I.i.	I.s.
Região	Centro-Oeste Nordeste	$1 / \exp\{\beta_{12}\}$	1,2593	1,0481	1,5129 *
	Centro-Oeste Norte	$1 / \exp\{\beta_{22}\}$	1,2541	1,0397	1,5126 *
	Centro-Oeste Sudeste	$1 / \exp\{\beta_{32}\}$	1,2229	1,0186	1,4682 *
	Centro-Oeste Sul	$1 / \exp\{\beta_{42}\}$	1,0841	0,9056	1,2978
	Norte Nordeste	$\exp\{\beta_{22}\} / \exp\{\beta_{12}\}$	1,0042	0,8274	1,2186
	Sudeste Nordeste	$\exp\{\beta_{32}\} / \exp\{\beta_{12}\}$	1,0298	0,8525	1,2439
	Sul Nordeste	$\exp\{\beta_{42}\} / \exp\{\beta_{12}\}$	1,1616	0,9642	1,3993
	Sudeste Norte	$\exp\{\beta_{32}\} / \exp\{\beta_{22}\}$	1,0255	0,8486	1,2393
	Sul Norte	$\exp\{\beta_{42}\} / \exp\{\beta_{22}\}$	1,1567	0,9589	1,3954
	Sul Sudeste	$\exp\{\beta_{42}\} / \exp\{\beta_{32}\}$	1,1280	0,9399	1,3537
Gênero	Masculino Feminino	$\exp\{\beta_{52}\}$	1,4390	1,2793	1,6187 *
Faixa Etária	34-48 anos 18-33 anos	$\exp\{\beta_{62}\}$	3,2000	2,7516	3,7215 *
	49-64 anos 18-33 anos	$\exp\{\beta_{72}\}$	3,2606	2,7695	3,8387 *
Faixa de Renda	34-48 anos 49-64 anos	$\exp\{\beta_{72}\} / \exp\{\beta_{62}\}$	1,0189	0,8879	1,1694
	Mais que 3 salários Até 3 salários	$\exp\{\beta_{82}\}$	1,4775	1,2992	1,6803 *
Escolaridade	Abaixo do Segundo grau Segundo grau ou Superior	$1 / \exp\{\beta_{92}\}$	1,5012	1,3160	1,7123 *

Fonte: Os autores (2014).

- Os indivíduos do gênero masculino apresentaram chance de 1,4 vezes em relação ao gênero feminino de se tornarem empreendedores estabelecidos, sugerindo que indivíduos homens são mais propensos a conduzirem um negócio por mais de 42 meses.
- As chances estimadas na comparação entre as faixas etárias foram maiores para as duas mais elevadas, 34-48 e 49-64 anos. Similarmente ao ocorrido no logito 1, a faixa de 34-48 anos predominou sobre a outras duas, entretanto foi seguida da faixa de 49-64 anos. Este resultado coincide com o que foi observado na Figura 5 (seção 4.1.3), cuja análise sugere que indivíduos mais velhos são mais propensos a conduzirem empreendimentos por mais de 42 meses.

- A variável renda familiar apresentou o mesmo resultado do logito 1, isto é, indivíduos na faixa salarial mais alta obtiveram maiores chances de se tornarem empreendedores estabelecidos em relação aos indivíduos da outra faixa.
- Quanto à escolaridade, ocorreu uma inversão das faixas ao compararmos com o ocorrido no logito 1; os indivíduos da faixa de escolaridade menor, ou seja, abaixo do segundo grau, obtiveram chance de se tornarem empreendedores estabelecidos de 1,5 vezes a chance da outra categoria.

A partir das razões de chances (*odds ratios*) de empreendedores iniciais em relação aos empreendedores estabelecidos, apresentadas na Tabela 6, têm-se as seguintes conclusões:

- Indivíduos que residem na Região Nordeste obtiveram maiores chances, aproximadamente 1,3, de se tornarem empreendedores iniciais, quando comparados aos indivíduos das regiões Sudeste e Sul.
- As mulheres apresentaram chance de 1,3 vezes a dos homens. Este resultado difere do ocorrido nos logitos 1 e 2, indicando que as mulheres têm mais propensão a iniciarem um novo negócio.
- No que se refere à idade, ocorreu a predominância de indivíduos da faixa etária de 18-33 anos sobre as demais faixas. A maior chance foi de 5 vezes na comparação com os indivíduos da faixa etária de 49-64 anos. Com isto, e como visto anteriormente na análise do logito 1, empreendimentos iniciais têm maior chance de serem conduzidos por pessoas mais jovens.
- Indivíduos com nível de escolaridade mais alto apresentaram maior chance de se tornarem novos empreendedores, aproximadamente 1,5 vezes a dos indivíduos com menor nível escolar.

