Universidade de São Paulo

Faculdade de Economia, Administração e Contabilidade

VARIÁVEIS RELEVANTES PARA AS EMPRESAS DE ALTO CRESCIMENTO NO
BRASIL

Carlos Roberto Francisco Bara

Orientador: Prof. Dr. Guilherme Ary Plonski

Versão Corrigida

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

São Paulo

2018

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Moacir de Miranda Oliveira Júnior
Chefe do Departamento de Administração

Prof. Dr. Moacir de Miranda Oliveira Júnior
Coordenador do Programa de Pós-Graduação em Administração

Carlos Roberto Francisco Bara

VARIÁVEIS RELEVANTES PARA AS EMPRESAS DE ALTO CRESCIMENTO NO
BRASIL

Tese apresentada ao Programa de Pós-Graduação em Administração do Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Guilherme Ary Plonski

Versão Corrigida

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

São Paulo

2018

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Bara, Carlos Roberto Francisco

Variáveis relevantes para as empresas de alto crescimento no Brasil /
Carlos Roberto Francisco Bara. – São Paulo, 2018.
149 p.

Tese (Doutorado) – Universidade de São Paulo, 2018.

Orientador: Guilherme Ary Plonski.

1. Empreendedorismo 2. Empresas de alto crescimento 3. Gazelas
4. Regressão logística I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 658.421

Para minha esposa Silvia
e meus filhos Ricardo, Felipe e Isabella,
pelo constante apoio e compreensão durante
essa jornada muito especial.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus e a São Judas Tadeu por terem me dado saúde e discernimento para fazer as escolhas que me trouxeram até aqui.

Meu agradecimento muito especial ao Prof. Dr. Guilherme Ary Plonski, que me guiou do ponto de vista acadêmico e pessoal, com seu sólido conhecimento da área; mais que um orientador, Prof. Ary foi capaz de me manter sempre focado e confiante, dirigindo-me adequadas palavras de encorajamento e incentivo, e não deixando espaço para qualquer tipo de esmorecimento durante as desafiadoras etapas de meu doutorado.

Ao Prof. Dr. Marcelo Hiroshi Nakagawa, meu colega de Insper, membro das bancas de qualificação e de defesa da tese, muito obrigado. Marcelo foi instrumental para definir certas escolhas acadêmicas e indicar frutíferos contatos de sua rede de relacionamento, que muito contribuíram para a elaboração da presente tese.

Agradeço ao Prof. Dr. Marcelo Caldeira Pedroso, membro das bancas de qualificação e de defesa da tese, pelos valiosos e construtivos comentários e recomendações. Ao Prof. Tales Andreassi, membro da banca de defesa da tese, agradeço as sólidas sugestões de melhoria desta versão corrigida.

Ao Prof. Dr. Luiz Paulo Fávero, que me mostrou como refinar o formulário de pesquisa e conduzir a análise de dados, muito obrigado.

A Endeavor e a Pedro Lipkin Peçanha Rosa, responsável pela área de Pesquisa e Mobilização da instituição, meu especial agradecimento. A grande experiência da Endeavor na área de empreendedorismo, materializada em sólidas recomendações para a coleta de dados, foi fundamental para a condução desta pesquisa. Meu muito obrigado também a Felipe Matos por todo o apoio recebido da Startup Farm.

Aos meus colegas do Insper, professores doutores Guilherme Fowler de Almeida Monteiro, Danny Pimentel Claro, Silvio Abrahão Laban, Guy Cliquet do Amaral Jr. e Giuliana Isabella muito obrigado pelos aconselhamentos acadêmicos, pessoais e permanente incentivo.

Agradecimento especial aos professores doutores José Afonso Mazzon, da FEA e Claudia Pavani, pesquisadora do Núcleo de Política e Gestão Tecnológica (PGT/USP), pelo constante apoio acadêmico e incentivo. A Sonia Vitorino, assistente administrativa do PGT/USP, pelo essencial suporte operacional, muito obrigado.

Enfim, agradeço a todos os que contribuíram com ideias, críticas e sugestões ou apenas como amigos dispostos a me ouvir, encorajar e a passar energia para que eu chegasse ao final dessa longa estrada.

RESUMO

Empreendedorismo tem sido objeto de incentivo no mundo e no Brasil, dada a sua significativa contribuição para o desenvolvimento econômico e social de uma nação. Observa-se que a maioria das empresas, existentes ou novas, evolui de forma lenta e gradual; no entanto, reduzida parcela apresenta um padrão diferente, com crescimento elevado em faturamento ou número de colaboradores: são as empresas de alto crescimento (EACs). Tais empresas são as responsáveis por grande parte da geração de empregos (Birch, 1981; Coad, Daunfeldt, Holzl, Johansson, & Nightingale, 2014; Henrekson & Johansson, 2010; OECD, 2010). A presente tese procurou identificar as variáveis que ajudam a explicar o desempenho das EACs no Brasil, classificadas conforme critério da OECD (2007). Foi conduzida uma pesquisa com 470 empresas brasileiras, que coletou mais de 30 variáveis preditoras categóricas ou métricas, utilizadas no modelo Regressão Logística. Foram identificadas algumas variáveis alinhadas com a literatura e outras menos intuitivas e documentadas. Comprovou-se o aumento da probabilidade de EACs quando se relacionavam com aceleradoras, recebiam premiações ou eram *spin-offs* de outras empresas. Em função das altas taxas de juros bancários e da cultura empreendedora no Brasil, surpreendeu o impacto positivo de empréstimos bancários e a percepção dos empreendedores sobre registrar marcas comerciais, bem como o impacto negativo da percepção sobre propaganda em mídia digital e doações de instituições de fomento, relacionadas às EACs. Análises adicionais com o subgrupo de EACs caracterizadas como gazelas foram feitas. Embora apresente limitações de *surveys* e outras, a tese confirmou parte dos resultados da literatura sobre empreendedorismo e identificou avenidas para futuras pesquisas.

Palavras-chave: Empreendedorismo, Empresas de Alto Crescimento, Gazelas, Regressão Logística.

ABSTRACT

Entrepreneurship has been object of encouragement in the world and in Brazil, given its significant contribution to the economic and social development of a nation. It is observed that the majority of companies, existing or new, are developing slowly and gradually; however, small share presents a different pattern, with high growth in sales or number of employees: they are the high growth firms (HGFs). These firms are responsible for a large part of job creation (Birch, 1981; Coad, Daunfeldt, Holzl, Johansson, & Nightingale, 2014, Henrekson & Johansson, 2010, OECD, 2010). This thesis aimed to identify the variables that help to explain the performance of HGFs in Brazil, according to OECD (2007) criterion. A survey with 470 Brazilian companies was conducted, collecting more than 30 categorical or metric predictor variables, used in the Logistic Regression model. Some identified variables were aligned to literature, but others less intuitive or documented. It was confirmed the increase in the probability of HGFs when they related to accelerators, received awards, or were spin-offs of other companies. As a consequence of the high banking interest rates and the entrepreneurship culture in Brazil, surprised the positive impact of bank loans and the entrepreneurs' perception of trademark registration, as well as the negative impact of perception on advertising in digital media and donations from development institutions, related to HGFs. Additional analyzes with the subgroup of HGFs characterized as gazelles were made. Although it presents limitations of surveys and others, the thesis confirmed part of the results of the literature on entrepreneurship and identified avenues for future researches.

Keywords: Entrepreneurship, High Growth Firms, Gazelles, Logistic Regression.

