



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 DO BRASIL

Projeto de Pesquisa apresentado à disciplina Laboratório de Estatística do Curso de Graduação em Estatística da Universidade Federal do Paraná, como requisito para elaboração do Trabalho de Conclusão de Curso.

Orientadora: Profa. Dra. Suely Ruiz Giolo

CURITIBA
2014

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	3
2.1	OBJETIVOS GERAIS	3
2.2	OBJETIVOS ESPECÍFICOS	3
3	MATERIAIS E MÉTODOS	4
3.1	MATERIAIS	4
3.1.1	Conjunto de Dados	4
3.1.1	Recursos Computacionais	5
3.2	MÉTODOS	5
3.2.1	Análise Descritiva	6
3.2.2	Modelo de Logitos Generalizados	6
3.2.3	Modelo de Regressão para dados Binários	7
4	CRONOGRAMA DE ATIVIDADES	9
	REFERÊNCIAS	10

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 e 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 2002, o empreendedorismo está sendo monitorado pelo projeto *Global Monitor Entrepreneurship* (GEM), desenvolvido pelo Instituto Brasileiro de Qualidade e Produtividade (IBQP), localizado em Curitiba, com o apoio do Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE). Este projeto tem por objetivo apresentar informações do empreendimento e do perfil do empreendedor, monitorando a dinâmica do empreendedorismo no mundo.

De acordo com o GEM, o Empreendedorismo é qualquer tentativa de criação de um novo negócio ou novo empreendimento, como, por exemplo, uma atividade autônoma, uma nova empresa, ou a expansão de empreendimento existente, por um indivíduo, grupos de indivíduos ou por empresas já estabelecidas (ANDREASSI *et al.*, 2012).

O propósito deste projeto é investigar as influências sociodemográficas dos empreendedores brasileiros utilizando informações do projeto *Global Monitor Entrepreneurship* (GEM), a mais abrangente pesquisa sobre empreendedorismo no Brasil, com o intuito de entender o perfil dos empreendedores brasileiros. Primeiramente, serão estudados os indivíduos classificados como: empreendedor inicial (que administra um novo negócio do qual é proprietário por menos de 42 meses), empreendedor estabelecido (que administra um novo negócio do qual é

proprietário por mais de 42 meses) e não empreendedor (não foi classificado como empreendedor na pesquisa). Uma segunda análise será estudar o perfil do empreendedor inicial segundo a motivação¹, ou seja, empreendedor motivado por oportunidade (indivíduo que, mesmo possuindo alternativa de emprego e renda, está envolvido com o empreendedorismo) e empreendedor motivado pela necessidade (indivíduo envolvido com empreendedorismo por não ter outra opção de emprego).

¹ A questão sobre os motivos que levaram o empreendedor a abrir um negócio é respondida somente por empreendedores classificados como iniciais.

2 OBJETIVOS

2.1 OBJETIVOS GERAIS

Os objetivos deste trabalho são: a) relacionar os diferentes estágios de empreendimentos (não empreendedores, empreendedores iniciais e empreendedores estabelecidos) com as características sociodemográficas dos indivíduos, considerando toda a amostra e, b) estudar as mesmas características sociodemográficas com as motivações (oportunidade ou necessidade) que levaram os empreendedores iniciais a abrir um novo negócio, neste caso, restringindo a amostra a apenas esta categoria. O estudo irá se concentrar em variáveis sociodemográficas, tais como: região, gênero, faixa etária, escolaridade e renda.

2.2 OBJETIVOS ESPECÍFICOS

- Proceder a uma análise descritiva dos dados a fim de delinear o perfil dos empreendedores segundo o estágio dos empreendimentos e a motivação;
- Complementar os resultados do relatório nacional da pesquisa GEM com análise de dados categorizados;
- Fazer uso do modelo de logitos generalizados a fim de estudar a associação dos estágios (não empreendedores, empreendedores iniciais e empreendedores estabelecidos) dos empreendimentos com as variáveis sociodemográficas;
- Utilizar modelos de regressão para dados binários, a fim de associar a motivação (oportunidade ou necessidade) dos empreendedores iniciais conforme as variáveis sociodemográficas;
- Após a realização das análises supracitadas, apresentar conclusões a respeito do perfil dos empreendedores em relação ao estágio do empreendimento e à motivação dos empreendedores, e fornecer os resultados para instituições que fomentam a atividade empreendedora.