Tabela 6 – Razões de chances (*odds ratios*) entre empreendedores iniciais e estabelecidos

Variáveis	Categorias	Logito 3 (Iniciais / Estabelecidos)			
		Odds ratios: π_{h1}/π_{h2}		I.C. 95%	
		Expressão	Valor	I.i.	I.s.
Região	Nordeste Centro-Oeste	$exp\{\beta_{13}\}$	1,1215	0,7016	1,1332
	Norte Centro-Oeste	$exp\{\beta_{23}\}$	1,0644	0,7492	1,1781
	Centro-Oeste Sudeste	$1 / exp\{\beta_{33}\}$	1,1789	0,9345	1,4872
	Centro-Oeste Sul	$1 / exp\{\beta_{43}\}$	1,1586	0,9182	1,4619
	Nordeste Norte	$exp\{\beta_{13}\} / exp\{\beta_{23}\}$	1,0536	0,8293	1,3386
	Nordeste Sudeste	$exp\{\beta_{13}\} / exp\{\beta_{33}\}$	1,3222	1,0406	1,68 *
	Nordeste Sul	$exp\{\beta_{13}\} / exp\{\beta_{43}\}$	1,2994	1,0282	1,642 *
	Norte Sudeste	$exp\{\beta_{23}\} / exp\{\beta_{33}\}$	1,2549	0,9862	1,5967
	Norte Sul	$exp\{\beta_{23}\} / exp\{\beta_{43}\}$	1,2332	0,9735	1,5623
	Sul Sudeste	$exp\{\beta_{43}\} / exp\{\beta_{33}\}$	1,0176	0,8035	1,2886
Gênero	Feminino Masculino	$1 / exp\{\beta_{53}\}$	1,2916	1,03	1,6196 *
Faixa Etária	18-33 anos 34-48 anos	$1 / exp\{\beta_{63}\}$	2,7595	2,3799	3,1995 *
	18-33 anos 49-64 anos	$1 / exp\{\beta_{73}\}$	4,9057	4,0997	5,8701 *
	34-48 anos 49-64 anos	$exp\{\beta_{63}\} / exp\{\beta_{73}\}$	1,7778	1,4597	2,1651 *
Faixa de Renda	Até 3 salários Mais que 3 salários	$1 / exp\{\beta_{83}\}$	1,1141	0,8985	1,3814
Escolaridade	Segundo grau ou Superior Abaixo do Segundo grau	$exp\{\beta_{93}\}$	1,5099	1,2861	1,7726 *

Fonte: Os autores (2014).

4.3 RESULTADOS DO MODELO PARA DADOS BINÁRIOS

O segundo estudo relacionou a motivação dos empreendedores iniciais na abertura de um negócio, seja por oportunidade ou necessidade, com as mesmas variáveis explicativas relatadas no primeiro estudo (Quadro 1). Neste caso, o modelo para dados binários utilizado foi o de regressão logística. Similarmente ao ocorrido no primeiro estudo, todas as cinco variáveis descritas no Quadro 1 foram selecionadas para o modelo final, porém a variável explicativa faixa etária apresentou valor p acima do nível de significância considerado (5%). No entanto optou-se por preservá-la no modelo, não somente para manter o padrão das análises realizadas anteriormente, mas também porque o modelo com estas variáveis apresentou o menor valor de AIC.

A Tabela 7 apresenta alguns dos modelos ajustados e seus respectivos valores AIC. Conforme citado anteriormente, o modelo contendo todas as cinco variáveis explicativas (último da Tabela 7) apresentou o menor valor AIC.

Tabela 7 - Resultados de alguns dos modelos de regressão logística ajustados aos dados

Modelos	g.l.	Deviance	Diferença de Deviances	Dif. de g.l.	AIC
<i>Nulo</i>	119	303,16	-	-	571,0318
X_1	115	216,67	86,49	4	492,5394
$X_1 X_2$	114	199,23	17,44	1	477,1026
$X_3 X_1, X_2$	112	189,37	9,86	2	471,2432
$X_4 X_1, X_2, X_3$	111	127,58	61,80	1	411,4453
$X_5 X_1, X_2, X_3, X_4$	110	107,63	19,94	1	393,5040

X_1 : Região, X_2 : Gênero, X_3 : Faixa etária, X_4 : Renda e X_5 : Escolaridade; g.l. = graus de liberdade.

Fonte: Os autores (2014).

As estimativas do modelo selecionado encontram-se na Tabela 8. Foram consideradas as mesmas categorias de referência do primeiro estudo, ou seja: região Centro-Oeste, gênero feminino, faixa etária de 18-33 anos, renda de 0 a 3 salários mínimos e escolaridade abaixo do segundo grau.

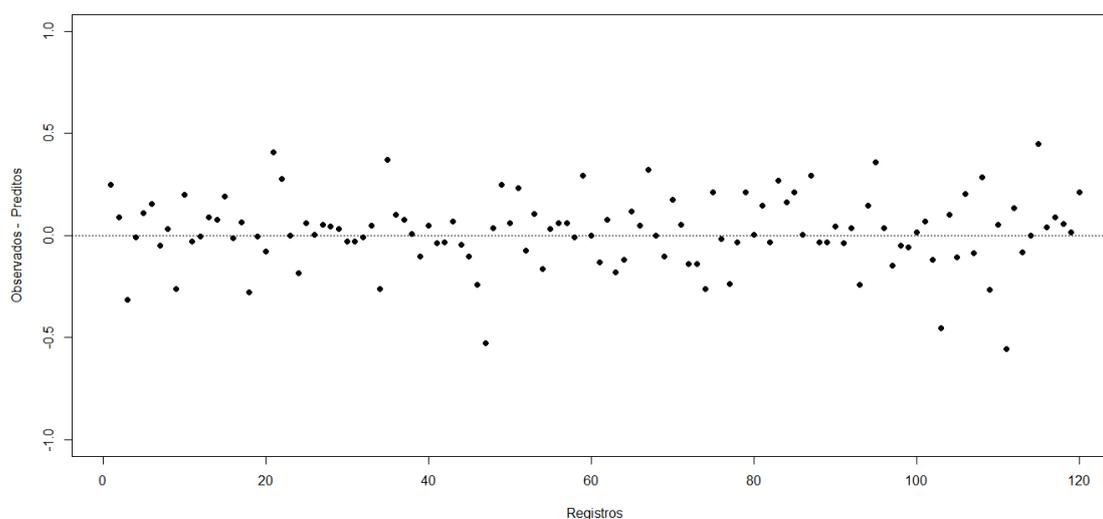
Tabelas 8 – Estimativas dos parâmetros do modelo de regressão logística final ajustado