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Tema	21
1.2	Objetivos e Questão de Pesquisa	24
1.3	Relevância	25
1.4	Estrutura da Tese	27
2	REFERENCIAL TEÓRICO.....	29
2.1	Empreendedorismo e Inovação	29
2.2	Ecossistema de Empreendedorismo.....	33
2.3	Empresas de Alto Crescimento e Gazelas.....	40
2.4	Variáveis Preditoras de Desempenho de EACs.....	42
2.5	Hipóteses	44
3	MÉTODO DE PESQUISA	49
3.1	Desenho da Pesquisa e Principais Variáveis	49
3.2	Amostragem	51
3.3	Regressão Logística Binária.....	56
3.4	Análise Fatorial.....	62
3.5	Lista de Variáveis	65
4	RESULTADOS	75
4.1	Análises Descritivas.....	75
4.2	Resultados da Análise Fatorial	84
4.3	Resultados da Regressão Logística Binária	91
4.4	Gazelas	109
4.5	Resultados Adicionais.....	112
4.6	Comentários sobre a Questão de Pesquisa e Hipóteses	121
5	CONSIDERAÇÕES FINAIS	127
5.1	Conclusões Principais	127
5.2	Implicações Gerenciais	131
5.3	Limitações	132
5.4	Estudos Futuros	133
	REFERÊNCIAS BIBLIOGRÁFICAS.....	135
	APÊNDICE 1	141

LISTA DE TABELAS

Tabela 1 – Ambientes Relacionados a Empreendedorismo	38
Tabela 2 – Fontes de Capital Relacionadas a Empreendedorismo	39
Tabela 3 – Amostragem de Empresas por Entidade	55
Tabela 4 – Relação entre KMO e Adequação da Análise Fatorial	64
Tabela 5 – Variáveis de Crescimento de Empresas	66
Tabela 6 – Variáveis Predictoras Demográficas	71
Tabela 7 – Variáveis Predictoras de Percepção dos Empreendedores	72
Tabela 8 – Variáveis Predictoras de Relacionamento das Empresas	73
Tabela 9 – Variáveis Predictoras do Perfil dos Empreendedores	74
Tabela 10 – CAGRs das EACs e Não EACs.....	77
Tabela 11 – Número de Colaboradores Médio das EACs e Não EACs	79
Tabela 12 – Variáveis Predictoras Demográficas: Estatísticas Descritivas	79
Tabela 13 – Variáveis Predictoras de Aporte de Capital: Estatísticas Descritivas	80
Tabela 14 – Variáveis Predictoras de Relacionamento: Estatísticas Descritivas	80
Tabela 15 – ANOVA: Comparação de Médias de Percepção dos Empreendedores de EACs e Não EACs.....	82
Tabela 16 – Percepção dos Empreendedores sobre a Importância de Marcas, Capital de Investidores e Empréstimos Bancários para EACs e Não EACs: Médias e Desvios	84
Tabela 17 – Percepção dos Empreendedores: Matriz de Correlação de Pearson.....	85
Tabela 18 – Percepção dos Empreendedores: Análise Fatorial – Eigenvalues	86
Tabela 19 – Percepção dos Empreendedores: Análise Fatorial – Matriz de Cargas Fatoriais .	88
Tabela 20 – Percepção dos Empreendedores: Análise Fatorial – KMO.....	88
Tabela 21 – Percepção dos Empreendedores: Análise Fatorial – Eigenvalues Rotacionados .	89
Tabela 22 – Percepção dos Empreendedores: Análise Fatorial – Matriz de Cargas Fatoriais Rotacionadas.....	90
Tabela 23 – EACs e Percepção dos Empreendedores: Regressão Logística Binária – Coeficientes	92
Tabela 24 – EACs e Percepção dos Empreendedores: Regressão Logística Binária – Matriz de Classificação.....	93
Tabela 25 – Novas Variáveis Predictoras Originadas dos Dados Primários	99
Tabela 26 – EACs e Variáveis Categóricas: Regressão Logística Binária – Coeficientes	100
Tabela 27 – EACs e Variáveis Categóricas: Regressão Logística Binária – Matriz de Classificação.....	101

Tabela 28 – EACs, Percepção dos Empreendedores e Variáveis Categóricas: Regressão Logística Binária – Coeficientes.....	103
Tabela 29 – EACs, Percepção dos Empreendedores e Variáveis Categóricas: Regressão Logística Binária – Matriz de Classificação.....	104
Tabela 30 – Gazelas: Estatísticas Descritivas.....	111
Tabela 31 – CAGR 2013/2017 de Faturamento, Percepção dos Empreendedores e Variáveis Categóricas: Regressão Linear – Coeficientes	118

LISTA DE FIGURAS

Figura 1 – Modelo do Processo de Empreendedorismo Afetando o Crescimento da Economia	22
Figura 2 – Domínios do Ecossistema de Empreendedorismo	34
Figura 3 – Curva de Probabilidade do Modelo Regressão Logística Binária	59
Figura 4 – Distribuição de Densidade do CAGR 2013/2017 de Faturamento das Empresas ..	78
Figura 5 – Percepção dos Empreendedores sobre a Importância de Ferramentas de Gestão para EACs e Não EACs: Box Plot.....	81
Figura 6 – Percepção dos Empreendedores sobre a Importância de Marcas, Capital de Investidores e Empréstimos Bancários para EACs e Não EACs: Box Plot.....	83
Figura 7 – Curva <i>ROC</i> de EACs em função da Percepção dos Empreendedores	94
Figura 8 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Fazer Propaganda em Mídia Digital	96
Figura 9 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Registrar Marcas Comerciais	96
Figura 10 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Obter Empréstimos Bancários	97
Figura 11 – Curva <i>ROC</i> de EACs em função de Variáveis Categóricas	101
Figura 12 – Curva <i>ROC</i> de EACs em função da Percepção dos Empreendedores e de Variáveis Categóricas	104
Figura 13 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Obter Empréstimos Bancários e Relacionamento com Aceleradoras	106
Figura 14 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Obter Empréstimos Bancários e Recebimento de Premiação	107
Figura 15 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Receber Doações de Instituições de Fomento e Recebimento de Premiação	108

LISTA DE ABREVIATURAS E SIGLAS

ABVCAP: Associação Brasileira de Private Equity & Venture Capital

Anprotec: Associação Nacional de Entidades Promotoras de Empreendimentos Inovadores

BACON: Algoritmo Estatístico Blocked Adaptive Computationally Efficient Outlier Nominators

CAGR: Compound Annual Growth Rate

EAC: Empresa de Alto Crescimento

EAESP: Escola de Administração de São Paulo da Fundação Getulio Vargas

EBSCO: Business Source Complete

EIBT: Empresa Inovadora de Base Tecnológica (EIBT)

FGV: Fundação Getulio Vargas

GEM: Global Entrepreneurship Monitor

GVcenn: Centro de Empreendedorismo e Novos Negócios

GVcepe: Centro de Estudos em Private Equity e Venture Capital

IBGE: Instituto Brasileiro de Geografia e Estatística, do Ministério do Planejamento, Desenvolvimento e Gestão

KMO: Estatística Kaiser-Meyer-Olkin

MIT: Massachusetts Institute of Technology

NTBF: New Technology Based Firm

OECD: Organisation for Economic Co-operation and Development

PGT/USP: Núcleo de Política e Gestão Tecnológica, da Universidade de São Paulo

PME: Pequena e Média Empresa

RAIS: Relação Anual de Informações Sociais, do Ministério do Trabalho

ROC: Curva Estatística Receiver Operating Characteristics

SEBRAE: Serviço Brasileiro de Apoio às Micro e Pequenas Empresas

VIF: Estatística Variance Inflation Factor

1 INTRODUÇÃO

Este capítulo inicial é composto por quatro seções; na primeira seção é apresentado o tema que foi desenvolvido nesta tese; na seção 2 são detalhados seus objetivos e a questão de pesquisa; segue a seção 3 onde são feitas as considerações sobre a relevância da pesquisa realizada; o capítulo é finalizado com a seção 4 que detalha como a tese está estruturada.