3 MATERIAIS E MÉTODOS

3.1 MATERIAIS

Os materiais utilizados para o desenvolvimento do presente estudo são compostos pelo banco de dados, fornecido pelo Instituto Brasileiro de Qualidade e Produtividade (IBQP), e por recursos computacionais, os quais 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*² e o *Babson College*³, o primeiro estudo abrangeu 10 países e, desde então, cerca de 100 países já participaram do GEM. Os objetivos principais da pesquisa 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, a pesquisa é conduzida pelo Instituto Brasileiro de Qualidade e Produtividade – IBQP, o qual entrevistou 10.000 indivíduos entre 18 e 64 anos de idade, coletando informações sobre suas características, mentalidade empreendedora e características do empreendimento (apenas empreendedores).

O banco de dados, fornecido pelo IBQP, contempla os indivíduos da pesquisa realizada no ano de 2012. O estudo será realizado em duas etapas. A primeira, considerando toda a amostra (10.000 indivíduos), utilizará as informações

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

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

sociodemográficas (região, gênero, faixa etária, nível de escolaridade e renda) e relaciona estas com a classificação do indivíduo referente ao estágio do empreendedorismo (não empreendedores, empreendedores iniciais e empreendedores estabelecidos). A segunda, considerando apenas os empreendedores iniciais (1.549 indivíduos), utilizará as mesmas informações sociodemográficas e relacionará estas com a motivação (oportunidade ou necessidade) que os levou a abrir um novo negócio.

3.1.2 Recursos Computacionais

O *software* R (R DEVELOPMENT CORE TEAM, 2013), versão 3.0.1, será utilizado na análise estatística dos dados.

3.2 MÉTODOS

Neste estudo, serão utilizados os métodos de análise de dados categorizados com o intuito de interpretar ou relacionar as informações relevantes contidas nos dados discretos oriundos da pesquisa.

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 com o fato de estar presente nos mais variados campos da ciência (PAULINO e SINGER, 2006).

As variáveis no estudo seguem duas classificações:

- Variáveis explicativas, ou ainda covariáveis: são aquelas que possibilitam o entendimento do fator de estudo;
- Variáveis resposta: são as que descrevem as particularidades de cada unidade amostral ou categoria, por meio de um modelo probabilístico.

A seguir são descritas brevemente as metodologias estatísticas que se pretende utilizar para a execução deste trabalho.

3.2.1 – Análise Descritiva

Para conhecer a distribuição do empreendedorismo no Brasil e o perfil dos empreendedores, serão construídas, inicialmente, tabelas de frequências e gráficos, assim como calculadas correlações entre as variáveis de interesse, para uma avaliação abrangente dos resultados e as premissas a serem verificadas e discutidas.

3.2.2 – Modelo de Logitos Generalizados

A primeira análise abordada será relacionar os diferentes estágios de empreendimentos com as características sociodemográficas. O estágio do empreendimento é uma variável de natureza nominal, ou seja, não numérica e não possui hierarquia, além disso, é uma variável politômica, com mais de duas categorias. O modelo sugerido para o estudo, nesta situação, é o de logitos generalizados.

Considerando um exemplo com duas variáveis explicativas e uma variável resposta, composta por três categorias, usaremos a seguinte codificação:

- h = categorias da 1ª variável explicativa;
- i = categorias da 2ª variável explicativa;
- k = 1 ... (r-1) categorias da variável resposta.

Assim, a probabilidade associada a cada uma das categorias da variável resposta ficará denotada por:

$$\pi_{hik}$$

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

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

Desse modo os logitos generalizados associados às categorias da variável resposta são:

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

sendo que o logito para a terceira categoria é obtido pela razão dos dois primeiros.