Coeficientes	Oportunidade vs Necessidade		
	Estimativas	Erro Padrão	Valor p
Intercepto	0,7228	0,1984	0,0003
Região Nordeste	-1,0567	0,1983	0,0000
Região Norte	-1,2459	0,1965	0,0000
Região Sudeste	-0,6892	0,2131	0,0012
Região Sul	-0,5794	0,2103	0,0059
Masculino	0,4355	0,1179	0,0002
Faixa etária: 34-48 anos	-0,0193	0,1303	0,8820
Faixa etária: 49-64 anos	-0,3143	0,1713	0,0666
Faixa de renda: mais que 3 salários	0,7673	0,1273	0,0000
Escolaridade: segundo grau ou superior	0,5873	0,1311	0,0000

Fonte: Os autores (2014).

Referente à adequação do modelo selecionado, observa-se, a partir da Figura 14, o quão próximo de zero estão as diferenças entre os valores das probabilidades observadas e as previstas, sugerindo um ajuste satisfatório do modelo aos dados.

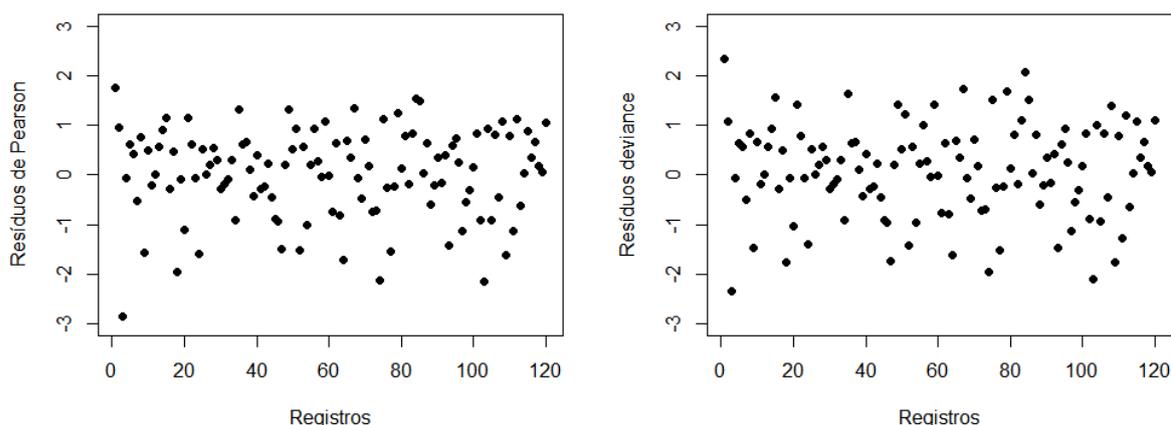
Figura 14 – Diferenças entre os valores observados e os valores previstos pelo modelo de regressão logística



Fonte: Os autores (2014).

Como complemento à análise de qualidade do modelo ajustado, observa-se na Figura 15 o comportamento dos resíduos de Pearson e resíduos *deviance*, cujos valores estão aleatoriamente distribuídos em torno de zero e dentro de uma faixa de variação aceitável, entre -3 e 3. Visualmente a análise indica ajuste satisfatório aos dados, com a ausência de pontos discrepantes (*outliers*).