1.1 Tema

O desenvolvimento econômico e social de um país está associado à melhoria da qualidade de vida de sua população, formando um círculo virtuoso que impacta diversos indicadores, dentre os quais a geração de empregos. Quando um país consegue criar postos de trabalho para a sua população economicamente ativa, conferindo autonomia aos indivíduos, soluciona um de seus principais problemas e consequentemente adquire melhores condições para se focalizar em outras demandas sociais.

Para gerar empregos de qualidade e em quantidade adequada, o fortalecimento da iniciativa privada tem se mostrado fundamental; segundo os dados do Cadastro Central de Empresas (CEMPRE) do IBGE (b 2017, p. 28) de 2015, as 4,6 milhões de empresas ativas foram responsáveis por empregar 40,2 milhões de pessoas, dos quais 33,6 milhões como assalariados e 6,6 milhões como sócios ou proprietários. Assim, seja pelo crescimento das empresas já existentes, como através da criação de novas, uma nação consegue estimular a oferta de postos de trabalho.

Isso explica em parte porque a grande maioria dos países busca disponibilizar ambientes favoráveis ao fortalecimento de empresas e novos negócios, dada sua contribuição para o desenvolvimento econômico e social e consequente criação de empregos. Nos países em desenvolvimento, como é o caso do Brasil, a oferta desses ambientes favoráveis assume relevância ainda maior, dada a necessidade premente de melhorar as condições de vida de parcelas expressivas de suas populações.

A relação entre a criação de ambientes favoráveis e o desenvolvimento de empresas tem sido muito estudada no contexto de Empreendedorismo, que é umas das forças que estimulam a geração de empregos, inovação e desenvolvimento econômico sadio e sustentável de comunidades (Wise e Valliere, 2014). A Figura 1 resume como o empreendedorismo pode

impactar o crescimento econômico nacional, segundo o modelo adotado pelo Global Entrepreneurship Monitor (GEM) desde 1999, proposto por Reynolds, Hay e Camp (1999) e reproduzido no relatório de 2014 (Singer et al., 2015, p. 19). O conceito de *Ecosistema de Empreendedorismo*, que aparece na figura, está associado aos mencionados ambientes favoráveis onde as empresas estão inseridas, em geral identificado a determinada área geográfica; esse conceito está detalhado na seção 2.2 desta tese.

Figura 1 – Modelo do Processo de Empreendedorismo Afetando o Crescimento da Economia

Fonte: Reynolds, P. D., Hay, M., & Camp, S. M., Global Entrepreneurship Monitor (1999, p. 10). Kauffman Center for Entrepreneurial Leadership at the Ewing Marion Kauffman Foundation. Livre tradução do autor.

Vale a pena ressaltar que um dos resultados finais do processo de empreendedorismo detalhado na Figura 1 é a geração de empregos, intrinsicamente relacionado ao crescimento econômico, ativando assim o mencionado círculo virtuoso que os países tanto almejam.

No entanto, não basta aos países fomentar o desenvolvimento de empresas, existentes ou novas; no relatório da reunião de Paris da Organisation for Economic Co-operation and Development (OECD), Mason e Brown (2014) defendem que é preciso aumentar o número de *empresas de alto crescimento* (EACs), pois são elas as responsáveis pela oferta da maior parte dos postos de trabalho. Embora as EACs representem uma parcela relativamente pequena das firmas, serão elas as grandes geradoras de empregos (Birch, 1981; Coad, Daunfeldt, Holzl, Johansson, & Nightingale, 2014; Henrekson & Johansson, 2010; OECD, 2010).

Assim, o foco de interesse acadêmico e gerencial deveria ir muito além do aumento do número de empresas ou da melhoria de desempenho, buscando entender de forma mais

profunda porque algumas firmas conseguem lograr alto crescimento, enquanto outras não (Monteiro, 2017).

Embora intuitivo, o conceito de EAC necessita uma definição mais precisa e a própria OECD (2007, p. 61) recomenda a seguinte:

“All enterprises with average annualised growth greater than 20% per annum, over a three year period should be considered a high-growth enterprise. Growth can be measured by the number of employees or by turnover.”

Portanto, para que uma empresa seja considerada de alto crescimento é necessário que apresente ao longo de um período predeterminado de três anos crescimento anual médio de 20% em faturamento ou em número de colaboradores. Essa métrica de crescimento anual médio é o chamado *Compound Annual Growth Rate* (CAGR), nesse caso aplicado a um período de três anos.

Para evitar distorções na identificação de EACs, a OECD (2007, p. 61) recomenda ainda que seja adotado o limite provisional mínimo de 10 colaboradores no ano inicial do período de crescimento avaliado, para ambos os indicadores, faturamento ou número de colaboradores. A OECD (2007) preconiza outros ajustes e critérios de exclusão que serão detalhados no capítulo 2 desta tese.

Além disso, a OECD (2007, p. 63), indica que as EACs com até cinco anos de idade sejam denominadas *gazelles* (gazelas), em analogia à rápida espécie de antílopes africanos. Embora a definição da OECD (2007) não seja a única encontrada na literatura acadêmica, uma vantagem em utilizá-la é a possibilidade de fazer comparações entre diversos países, razão pela qual foi adotada nesta tese.

Assim, estudar e entender porque algumas empresas, sejam existentes ou novas, conseguem apresentar resultados de alto crescimento de faturamento ou de número de colaboradores, que permitam enquadrá-las como EACs, é o objetivo central desta tese. Para tanto, procurou-se identificar quais variáveis ajudam a explicar o alto crescimento das empresas nacionais, quantificando a contribuição dos principais *drivers* que as caracterizam.

Neste contexto, a presente tese buscou aportar contribuições para a literatura acadêmica, ajudando a construir a teoria sobre EACs. Como benefício adicional, trazer subsídios à

elaboração de melhores políticas públicas, favorecendo o desenvolvimento econômico e social do Brasil.

1.2 Objetivos e Questão de Pesquisa

A presente tese tem um objetivo principal, foco da pesquisa conduzida, e alguns objetivos secundários, que embora coadjuvantes, também apresentam relativa importância. O objetivo principal refere-se a identificar as variáveis relevantes que contribuem para que algumas firmas possam alcançar resultados que as classifiquem como empresas de alto crescimento (EACs) no Brasil. Os objetivos secundários referem-se a avaliar um painel relevante de empresas, para aprofundar o conhecimento sobre como elas são estruturadas, a percepção dos empreendedores sobre ferramentas de gestão e especificamente identificar eventuais gazelas brasileiras; o objetivo principal e os objetivos secundários são a seguir detalhados.

O principal objetivo desta tese foi o de estudar quais são as variáveis relevantes ou principais *drivers* que fazem com que uma empresa brasileira apresente alto crescimento de faturamento ou do número de colaboradores, ao longo de um determinado período, que permitam caracterizá-la como de alto crescimento. O critério escolhido para classificar as firmas como EACs foi o da OECD (2007), transcrito na seção 1.1 desta tese: empresas que apresentem durante um período de três anos um crescimento médio em faturamento ou do número de colaboradores – CAGR – superior a 20% ao ano, que equivale a 72,8% entre o ano final *xx* e o ano inicial *xx-3*.