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

$$\text{logit}_{hik} = \beta_{0k} + \beta'_{k}X_{hi}.$$

Percebe-se que para cada logito do modelo existem diferentes interceptos (β_{0k}) e parâmetros de regressão ($\beta'_{k}X_{hi}$), ou seja, vários parâmetros serão estimados para cada subpopulação. Desta maneira, precisa-se de um tamanho amostral compatível para alocar tais logitos e seus graus de liberdade (GIOLO, 2014).

Em relação à seleção das covariáveis, pretende-se utilizar o método *stepwise do tipo* seleção regressiva (*backward selection*). Tal método consiste na retirada sequencial de termos do modelo até não se conseguir uma melhora significativa no aperfeiçoamento do ajuste (PAULINO e SINGER, 2006).

3.2.3 – Modelo de Regressão para Dados Binários

A segunda análise pretende confrontar a motivação dos empreendedores iniciais na abertura de um negócio, seja por oportunidade ou necessidade, contra as mesmas características sociodemográficas relatadas na primeira abordagem. Neste caso, a variável resposta, motivação, é binária ou dicotômica, isto é, quando a variável resposta assume apenas dois valores possíveis, por exemplo, $y_i = 0$ e $y_i = 1$, denominados "fracasso" e "sucesso", respectivamente. Neste cenário, o uso de um modelo de regressão para dados binários é o mais recomendado, como, por exemplo, o modelo de regressão logística.

O modelo de regressão logística é frequentemente usado em estudos epidemiológicos para modelar a relação entre uma variável binária e um conjunto de variáveis de previsão. Tradicionalmente, a regressão logística assume que as observações representam uma amostra aleatória de uma população (ou seja, independentes e identicamente distribuídos). É semelhante ao modelo de regressão linear. No entanto, no modelo logístico a variável resposta é binária (HOSMER e

LEMESHOW, 1989), com $\theta(\mathbf{x}) = P(Y = 1 | \mathbf{x})$ expressa por:

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

Caso este modelo não apresente ajuste satisfatório aos dados, pretende-se considerar outros modelos com funções de ligação alternativas como, por exemplo, as funções: *probit*, *cauchy* e *complementar log log*.

Em relação à seleção das covariáveis, pretende-se utilizar o mesmo método citado na primeira abordagem.

4 CRONOGRAMA DE ATIVIDADES

ATIVIDADES	Fev/2014	Mar/2014	Abr/2014	Mai/2014	Jun/2014
1 Projeto de Pesquisa					
Definição do tema de estudo	■				
Definição do conjunto de dados e dos métodos estatísticos	■				
Elaboração e entrega do projeto de pesquisa ao orientador	■	■			
2 Elaboração do Trabalho de Conclusão de Curso					
Revisão de literatura sobre o tema	■	■			
Análise dos dados e discussão dos resultados obtidos		■	■	■	
Redação do trabalho de conclusão de curso			■	■	
Leitura do trabalho pelo orientador e correções				■	■
Entrega do trabalho redigido aos membros da banca					■
3 Defesa do Trabalho de Conclusão de Curso					
Preparação e apresentação do trabalho de conclusão de curso					■
4 Elaboração da Versão Final do Trabalho de Conclusão de Curso					
Elaboração da versão final do TCC					■
Entrega da versão final do trabalho ao orientador					■

REFERÊNCIAS

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

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

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

ANDREASSI, T., et al. **Global Entrepreneurship Monitor Empreendedorismo no Brasil 2012**, Instituto Brasileiro de Qualidade e Produtividade no Paraná, 1.ed., 164p, 2012.

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/>.

PAULINO, D.D.M., SINGER, J.M., **Análise de Dados categorizados**, São Paulo: Blucher, 1.ed., 629p, 2006.

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

GIOLO S.R., **Introdução à Análise de Dados Categóricos com Aplicações**. Disponível em: <<http://www.ufpr.br/~Giolo/CE073/>>. Acesso em 22 fev. 2014, (Material da disciplina de Análise de Dados Categóricos, Curso de Estatística, Universidade Federal do Paraná), 180p, 2014.