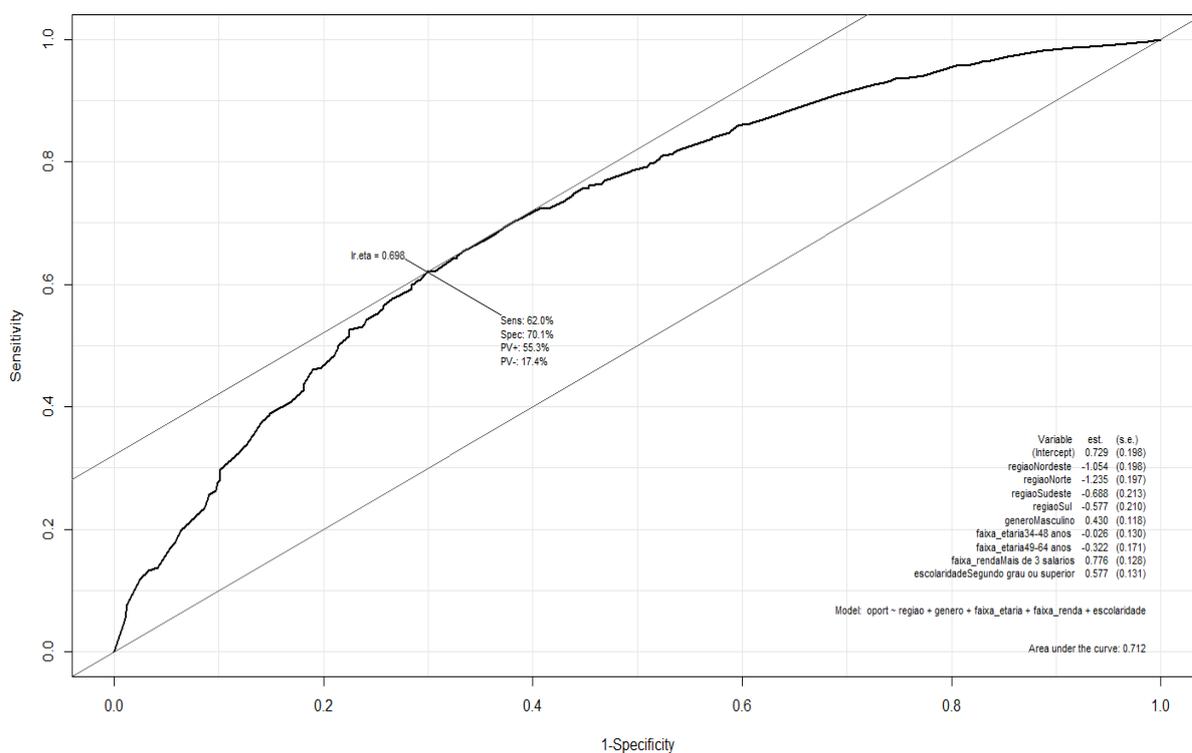
Figura 15 – Distribuição dos resíduos de Pearson e *deviance* associados ao modelo final



Fonte: Os autores (2014).

Dando continuidade à análise de qualidade do modelo, a curva ROC associada ao modelo encontra-se na Figura 16. De acordo com essa figura, há evidências de que o modelo apresenta ajuste satisfatório, pois a área sob a curva foi de 0,712, bem como sensibilidade e especificidade associadas ao ponto de corte 0,698 foram 62% e 70,1%, respectivamente. Em termos das probabilidades preditas pelo modelo, tal ponto de corte corresponde ao valor que proporcionou os melhores percentuais de acerto quanto à motivação de iniciar um novo negócio por parte dos empreendedores iniciais.

Figura 16 – Curva ROC associada ao modelo de regressão logística ajustado



Fonte: Os autores (2014).

A partir das razões de chances (*odds ratios*) entre os empreendedores iniciais motivados pela oportunidade em relação àqueles motivados pela necessidade, apresentadas na Tabela 9, têm-se as conclusões a seguir.

- Indivíduos que residem na Região Centro-Oeste predominaram sobre as demais regiões na chance de iniciarem um novo empreendimento motivados pela oportunidade, com destaque na comparação com os indivíduos da Região Norte, tal chance foi de 3,5 vezes. Nas comparações entre as demais regiões, prevaleceram as regiões Sudeste e Sul.
- Em se tratando do gênero, os homens apresentaram maior chance de iniciar um empreendimento motivados por oportunidade. Esta chance corresponde a 1,5 vezes a das mulheres.

Tabela 9 – Razões de chances (*odds ratios*) entre empreendedores iniciais motivados por oportunidade e os motivados por necessidade

Variáveis	Categorias	Oportunidade / Necessidade			
		Odds ratios: π_1/π_2		I.C. 95%	
		Expressão	Valor	I.i.	I.s.
Região	Centro-Oeste Nordeste	$1 / \exp\{\beta_1\}$	2,8769	1,9506	4,2431 *
	Centro-Oeste Norte	$1 / \exp\{\beta_2\}$	3,4762	2,3651	5,1094 *
	Centro-Oeste Sudeste	$1 / \exp\{\beta_3\}$	1,9921	1,3120	3,0249 *
	Centro-Oeste Sul	$1 / \exp\{\beta_4\}$	1,7851	1,1822	2,6954 *
	Nordeste Norte	$\exp\{\beta_1\} / \exp\{\beta_2\}$	1,2083	0,8806	1,6580
	Sudeste Nordeste	$\exp\{\beta_3\} / \exp\{\beta_1\}$	1,4441	1,0040	2,0772 *
	Sudeste Norte	$\exp\{\beta_3\} / \exp\{\beta_2\}$	1,7450	1,2193	2,4974 *
	Sul Nordeste	$\exp\{\beta_4\} / \exp\{\beta_1\}$	1,6117	1,1309	2,2967 *
	Sul Norte	$\exp\{\beta_4\} / \exp\{\beta_2\}$	1,9474	1,3737	2,7608 *
	Sul Sudeste	$\exp\{\beta_4\} / \exp\{\beta_3\}$	1,1160	0,7578	1,6436
Gênero	Masculino Feminino	$\exp\{\beta_5\}$	1,5458	1,2269	1,9475 *
Faixa Etária	18-33 anos 34-48 anos	$1/\exp\{\beta_6\}$	1,0195	0,7897	1,3163
	18-33 anos 49-64 anos	$1/\exp\{\beta_7\}$	1,3692	0,9787	1,9156
	34-48 anos 49-64 anos	$\exp\{\beta_6\} / \exp\{\beta_7\}$	1,3430	0,9584	1,8819
Faixa de Renda	Mais que 03 salários De 01 a 03 salários	$\exp\{\beta_8\}$	2,1539	1,6783	2,7644 *
Escolaridade	Segundo grau ou Superior Abaixo do Segundo grau	$\exp\{\beta_9\}$	1,7990	1,3914	2,3261 *

Fonte: Os autores (2014).