A partir de uma amostra representativa de empresas, das quais parte foi classificada como EACs e as restantes como não EACs, o passo seguinte foi identificar e avaliar o impacto de variáveis críticas que constituem os *drivers* para levá-las a pertencerem a cada um desses dois agrupamentos. Assim, a partir da amostra de firmas que passaram pelos filtros de exclusões recomendados pela OECD (2007), detalhados na seção 2.3, a presente tese se focou em seu objetivo principal, que foi identificar as EACs e analisar o impacto das variáveis que contribuíram para o seu desempenho; buscou assim avançar as fronteiras do conhecimento, procurando identificar quais são as variáveis relevantes para o alto crescimento das empresas no Brasil.

Como objetivos secundários, esta tese buscou conhecer, entender e descrever como as empresas estão sendo gerenciadas no Brasil, através da análise de um painel de firmas representativo desse universo. O levantamento quantitativo em que se baseou a presente

pesquisa permitiu a coleta de diversas variáveis das empresas sobre: aspectos demográficos, percepção dos empreendedores sobre as ferramentas de gestão mais utilizadas, apoios recebidos, relacionamento com aceleradoras, aportes de capital, dentre outras; tais características das empresas serviram de base para entender melhor a atividade empreendedora no Brasil.

O subgrupo de gazelas, parcela de EACs com até cinco anos de idade (OECD, 2007, p. 63), também foi objeto de estudos adicionais. Essas empresas são de particular interesse, uma vez que além de serem de alto crescimento são também novas, com idade entre três e cinco anos, o que as torna ainda mais valiosas como unidade de pesquisa.

Embora devam ser considerados objetivos secundários ao principal, espera-se que uma melhor descrição e compreensão das empresas brasileiras, através de uma amostragem representativa, tenha o potencial de agregar valor ao conhecimento acadêmico sobre a área de empreendedorismo no Brasil. Portanto, o objetivo principal e os objetivos secundários desta tese são complementares e juntos buscam aportar de forma sinérgica um melhor entendimento das empresas de alto crescimento (EACs) no Brasil.

Nesse contexto, a *questão de pesquisa* da presente tese é a seguinte:

Quais são as variáveis relevantes que ajudam a explicar o desempenho das empresas de alto crescimento (EACs) no Brasil?

1.3 Relevância

No Brasil existem numerosos estudos acadêmicos sobre empreendedorismo, que buscam identificar fatores e variáveis que impactam no desempenho das empresas. Entretanto a grande maioria de artigos publicados é embasada em pesquisas qualitativas, sobretudo utilizando o método de estudo de casos, como revelou pesquisa na base de artigos acadêmicos *Business Source Complete (EBSCO)*, conduzida pelo autor desta tese. Embora tais estudos sejam de grande relevância e aportem valiosa contribuição para o avanço da ciência e conhecimento em nosso país, o fato de serem calcados em técnicas de pesquisas qualitativas limitam suas abrangências.

Mais especificamente, em relação a empresas de alto crescimento (EACs), o número de publicações brasileiras é escasso; Monteiro (2017) publicou ampla revisão da literatura sobre EACs, usando metodologia de busca através das ferramentas *Pro Quest Dissertations and Theses*, *Google Scholar*, *EconLit* e *Business Source Complete*, além de referências obtidas em trabalhos anteriores de revisão (ex. Coad et al., 2014; OECD, 2010); a revisão citou mais 80 artigos, nenhum dos quais tinha sido realizado ou publicado no Brasil.

Como comentado na seção 1.2, esta tese teve como objetivo principal estudar o impacto de variáveis no crescimento elevado de empresas, através de estudo quantitativo, com uma amostra representativa de firmas, utilizando técnicas adequadas de estatística. Dado o limitado número de artigos baseados em pesquisas quantitativas ou sobre EACs no Brasil, acredita-se que esta tese apresente relevância acadêmica e possa preencher parte da referida lacuna.

Por outro lado, entender como as empresas brasileiras podem aumentar a probabilidade de apresentarem alto crescimento pode contribuir para a economia do país, o que sinaliza a relevância gerencial e social desta pesquisa.

Com efeito, essa relevância gerencial e social está diretamente associada à própria existência de empreendedorismo; pode-se afirmar que quando um grupo de empreendedores se reúne e decide investir em uma empresa, a agenda comum é alcançar um desempenho econômico e financeiro satisfatório, que permita que ela cresça de forma sustentável e alcance o almejado sucesso. Aqui o conceito de sucesso está sendo usado como sinônimo de êxito, de forma vaga, representado pelo faturamento crescente da empresa. Obter sucesso é o objetivo predominante dos empreendedores que vão dedicar tempo e capital a uma empreitada no mundo dos negócios privados, cientes das incertezas e riscos que uma empresa representa, ainda mais em país complexo como o Brasil.

Assim, na medida em que busca separar os *drivers* que contribuem para um alto crescimento das empresas de outras variáveis de menor influência, a presente tese espera ajudar de forma concreta os empreendedores a focarem suas atuações gerenciais. A título de exemplo, assumindo que a variável localização em parques tecnológicos ou incubadoras seja um *driver* para o crescimento elevado de empresas, os empreendedores deveriam buscar espaços em tais polos mais arduamente. Esse exemplo foi escolhido de propósito, para mostrar que mesmo para um *driver* que intuitivamente seja conhecido, confirmar que é realmente relevante e ainda saber sua contribuição para o aumento da probabilidade de uma empresa atingir

crescimento elevado, pode transformar uma mera percepção gerencial em um aprendizado mais consistente.

Esta tese espera ainda ter contribuído para entender o papel de certas políticas públicas e sua efetiva influência no desempenho das empresas. Foram diversas as variáveis testadas para identificar os *drivers* de crescimento elevado; tais variáveis, uma vez conhecidos seus impactos positivos e negativos de forma quantificada e respectivos níveis de significâncias estatísticas, podem contribuir para os agentes formuladores de políticas públicas orientarem suas atuações.

Voltando ao exemplo do item anterior, conhecido o impacto de parques tecnológicos e incubadoras no crescimento elevado de empresas, espera-se que seja mais fácil para os gestores de políticas públicas criarem uma prescrição para o setor, fortalecendo a defesa de mais investimentos nesse tipo de instituição.

1.4 Estrutura da Tese

Esta tese foi estruturada em cinco capítulos; no capítulo 1 está apresentada a introdução, questão de pesquisa e relevância do tema; o capítulo 2 trás a revisão da literatura acadêmica relacionada a empreendedorismo e a empresas de alto crescimento; no capítulo 3 o método de pesquisa é apresentado e discutido; no capítulo 4 foram descritos os resultados empíricos obtidos a partir da amostra coletada; por último, o capítulo 5 trás as considerações finais, detalhando as conclusões, implicações, limitações e perspectivas para estudos futuros.

Após os cinco capítulos, as referências bibliográficas são listadas e o formulário de pesquisa desta tese é apresentado como Apêndice 1.

2 REFERENCIAL TEÓRICO

Este capítulo trás uma revisão da literatura acadêmica focado em empresas de alto crescimento (EACs), tema central desta tese, sendo composto por cinco seções; na primeira, os conceitos de empreendedorismo e inovação são apresentados e discutidos; a seção 2 trata de ecossistema de empreendedorismo, com ênfase no cenário brasileiro; na seção 3 são analisadas as métricas de desempenho de empresas em geral e como são identificadas as EACs; na seção 4 são apresentadas as variáveis preditoras ou explicativas mais frequentemente consideradas na literatura internacional que ajudam a explicar o desempenho de empresas; o capítulo é encerrado pela seção 5, onde as hipóteses de presente tese são apresentadas.

2.1 Empreendedorismo e Inovação

O estudo de empreendedorismo e da influência de ecossistemas nos quais tais empresas estão inseridas tem sido objeto de preocupação acadêmica e gerencial nos últimos anos, tanto em países desenvolvidos quanto em desenvolvimento.