- Quanto à idade, a comparação entre os indivíduos motivados por oportunidade e os motivados pela necessidade não mostrou diferenças significativas entre as faixas etárias ao nível de 5%.
- Os indivíduos com renda familiar acima de três salários mínimos apresentaram chance de se tornarem empreendedores iniciais, motivados por oportunidade, de 2,2 vezes a dos indivíduos com renda abaixo de três salários mínimos.
- Considerando a escolaridade, os indivíduos com o nível mais alto apresentaram chance maior de empreender motivados pela oportunidade do que os com nível escolar mais baixo. Esta chance foi de 1,8 vezes.

5 CONCLUSÕES

O propósito deste estudo foi o de investigar, por meio de metodologias estatísticas, a influência de variáveis sociodemográficas sobre o empreendedorismo brasileiro. Para esse fim, foram utilizadas informações do projeto *Global Monitor Entrepreneurship* (GEM), a mais abrangente pesquisa sobre empreendedorismo no Brasil.

Primeiramente, foi utilizado o modelo de logitos generalizados para analisar a associação de diversas características sociodemográficas (registradas para 9.354 indivíduos brasileiros) com a classificação do estágio de empreendedores dos mesmos (empreendedor inicial, empreendedor estabelecido e não empreendedor). A segunda análise, restrita aos 1.531 empreendedores iniciais, consistiu em estudar, por meio do modelo de regressão logística, a associação das mesmas características sociodemográficas com a motivação de empreender (motivação por oportunidade ou por necessidade). As características sociodemográficas (variáveis explicativas), consideradas em ambos os modelos, foram: regiões geográficas, gênero, faixa etária, escolaridade e renda familiar. Ambos os modelos apresentaram ajustes satisfatórios aos dados.

O primeiro estudo evidenciou que os brasileiros residentes na região Centro-Oeste do Brasil apresentaram maiores chances de se tornarem empreendedores em relação aos residentes nas demais regiões. Quanto ao gênero, observou-se que os homens têm maiores chances de se tornarem empreendedores estabelecidos, enquanto as mulheres têm maiores chances de se tornarem empreendedores iniciais (iniciar um novo negócio). Tal resultado diferiu do da pesquisa realizada com empreendedores no norte do estado do Paraná por Gimenez et al. (2000), a qual indicou semelhanças no nível de atividade empreendedora entre os gêneros. Esta diferença se deve, provavelmente, às mudanças acentuadas que vêm ocorrendo desde então no comportamento do gênero feminino.

Ainda no primeiro estudo, observou-se que os indivíduos mais jovens (faixa de 18-33 anos) apresentaram maiores chances de se tornarem empreendedores iniciais, diferente dos resultados encontrados por Hart e O'Reilly (2006), em que indivíduos na faixa etária de 35-44 anos, residentes no Reino Unido e Irlanda do Norte, apresentaram maiores chances de se tornarem empreendedores iniciais. No

Brasil, os brasileiros na faixa de 34-48 anos apresentaram maiores chances de serem empreendedores estabelecidos.

Quanto à renda familiar, foi observada uma chance maior de serem encontrados tanto empreendedores iniciais quanto estabelecidos entre os com faixa salarial maior do que três salários mínimos, o que parece razoável tendo em vista que, em geral, a faixa salarial deva ser inferior entre os não empreendedores.

Quanto à escolaridade, observou-se uma chance maior de serem encontrados indivíduos estabelecidos entre os com menor escolaridade (abaixo do segundo grau). Duas possíveis justificativas para este fato são: 1ª) entre os empreendedores estabelecidos, há uma quantidade maior de indivíduos na faixa etária mais alta (conforme citado acima), indicando falta de disposição ou necessidade de se especializarem mais em termos de escolaridade e, 2ª) indivíduos mais jovens são maioria entre os empreendedores iniciais e, devido às exigências do mercado de trabalho atual, estes necessitam de nível de escolaridade mais alto dado a competição.

No segundo estudo constatou-se, assim como no primeiro estudo, que no ano de 2012 os brasileiros da Região Centro-Oeste apresentaram, sobre os das demais regiões, maiores chances de começar um novo negócio motivados pela oportunidade, com destaque da Região Centro-Oeste sobre a Região Norte, cuja chance foi de 3,5 vezes.

Na relação entre os gêneros foram obtidas as mesmas conclusões descritas por Bholá et al. (2006), que nos países da união europeia e Estados Unidos os homens são mais propensos a iniciar um negócio por oportunidade. Relacionando este fato com o primeiro estudo, parece sugestivo que um negócio iniciado por oportunidade tem grandes chances de se consolidar, visto que os homens apresentaram maiores chances de se tornarem empreendedores estabelecidos.

Ainda no segundo estudo, quando se analisou a renda familiar, pôde-se observar maiores chances de serem encontrados empreendedores motivados pela oportunidade entre os com renda acima de três salários mínimos, o que parece coerente, pois quanto maior a renda do indivíduo menor a necessidade de busca por ganhos extras e, em consequência, maiores serão as chances de vislumbrar novas oportunidades.

Os resultados encontrados indicaram, ainda, que brasileiros com nível mais elevado de escolaridade tendem a apresentar maiores chances de iniciar um empreendimento motivados por oportunidade do que por necessidade.

De modo geral, este estudo sugere a necessidade de esforços de todas as organizações públicas e privadas brasileiras, a fim de promover e dar apoio ao empreendedorismo. Merecendo um maior destaque o público feminino e o público jovem, como visto são à maioria dentre os empreendedores iniciais, no entanto, tiveram poucas chances frente aos públicos de outras características para se tornarem empreendedores estabelecidos.

REFERÊNCIAS

BHOLA, R.; VERHEUL, I.; THURIK, R.; GRILO, I. Explaining engagement levels of opportunity and necessity entrepreneurs, **SCALES: Scientific Analysis of Entrepreneurship and SMSs**, 2006. Disponível em: <http://repub.eur.nl/pub/9705/EIM20070208003_H200610.pdf>. Acesso em: 27 mai. 2014.

BLAND J. M.; ALTMAN D. G. The odds ratio. **British Medical Journal**, 2000. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1127651/>>. Acesso em: 23 mai. 2014.

BRITO, F.; WEVER, L. **Empreendedores Brasileiros**. 4.ed. Rio de Janeiro: Negócios, 2003. 246 p.

DORNELAS, J. C. A. **Empreendedorismo transformando ideias em negócios**. 3. ed. Rio de Janeiro: Elsevier, 2008. 231 p.

ENDEAVOR BRASIL. **Empreendedores Brasileiros: perfis e percepções 2013**. Disponível em: <http://www.endeavor.org.br/pesquisas/empreendedores_brasileiros_perfis_percepcoes_relatorio_completo.pdf>. Acesso em: 12 abr. 2014.

FILION, L. J. Empreendedorismo: empreendedores e proprietário-gerente de pequenos negócios. **Revista de Administração**, v. 34, n. 2, p. 05-28, 1999.

GIMENEZ, F. A. P. et al. Gênero e empreendedorismo: um estudo comparativo no Paraná, **Anais do I EGEPE**, v.1, p. 144-152, 2000.

GIOLO S. R. **Introdução à Análise de Dados Categóricos com Aplicações**. Material Didático, 2013. 180 p. Não publicado.

GEM - GLOBAL ENTREPRENEURSHIP MONITOR. **Empreendedorismo no Brasil 2012 - Global Entrepreneurship Monitor**. 1.ed. Instituto Brasileiro de Qualidade e Produtividade no Paraná, 2012. 164 p.

HART, M.; O'REILLY, M. **Global Entrepreneurship Monitor 2005**. 1. ed. Kingston University e ERINI, 2006. 100 p.

HOSMER JR, D. W.; LEMESHOW, S. **Applied Logistic Regression**. New York: John Wiley & Sons, 3.ed., 1989. 369 p.

McCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**, London: Chapman and Hall, 1989.

PAULA, G. A. **Modelo de Regressão com apoio computacional**. São Paulo: Instituto de Matemática e Estatística, 2013. 428 p. Disponível em: <http://www.ime.usp.br/~giapaula/texto_2013.pdf>. Acesso em: 26 abr. 2014.

PAULINO, C. D. E; SINGER, J. M. **Análise de dados categorizados**. 1.ed. São Paulo: Edgard Blücher, 2006. 648 p.

R DEVELOPMENT CORE TEAM. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <URL <http://www.R-project.org/>>.

SAKAMOTO, Y., ISHIGURO, M., AND KITAGAWA G. **Akaike Information Criterion Statistics**. 1.ed. D. Reidel Publishing Company, 1986. 290 p.

SAY, J. B. **Treatise on political economy: on the production, distribution and consumption of wealth**. 6. ed. New York: A. M. Kelley Publishers, 1971, 492 p. Disponível em: <<https://mises.org/books/politicaecon.pdf>>. Acesso em: 25 Mai. 2014.

SCHUMPETER, J. **The Theory of Economic Development**. 5.ed. Cambridge: Harvard University Press, 1934. 433 p.

THURIK, A. R.; CAREE, M. A.; VAN STEL, A. J.; AUDRETSCH, D. B. Does self-employment reduce unemployment? **Journal of Business Venturing**, v. 23, n. 6, p.

673-686, 2008. Disponível em <<http://ideas.repec.org/a/eee/jbvent/v23y2008i6p673-686.html>>. Acesso em: 20 fev. 2014.

VAN STEL, A. J., CARREE, M. A.; THURIK, A. R. The effect of entrepreneurial activity on national economic growth. **Small Business Economics**, v. 24, n. 3, p. 311-321, 2005. Disponível em: <<http://ideas.repec.org/p/esi/egpdis/2005-04.html>>. Acesso em: 10 fev. 2014.