Os conceitos de empreendedorismo e empreendedor utilizados nesta tese, que tem como foco as empresas de alto crescimento, são tratados nesta seção de forma direta. Dessa forma, a definição de Schumpeter (1982), em seu seminal livro *The Theory of Economic Development* publicado em 1934, que associa empreendedorismo à “destruição criativa”, ou seja, criar algo novo e melhor que destrói o anterior, não foi necessariamente seguida nesta tese. Buscou-se uma definição mais alinhada com o perfil das empresas que fizeram parte da amostra utilizada na presente pesquisa, que é apresentada a seguir.

Para Nakagawa (2008, p. 54), empreendedorismo é feito com base em inovação, que se expressa através do lançamento de novos produtos e serviços, utilização de novos métodos de produção, abertura de novos mercados, busca de novos fornecedores e insumos e novas estruturas de mercado; o conceito de inovação foi recentemente ampliado, conforme tipologia estabelecida no respeitado Manual de Oslo (OECD, 2005), no qual quatro tipos de inovação foram definidos: de produto, de processo, organizacional e de marketing (Nakagawa, 2008). Essa segunda definição está mais alinhada com o perfil das empresas utilizadas na presente pesquisa.

Essa relação entre empreendedorismo e inovação permeou a presente pesquisa dado que são conceitos que se complementam e estão relacionados com o desenvolvimento econômico das nações. Como explica Pavani (2003), a inovação é um tema central para qualquer país que busque o desenvolvimento, dado que sua capacidade de inovar está intrinsecamente ligada à competitividade e ao crescimento econômico; afirma ainda aquela autora que as inovações ocorrem nas empresas, onde são transformadas em produtos e serviços, fazendo com que ganhem significado econômico. Essa constatação foi confirmada em pesquisa com 125 empresas brasileiras, elaborada por Andreassi e Sbragia (2002), que mostrou forte associação entre investimentos em Pesquisa e Desenvolvimento (P&D) e a participação de novos produtos no faturamento total.

Nesse contexto, o objetivo principal desta tese foi estudar as empresas de alto crescimento (EACs) e não apenas as empresas nascentes; são vários os estágios de crescimento de uma empresa e diversos autores segmentaram tais fases de formas similares; dois exemplos são Bygrave e Timmons (1992) e Bell e McNamara (1991); ambos definiram cinco os estágios de crescimento, sendo que para a primeira dupla de autores são os seguintes:

- 1) Semente / pesquisa e desenvolvimento
- 2) Startup
- 3) Estágio inicial
- 4) Expansão
- 5) Maturidade

Para Bell e McNamara (1991), os cinco estágios são:

- 1) Conceito
- 2) Semente
- 3) Desenvolvimento de produto
- 4) Desenvolvimento de mercado
- 5) Equilíbrio

As atividades que ocorrem em cada um dos estágios das duas tipologias são semelhantes, passando pelos processos de ideação, criação da empresa, plano de negócios, teste de produtos e serviços, produção, primeiros clientes e vendas, aumento do faturamento e estabilidade das operações. Cabe salientar que Bygrave e Timmons (1992) utilizaram o

conceito de *Startup* para explicar um dos estágios de crescimento das empresas; esse conceito está detalhado nos próximos parágrafos.

Com base nos autores citados, Pavani (2003) propôs uma segmentação híbrida dos estágios de crescimento de uma empresa, a seguir detalhada:

- 1) Concepção / Criação: estágio onde ocorre o desenvolvimento de estudos, pesquisas preliminares, planos de negócios, projetos, pesquisas de mercado, criação jurídica; nessa fase o faturamento é zero e recursos financeiros são consumidos;
- 2) Startup: estágio onde ocorre a finalização de protótipos e iniciam-se os processos de contratação de profissionais, compra de equipamentos e busca de primeiros clientes; o faturamento é baixo e instável; mais recursos financeiros são consumidos;
- 3) Estágio inicial: estágio onde ocorre a estruturação das atividades de produção, marketing e vendas; a empresa continua a investir na contratação de profissionais e compra de equipamentos; ponto de equilíbrio ainda difícil de atingir e mais recursos financeiros são consumidos;
- 4) Crescimento / Expansão: estágio onde ocorre aumento do faturamento e empresa foca em atividades de marketing, vendas e desenvolvimento de novos produtos e serviços; equilíbrio operacional deve ter sido atingido e recursos financeiros são direcionados ao aumento de capital de giro;
- 5) Maturidade: estágio onde ocorrem atividades voltadas para preservar a competitividade da empresa.

Os estágios de uma empresa servem para esclarecer a relação entre empreendedorismo e startups; de acordo com citadas segmentações, startup representa um estágio inicial das empresas, uma fase transitória.

A expressão startup começou a ser utilizado a partir dos anos 90 em empreendedorismo nos Estados Unidos da América e logo chegou ao Brasil. Na literatura existem diversos conceitos de startup; embora mais recente, destaca-se a definição de Steve Blank (2006), para quem startup é uma organização constituída para pesquisar um modelo de negócio que possa ser repetido e escalável; segundo esse autor, existem quatro categorias de startups:

- startups que trazem um novo produto para um mercado existente;
- startups que trazem um novo produto para um novo mercado;

- startups que trazem um novo produto para um mercado existente, na tentativa de penetrá-lo, através de nova segmentação a preços baixos;
- startups que trazem um novo produto para um mercado existente, na tentativa de penetrá-lo através de nova segmentação em um nicho (Blank, 2006).

Pode-se afirmar que o conceito de startup está associado a novos produtos e serviços, que entreguem novos atributos, mirem novos mercados ou novas segmentações de mercados existentes. Assim, caminham juntos os conceitos de empreendedorismo, startups e inovação, embora tenham recortes diferentes.

Outro conceito próximo, muito utilizado em estudos e pesquisas acadêmicas, é o de empresas inovadoras de base tecnológica (EIBT) ou, como são conhecidas nos EUA, *New Technology Based Firms* (NTBF). As definições de EIBT e startup são diferentes, embora em vários contextos possam aparecer de forma sobreposta. Isso se deve ao fato que todas as empresas nascentes com as características de uma EIBT são startups, durante certo período de tempo; entretanto, a recíproca não é verdadeira: nem toda startup pode ser considerada uma EIBT.

A EIBT deve necessariamente trazer uma nova tecnologia para o mercado, enquanto que uma startup, embora tenha a inovação em seu DNA, pode inovar através de um novo modelo de negócios ou um novo produto mais simples e mais barato, como caracterizado pela tipologia de Blank (2006) anteriormente detalhada. Como esclarece Nakagawa (2008), uma EIBT é um tipo singular de empresa que é ao mesmo tempo inovadora e de base tecnológica; existem empresas inovadoras que não são de base tecnológica e empresas de base tecnológica que não são inovadoras (Nakagawa, 2008).

São exemplos de startups e de EIBT: uma nova empresa de prestação de serviços de Tecnologia da Informação (TI), com uma inovadora forma de tarifar seus clientes, pode ser uma startup, mas não é uma EIBT; uma empresa que desenvolve uma nova tecnologia para transformar entulho de demolições em tijolos para novas obras pode ser uma EIBT e ao mesmo tempo uma startup; o mesmo vale para uma nova empresa de biotecnologia, que consegue reduzir a proliferação de mosquitos transmissores de doenças.