APÊNDICES

APÊNDICE A.1 – Comandos R para ajuste do modelo de Logitos Generalizados

```

# Pacotes utilizados
require(VGAM)
library(lattice)
require(ggplot2)
require(GGally)
# Dados
primeiro_estudo <- read.delim("dados1.txt", header=TRUE)
attach(primeiro_estudo)
str(primeiro_estudo)
summary(primeiro_estudo)
# Modelo Nulo
m0 <- vglm(cbind(iniciais,estabelecidos,N_.empreendedor)~1,family=multinomial,
           link='logit',data=primeiro_estudo)
m0
summary(m0)
Nulo<- cbind(m0@df.residual,m0@criterion$deviance)
# Modelo com Região
m1 <- vglm(cbind(iniciais,estabelecidos,N_.empreendedor)~factor(regiao),
           family=multinomial,link='logit',data=primeiro_estudo)
m1
summary(m1)
mod1<- cbind(m1 @df.residual,m1 @criterion$deviance)
# Modelo com Região e Gênero
m2 <- vglm(cbind(iniciais,estabelecidos,N_.empreendedor)~factor(regiao)+
           factor(genero),family=multinomial,link='logit',data=primeiro_estudo)
m2
summary(m2)
mod2<- cbind(m2 @df.residual,m2@criterion$deviance)
# Modelo com Região, Gênero e Faixa etária
m3 <- vglm(cbind(iniciais,estabelecidos,N_.empreendedor)~factor(regiao)+
           factor(genero)+factor(faixa_etaria),family=multinomial,link='logit',
           data=primeiro_estudo)
m3
summary(m3)
mod3<- cbind(m3 @df.residual,m3@criterion$deviance)
# Modelo com Região, Gênero, Faixa etária e Renda
m4 <- vglm(cbind(iniciais,estabelecidos,N_.empreendedor)~factor(regiao)+
           factor(genero)+factor(faixa_etaria)+factor(faixa_renda),family=multinomial,
           link='logit',data=primeiro_estudo)
m4
summary(m4)
mod4<- cbind(m4 @df.residual,m4@criterion$deviance)
# Modelo com Região, Gênero, Faixa etária, Renda e Escolaridade
m5 <- vglm(cbind(iniciais,estabelecidos,N_.empreendedor)~factor(regiao)+
           factor(genero)+factor(faixa_etaria)+factor(faixa_renda)+factor(escolaridade),
           family=multinomial,link='logit',data=primeiro_estudo)
m5
summary(m5)

```

```

mod5<- cbind(m5@df.residual,m5@criterion$deviance)
# Modelo com Região, Gênero, Faixa etária, Renda e Escolaridade INVERSA DE Y
m6 <- vglm(cbind(N_.empreendedor,iniciais,estabelecidos)~factor(regiao)+
          factor(genero)+factor(faixa_etaria)+factor(faixa_renda)+factor(escolaridade),
          family=multinomial,link='logit',data=primeiro_estudo)
m6
# AIC
AIC_m0 <-AIC(m0)
AIC_m1 <-AIC(m1)
AIC_m2 <-AIC(m2)
AIC_m3 <-AIC(m3)
AIC_m4 <-AIC(m4)
AIC_m5 <-AIC(m5)
Nulo<-c(Nulo,AIC_m0)
mod1<-c(mod1,AIC_m1)
mod2<-c(mod2,AIC_m2)
mod3<-c(mod3,AIC_m3)
mod4<-c(mod4,AIC_m4)
mod5<-c(mod5,AIC_m5)
tab<-cbind(Nulo,mod1,mod2,mod3,mod4,mod5)
rownames(tab)<-c("gl", "Deviance", "AIC")
tab<-t(tab)
write.table(tab,"tab.csv",sep=";",dec="," )
# P-Valor para os Coeficientes Logitos 1 , 2 e 3
ctable <- coef(summary(m5))
ctable
pvals <- 2 * pt(abs(ctable[, "z value"]), df.residual(m5), lower.tail = FALSE)
pvals
pvalor12<- cbind(ctable, pvals)
write.table(pvalor12,"pvalor12.csv",sep=";",dec="," )
ctable3 <- coef(summary(m6))
ctable3
pvals3 <- 2 * pt(abs(ctable3[, "z value"]), df.residual(m6), lower.tail = FALSE)
pvals3
pvalor3<- cbind(ctable3, pvals3)
write.table(pvalor3,"pvalor3.csv",sep=";",dec="," )
# Resíduos
# Diferenças Obs - Pred
obs<-m5@y
pred<-fitted(m5)
t1<-(obs[,1] - pred[,1]) #Iniciais
t2<-(obs[,2] - pred[,2]) #Estabelecidos
t3<-(obs[,3] - pred[,3]) #Não Empreendedores
par(mfrow=c(1,3))
plot(t1, ylim=c(-1,1),ylab="Observados - Preditos",xlab="Registros",
     main = "INICIAIS: Valores Observados vs Preditos",pch=19)
abline(0,0,lty=3)
plot(t2, ylim=c(-1,1),ylab="Observados - Preditos",xlab="Registros",
     main = "ESTABELECIDOS: Valores Observados vs Preditos",pch=19 )
abline(0,0,lty=3)