Outra fonte de frequentes erros de interpretação, sobretudo no mundo gerencial, é o de confundir Tecnologia com Tecnologia da Informação (TI); todas as empresas, novas ou já existentes, usam TI para gerenciar suas próprias operações; entretanto, apenas parte dessas empresas tem em TI seu *core business*. Nos dias atuais são muito frequentes startups baseadas

em TI, lançando no mercado inovadores aplicativos. Claramente trata-se de uma startup e ao mesmo tempo de uma EIBT. Entretanto, no exemplo citado anteriormente, a empresa para transformar entulho de demolições em tijolos para novas obras pode ser uma startup e ao mesmo tempo uma EIBT, mas seu *core business* passa distante de TI.

A presente tese não teve como unidade de análise o universo de startups, por entender que se trata de um estágio inicial pelo qual as empresas passam, durante uma janela temporal relativamente curta; tampouco procurou identificar se era uma EIBT ou uma empresa de TI; seu objetivo principal foi estudar as empresas de alto crescimento em comparação com demais empresas, independentemente de seu tipo ou setor; sua unidade de análise foi o universo de empreendedorismo, com foco no quarto estágio de crescimento de uma empresa, conforme as tipologias de Bygrave e Timmons (1992), Bell e McNamara (1991) e Pavani (2003), sem assumir a complexa tarefa de identificar se era startup, EIBT ou de TI.

2.2 Ecosistema de Empreendedorismo

O conceito de um sistema que incentive as atividades empreendedoras não é novo e pode ser atribuído a Schumpeter (1982), em seu citado livro de 1934; a partir dos anos 70, alguns autores passaram a se referir ao sistema empreendedor como um *ecossistema*, de forma análoga ao da ecologia, que estuda o meio ambiente e os seres vivos.

O professor Daniel Isenberg, do Babson College (EUA), um dos principais pesquisadores do assunto, entende que o incentivo à atividade empreendedora é peça chave para o desenvolvimento econômico das cidades e países, sendo a metáfora predominante dessa estratégia o ecossistema de empreendedorismo (Isenberg, 2014); em artigo recente, publicado em conjunto com o professor Vicent Onyemah da mesma instituição, foi atualizada a representação gráfica que mostra os amplos domínios do ecossistema de empreendedorismo, ilustrados de forma simplificada na Figura 2 (Isenberg e Onyemah, 2016).

Como se pode depreender da figura a seguir, o conceito de ecossistema de empreendedorismo envolve diversas áreas e partes interessadas, que atuam de forma não coordenada, mas complementar e sinérgica, criando um ambiente favorável ao desenvolvimento de empresas. Dessa forma, diversas instituições governamentais e privadas – com e sem fins lucrativos –

que possam oferecer capital humano, financeiro e adequada infraestrutura a países, regiões e cidades, formam o conjunto de atores de um ecossistema de empreendedorismo.

Figura 2 – Domínios do Ecossistema de Empreendedorismo

Fonte: adaptado de Daniel Isenberg e Vicent Onyemah, 2016.

Andersson e Klepper (2013) compararam ecossistemas de empreendedorismo de quatro países: Suécia, Dinamarca, Brasil e Estados Unidos, com foco nas características e desempenho de novas empresas independentes ou originadas por separação de outras já existentes, conhecidas como *spin-offs*; segundo esses autores, o ambiente para a formação de novas empresas nesses países não apresentou grandes diferenças; as informações utilizadas sobre o Brasil foram originadas em outro artigo, publicado por Hirakawa, Muendler, e Rauch (2010), elaborado a partir de dados da RAIS (Relação Anual de Informações Sociais) do Ministério do Trabalho, onde *spin-offs* criadas no período entre 1995 e 2001 foram estudadas. Concluíram que os desempenhos de *spin-offs* nos países pesquisados foram semelhantes e superiores, comparados a outros tipos de novas empresas, especialmente se as firmas que criaram as *spin-offs* continuassem em operação.

Segundo um dos principais atores do ecossistema de empreendedorismo no Brasil, a Associação Nacional de Entidades Promotoras de Empreendimentos Inovadores – Anprotec – “a oferta de ambientes propícios à criação e ao crescimento de negócios e soluções inovadoras tem se mostrado, ao longo dos anos, uma ferramenta importante de impulso ao desenvolvimento econômico, tecnológico e social” (Anprotec a, 2016); essa importância é ainda maior em países em desenvolvimento, como o Brasil.

Como a maioria dos países em desenvolvimento, o Brasil apresenta muitas oportunidades e desafios para todas as empresas que fazem parte do seu cenário econômico. De dimensões continentais e população estimada em 207,7 milhões de habitantes em julho/2017 (IBGE a, 2017), o Produto Interno Bruto (PIB) no Brasil atingiu a 9ª posição no ranking global em 2016 com US\$ 1,8 trilhões; a renda per capita estimada ocupou a 72ª posição com cerca de US\$ 8.700 (The World Bank, 2016). Essa diferença entre as posições nos dois rankings pode ser interpretada sob duas perspectivas: como reflexo dos grandes problemas relacionados ao desenvolvimento do país e como medida de seu grande potencial de crescimento.

Embora o Brasil seja um país em desenvolvimento, suas características tornam seu mercado interno muito atrativo em praticamente todos os setores econômicos e, conseqüentemente muito competitivo. Com efeito, existem no Brasil 4,6 milhões de empresas (IBGE b, 2017 p. 28), de capital nacional, de grande, médio ou pequeno porte, disputando arduamente quase todos os segmentos de mercado com subsidiárias de multinacionais.

Potencializando esse quadro, o Brasil experimentou um crescimento robusto de sua economia nos últimos 15 anos até 2014, sustentado pela incorporação de grande parcela da população de classes sociais menos favorecidas ao mercado de consumo. Nos últimos dois anos – 2015 e 2016 – uma crise política e econômica interrompeu esse processo de crescimento e melhoria da distribuição de renda; aparentemente as causas dessa crise parecem ter sido controladas e o crescimento foi retomado em 2017, ainda que em ritmo lento. Não obstante situações conjunturais típicas de economias em desenvolvimento, o mercado interno brasileiro segue muito atrativo, razão pela qual continua recebendo sólidos investimentos por parte das grandes empresas.

Tal dinamismo econômico estimula uma profusão de oportunidades de novos negócios, parte dos quais são aproveitados pelas grandes corporações multinacionais e nacionais, enquanto outros são negligenciados ou ignorados pelas empresas existentes. Aproveitando-se de

oportunidades não prioritárias para as grandes empresas e, por vezes incentivadas ou mesmo patrocinadas por elas, o empreendedorismo vem florescendo no Brasil nos últimos anos; empresas, tanto existentes quanto novas, estão sendo estimuladas para se desenvolverem lançando produtos e serviços inovadores, melhorando processos e criando novos modelos de negócios em todos os setores econômicos.

Mais recentemente está havendo uma mudança de foco e grande quantidade de empresas está se voltando cada vez mais para o setor corporativo, rotulado como mercado Business-to-Business (B2B). O risco associado a uma empresa que desenvolve produtos e serviços para o mercado B2B costuma ser menor, sobretudo quando pode contar com o apoio, patrocínio e até mesmo certo faturamento garantido para uma grande corporação.

Por outro lado, as grandes empresas acabam também se beneficiando bastante do convívio com as pequenas e médias empresas, que aportam inovação, criatividade e novas soluções para problemas relevantes, rejuvenescendo-as a partir dessa inusitada relação. Prova dessa tendência é que grandes empresas como o Banco Itaú®, Banco Bradesco®, Nestlé®, Hospital Israelita Albert Einstein®, Google® e muitas outras estão investindo de forma consistente em atrair e apoiar firmas inovadoras que possam arejar seus negócios. Esse apoio frequentemente ocorre através de suporte financeiro e com a criação de espaços de coworking, oferecidos para a instalação de empresas que tenham afinidade com suas atividades, de forma subsidiada.

A oferta de financiamento para as empresas, existentes ou novas, é uma das principais ferramentas de incentivo ao empreendedorismo; é uma condição necessária, porém não suficiente, como ilustra a Figura 2. Portanto, outra forma de evidenciar a vitalidade do empreendedorismo brasileiro é através do crescente investimento de fundos de *private equity* e *venture capital*. Segundo relatório da ABVCAP e KPMG (2016), até o ano de 2015 o capital comprometido acumulado com o Brasil foi de R\$ 153 bilhões, dos quais R\$ 102 bilhões foram efetivamente investidos em empresas; o CAGR no período 2011 a 2016 desses dois indicadores foram respectivamente 25% e 27% respectivamente. Cerca de 95% desse montante foi investido por fundos de *private equity*, mas como afirma o relatório da ABVCAP e KPMG (2016): “o capital comprometido de R\$ 5,3 bilhões e disponível para investimentos de R\$ 1,7 bilhão na modalidade *venture capital* é bem relevante, pois se tratam de investimentos com valor médio bem menor que os de *private equity*” (ABVCAP e KPMG, 2016, p. 13).

Outro indicador relevante desse mesmo relatório do ano anterior comparou a parcela de investimentos anuais em relação ao PIB em 2014 dos países conhecidos por BRICS, obtendo-se: 0,24% no Brasil, 0,15% na China, 0,20% na Índia, menos de 0,01% na Rússia e 0,10% na África do Sul (ABVCAP & KPMG, 2015). Dos dois relatórios pode-se concluir que o Brasil tem atraído capital para investimentos de forma satisfatória e que existe certa disponibilidade de recursos para financiar o empreendedorismo.

Embora o ecossistema de empreendedorismo brasileiro tenha melhorado nos últimos anos, muitos problemas ainda persistem. Alguns deles foram identificados por Cahen, Lahiri, e Borini (2016) em artigo sobre a internacionalização de empresas inovadoras brasileiras baseadas em tecnologia (EIBT). Afirmam esses autores que existem três importantes barreiras: a institucional externa, a capacitação organizacional interna e a de recursos humanos. Através de *survey* junto a 92 empresas brasileiras ligadas à Anprotec, esse artigo mostrou que o ecossistema brasileiro necessita eliminar barreiras e identificar os *drivers* para favorecer a internacionalização de empresas de base tecnológica.

Nos domínios do ecossistema de empreendedorismo, ilustrados na Figura 2, coexistem diversas partes interessadas. Visando esclarecer alguns conceitos e definições frequentemente utilizados na presente tese, foram preparadas duas tabelas explicativas; a Tabela 1 descreve os principais ambientes relacionados e favoráveis ao empreendedorismo, enquanto que a Tabela 2 detalha as principais fontes de capital que costumam financiar as atividades do setor.

Tabela 1 – Ambientes Relacionados a Empreendedorismo

Instituição	Conceito	Exemplos	Fonte
Parque Tecnológico	Complexo produtivo industrial e de serviços de base científico-tecnológica. São planejados e têm caráter formal, concentrado e cooperativo, agregando empresas cuja produção se baseia em Pesquisa e Desenvolvimento (P&D). Atuam como promotores da cultura de inovação, da competitividade e da capacitação empresarial, fundamentados na transferência de conhecimento e tecnologia, com o objetivo de incrementar a produção de riqueza de uma determinada região. Alguns parques tecnológicos estão ligados a universidades.	Parques de São José dos Campos, Sorocaba, UFRJ	Anprotec b (2016)
Incubadora	Entidade que tem por objetivo oferecer suporte a empreendedores para que possam desenvolver ideias inovadoras e transformá-las em empreendimentos de sucesso. Oferece infraestrutura e suporte gerencial, orientando os empreendedores quanto à gestão do negócio e sua competitividade. Algumas incubadoras estão ligadas a universidades.	Cietec (USP), InCamp (Unicamp)	Anprotec b (2016)
Aceleradora	Entidades focadas em empresas que tenham o potencial para crescerem muito rápido. Buscam empresas escaláveis e não somente uma pequena empresa promissora. Estudam o modelo de negócios e apostam em boas ideias, com menos formalismo que as Incubadoras. São lideradas por empreendedores ou investidores experientes; em geral usam capital privado; apoiam as empresas em sessões de mentoring.	StartupFarm, Aceleratech, Wayra	Sebrae (2016)
Espaços de Coworking	Nova forma de pensar o ambiente de trabalho empreendedor. Reúnem empresas e pessoas a fim de trabalhar em um ambiente inspirador, especialmente pensado para o trabalho autônomo, muito networking com pessoas de diversas áreas e toda a estrutura para receber clientes, com um custo menor do que teria ao alugar uma sala comercial.	Cubo (Itaú), Creative Coworking	Coworking (2016)

Fonte: Elaborado pelo autor a partir das fontes acima citadas.

Da Tabela 1 pode-se deduzir que a simples localização de uma empresa pode trazer grandes diferenças quanto ao ecossistema de empreendedorismo em que está inserida; quando uma empresa está instalada em um parque tecnológico, incubadora ou espaço de coworking as possibilidades de *networking* aumentam exponencialmente. A rede de relacionamento pode influenciar a forma como a empresa opera e se desenvolve, impactando o seu desempenho.

Essa influência positiva, quando uma empresa está localizada em um ambiente favorável, é associada ao chamado efeito “filtro de entrada”; esse efeito decorre do fato que somente empresas promissoras são admitidas para integrar os espaços de incubadoras ou parques tecnológicos, uma vez que a demanda é maior que a oferta. Ficam dadas as condições de interdependência: a empresa era inovadora e com grande potencial para atingir alto desempenho e por isso foi aceita para integrar um ambiente de inovação ou seu desempenho foi superior em decorrência de sua integração a tal ambiente? A presente pesquisa buscou controlar o efeito filtro de entrada durante a fase de análise de dados.

Em relação às fontes de capital, a Tabela 2 esclarece as nuances de perfil e foco de alguns tipos de instituições que têm investido de forma consistente em empreendedorismo; tal segmentação está associada aos estágios de desenvolvimento das empresas da seção 2.1.

Tabela 2 – Fontes de Capital Relacionadas a Empreendedorismo

Instituição	Conceito	Exemplos	Fonte
Investidor Angel	Pessoa física que pode investir, por meio de capital próprio, na empresa nascente, desde que ela apresente um alto potencial de crescimento. Acompanha e apoia no sentido de aumentar as chances de sucesso, ajudando o empreendedor e a empresa com a própria experiência, com o conhecimento no campo dos negócios, e com uma valiosa rede de relacionamentos. Não tem caráter filantrópico ou assistencialista; o investidor-anjo vê a possibilidade de investir na empresa como negócio, pensando no lucro.	Pierre Schurmann, Mario Ghio, Ricardo Levinzon	Endeavor (2016)
Capital Semente	Buscam negócios na fase embrionária, como o Angel; a diferença é que, em geral, o investidor-anjo é uma pessoa física e o capital semente vem de uma pessoa jurídica, por meio de um fundo de investimento. Outra diferença com o Angel, é que focam em empresas que já possuem clientes e produtos definidos, mas dependem de investimento para expandirem o consumo e se estabelecerem no mercado.	Horizonti, Criatec, Nascenti	Endeavor (2016)
Venture Capital	Fundos de investimentos de risco focados em pequenas e médias empresas (PME). Em geral, investem em empresas de médio porte, que já tem um faturamento expressivo, mas que ainda precisam dar um salto de crescimento. Com o investimento, o objetivo é ajudar essas empresas a expandir e alcançar o seu potencial máximo.	Eccelera, Rio Bravo, FIR	Endeavor (2016)
Private Equity	Fundos de investimentos de risco focados em investir em operações de fusões e aquisições de grandes empresas. Buscam empresas de capital aberto ou prestes a abrir seu capital.	Advent, GP, Pátria	Endeavor (2016)

Fonte: Elaborado pelo autor a partir das fontes acima citadas.