```

```

plot(t3, ylim=c(-1,1),ylab="Observados - Preditos",xlab="Registros",
     main = "NÃO EMPREENDEDORES: Valores Observados vs Preditos", pch=19)
abline(0,0,lty=3)
# Brutos Observados - Preditos
par(mfrow=c(1,3))
plot(residuals(m5)[,1], pch=16, ylim=c(-4,4), ylab="Resíduos Brutos",
     xlab="Registros")
abline(h=0, lty=3); title("Iniciais | Não Empreendedores")
plot(residuals(m5)[,2], pch=16, ylim=c(-4,4), ylab="Resíduos Brutos",
     xlab="Registros")
abline(h=0, lty=3); title("Estabelecidos | Não Empreendedores")
plot(residuals(m6)[,1], pch=16, ylim=c(-4,4), ylab="Resíduos Brutos",
     xlab="Registros")
abline(h=0, lty=3); title("Iniciais | Estabelecidos")
# Pearson
par(mfrow=c(1,3))
plot(residuals(m5,'pearson')[,1], pch=16, ylim=c(-4,4), ylab="Resíduos
     Pearson",xlab="Registros")
abline(h=0, lty=3); title("Iniciais | Não Empreendedores")
plot(residuals(m5,'pearson')[,2], pch=16, ylim=c(-4,4), ylab="Resíduos
     Pearson",xlab="Registros")
abline(h=0, lty=3); title("Estabelecidos | Não Empreendedores")
plot(residuals(m6,'pearson')[,1], pch=16, ylim=c(-4,4), ylab="Resíduos
     Pearson",xlab="Registros")
abline(h=0, lty=3); title("Iniciais | Estabelecidos")
# Intervalo de Confiança
# Erro Padrão
se <- sqrt(diag(vcov(m5)))
se
# Covariâncias
cov <- vcov(m5)
cov
b <- coef(m5)
se <- sqrt(diag(vcov(m5)))
se
cbind(LL = b - qnorm(0.975) * se, UL = b + qnorm(0.975) * se)
qnorm(0.975)

```

APÊNDICE A.2 – Comandos R para ajuste do modelo para Dados Binários

```

# Pacotes utilizados
require(VGAM)
library(lattice)
require(ggplot2)
require(GGally)
# Dados
segundo_estudo <-
read.delim("C:/Users/Adriano/Dropbox/TCC_Adriano_Denis_2014/dados
categorizados/segundo_estudo.txt")
attach(segundo_estudo)
# Modelo Nulo
modelo00<-glm(cbind(oport,nec)~1,
              family=binomial(link="logit"),data=segundo_estudo)
# Modelo com Região
modelo01 <- glm(cbind(oport,nec)~factor(regiao),
               family=binomial(link="logit"),data=segundo_estudo)
# Modelo com Região e Gênero
modelo02 <-glm(cbind(oport,nec)~factor(regiao)+factor(genero),
               family=binomial(link="logit"),data=segundo_estudo)
# Modelo com Região, Gênero e Faixa etária
modelo03 <-
glm(cbind(oport,nec)~factor(regiao)+factor(genero)+factor(faixa_etaria),
    family=binomial(link="logit"),data=segundo_estudo)
# Modelo com Região, Gênero, Faixa etária e Renda
modelo04 <-
glm(cbind(oport,nec)~factor(regiao)+factor(genero)+factor(faixa_etaria)+factor(faixa_
renda),
    family=binomial(link="logit"),data=segundo_estudo)
# Modelo com Região, Gênero, Faixa etária, Renda e Escolaridade
modelo05 <-
glm(cbind(oport,nec)~factor(regiao)+factor(genero)+factor(faixa_etaria)+factor(faixa_
renda)+factor(escolaridade),
    family=binomial(link="logit"),data=segundo_estudo)
# AIC
AIC_m0 <-AIC(modelo00)
AIC_m1 <-AIC(modelo01)
AIC_m2 <-AIC(modelo02)
AIC_m3 <-AIC(modelo03)
AIC_m4 <-AIC(modelo04)
AIC_m5 <-AIC(modelo05)
Nulo<-c(modelo00,AIC_m0)
mod1<-c(modelo01,AIC_m1)
mod2<-c(modelo02,AIC_m2)
mod3<-c(modelo03,AIC_m3)
mod4<-c(modelo04,AIC_m4)
mod5<-c(modelo05,AIC_m5)
tab<-cbind(Nulo,mod1,mod2,mod3,mod4,mod5)
rownames(tab)<-c("gl", "Deviance", "AIC")

```

```

tab<-t(tab)
write.table(tab,"tab.csv",sep=";",dec=","")
# P-Valor para os Coeficientes
anova(modelo05,test="Chisq")
# Resíduos
# Diferenças Obs - Pred
plot(ajust1$y-ajust1$fitted.values, ylim=c(-1,1), ylab="obs-pred")
abline(0,0,lty=3)
# Resíduos pearson e deviance
dev<-residuals(modelo05,type='deviance')
rpears<-residuals(modelo05,type='pearson')
par(mfrow=c(1,2))
plot(rpears,pch=16, ylab="Resíduos de Pearson",ylim=(-3,3))
plot(dev,pch=16,ylab="Resíduos deviance",ylim=(-3,3))
# Intervalo de Confiança
# Erro Padrão
se <- sqrt(diag(vcov(modelo05))
se
# Covariâncias
cov <- vcov(modelo05)
cov
b <- coef(modelo05)
se <- sqrt(diag(vcov(m5)))
se
cbind(LL = b - qnorm(0.975) * se, UL = b + qnorm(0.975) * se)
qnorm(0.975)
# Curva ROC
segundo_estudo1 <-
read.delim("C:/Users/Adriano/Dropbox/TCC_Adriano_Denis_2014/dados
categorizados/segundo_estudo1.txt")
attach(segundo_estudo1)
# Segundo_estudo1 = arquivo com 1 indivíduo por linha
require(Epi)
ROC(form=oport~regiao+genero+faixa_etaria+faixa_renda+escolaridade,family=bino
mial(link="logit")
,plot="ROC")

```