Em princípio, existe capital de risco para as empresas que são inovadoras, desde que apoiadas em sólidos modelos de negócios e gerenciadas por uma comprometida e qualificada equipe de sócios. Ainda que fontes de financiamento sejam mais escassas no Brasil que em países desenvolvidos, a oferta de investidores privados está crescendo junto como setor de forma consistente.

Deve-se ressaltar que as Tabelas 1 e 2 visaram esclarecer como funcionam os principais ambientes e fontes de financiamento que atuam em empreendedorismo, de forma didática e sintética; na prática, porém, os limites de atuação de cada entidade são tênues, sendo frequentes sobreposições e variações de escopo e abrangência. Como exemplos, existem incubadoras que trabalhem com empresas incubadas residentes e não residentes, e incubadoras que podem fazer aportes de capital, através de empresas coligadas; ou seja, deve-

se ter presente que as entidades mencionadas nas duas tabelas não são estanques e sim integradas e sinérgicas.

2.3 Empresas de Alto Crescimento e Gazelas

São escassos os estudos acadêmicos com pesquisas empíricas sobre empresas de alto crescimento (EACs), conforme revisão da literatura publicada por Demir, Wennberg e McKelvie (2017); com foco na base de artigos *ISI Web of Knowledge*, complementado por buscas em mais 18 importantes revistas da área, os autores identificaram apenas 39 publicações durante o período 1985 a 2015.

Os artigos detalhados nessa citada revisão utilizaram variadas métricas de desempenho de empresas para identificar as EACs e as gazelas; boa parte deles utilizou o crescimento de faturamento ou de número de colaboradores, seguindo o critério da OECD (2007) ou métricas similares; outra parte mediu a rentabilidade das empresas, utilizando bases de dados secundários; alguns estudos simplesmente foram baseados em sobrevivência das empresas, sobretudo quando se tratavam de empresas nascentes; não foi verificado claro consenso acadêmico em como identificar as empresas de alto crescimento (Demir, Wennberg e McKelvie, 2017).

Essa situação parece ser consequência da complexidade em medir o desempenho de empresas, principalmente de empresas nascentes; pode-se citar como exemplo uma nova empresa, oferecendo ao mercado um produto inovador, por meio de um novo modelo de negócios, poucos meses de operação, faturamento relativamente baixo, custos elevados, operando no vermelho ao nível de *bottom-line*; como avaliar o desempenho dessa empresa? Ou como frequentemente citado em artigos de revistas acadêmicas ou de negócios: é uma empresa de sucesso ou não?

Quando a métrica muda para taxa de sobrevivência das empresas como *proxy* de desempenho, critério aparentemente mais objetivo e utilizado em vários artigos acadêmicos, vantagens e desvantagens emergem.

Um exemplo é o artigo de Van Gelderen et al. (2006) sobre fatores de sucesso e de risco de empresas em sua fase bem inicial; esses autores usaram sobrevivência após três anos, classificando como sucesso as empresas que estavam estabelecidas e não sucesso as que tinham sido abandonadas ou encerradas; de uma amostra com 310 empresas nascentes na

Holanda, 63% estavam em operação e os restantes 37% haviam desistido de seguir adiante (Van Gelderen et al., 2006). Outro exemplo vem de Andersson e Klepper (2013), que consideraram taxa de sobrevivência a 3, 6, 9 e 12 anos na Suécia. Por ambos os estudos terem sido feitos em países desenvolvidos, a sobrevivência pode ser avaliada em prazos relativamente longos.

Os estudos de Robb e Farhat (2013), com 4.928 startups dos Estados Unidos da América que fizeram parte do Kauffman Firm Survey, e de Arruda et al. (2014), com 221 empresas no Brasil, ajudam a esclarecer a questão sobre o prazo ideal para avaliar a sobrevivência de uma empresa. Segundo esses artigos a taxa de sobrevivência de empresas é baixa nos dois países, porém bem distintas; no estudo americano a taxa de sobrevivência das empresas fundadas em 2004 foi 45% em 2011, após sete anos (Robb & Farhat, 2013); no estudo brasileiro as taxas foram 75% em um ano, 50% em quatro anos e 25% em treze anos (Arruda et al., 2014). Pode-se concluir que no Brasil a adoção de um determinado prazo para avaliar a taxa de sobrevivência é tarefa difícil, sujeita a aportar imprecisões a qualquer estudo acadêmico.

Além da dificuldade em estabelecer o número de anos de sobrevivência de uma empresa para identificar sucesso versus não sucesso, sua mera existência é um indicador limitado de desempenho econômico. Assim, pode-se concluir que sobrevivência é uma métrica útil, porém insuficiente para avaliar o desempenho de uma empresa, em qualquer país do mundo. Não basta sobreviver e sim prosperar e crescer.

O faturamento e o número de colaboradores parecem ser métricas mais adequadas para avaliar o desempenho das empresas. Quando uma empresa tem operações comerciais, emite notas fiscais e cria empregos diretos ou indiretos, um novo patamar de desenvolvimento é atingido. Tem limitado valor uma empresa, existente ou nova, ter ótimas intenções, grandes planos, criativos modelos de negócios e buscar desenvolver produtos e serviços inovadores, se tudo isso não for transladado para o faturamento e a consequente geração de empregos ao longo dos anos.

Dessa forma, para se medir o desempenho de uma empresa uma primeira métrica é seu faturamento e as taxas de crescimento anuais; uma segunda, parcialmente correlacionada e complementar, é o número de colaboradores, também medido ano a ano.

Sabe-se das dificuldades burocráticas e custos elevados que implicam as contratações de funcionários no Brasil; por isso o monitoramento do número de colaboradores ao longo dos

anos é uma forma clara e relativamente robusta de fazer inferências sobre o desempenho de uma empresa.

Essas duas métricas, faturamento e número de colaboradores, devem ser medidas ano-a-ano, sendo importantes não só por suas magnitudes absolutas como, sobretudo, pelas respectivas taxas de crescimentos anuais e acumulada no período.

Existem vários artigos na literatura que discutem a relação entre o crescimento do número de colaboradores e a produtividade das empresas, dado o potencial conflito entre as duas métricas, como explica Monteiro (2017); no entanto, de um modo geral, a relação entre o crescimento da produtividade e o crescimento das empresas é positiva (Bravo-Biosca, 2011); Erken, Donselaar e Thurik (2016) defendem que empreendedorismo apresenta um impacto estável e relevante para o crescimento da produtividade da economia; Du e Temouri (2015) sustentam que as EACs apresentam maior probabilidade de atingirem alto crescimento da produtividade.

A dinâmica da atividade empreendedora é grande e por essa razão as empresas podem registrar alto crescimento em um determinado período de tempo, e não manter essa condição ao longo da vida ou em períodos sucessivos (Daunfeldt e Halvarsson, 2015); embora temporário, o período de alto crescimento é uma transição observável (Monteiro, 2017), que a presente pesquisa procurou capturar.

2.4 Variáveis Predictoras de Desempenho de EACs

Muitas são as variáveis predictoras que vêm sendo objeto de estudos acadêmicos na ampla literatura sobre desempenho de empresas; tais variáveis podem ser classificadas em áreas específicas, visando agrupá-las de acordo com sua natureza; não existe um padrão único e a lista a seguir serve para exemplificar uma forma de classificação:

- Sócios: número, idade, gênero, formação acadêmica, classe social, perfil empreendedor, dedicação em tempo integral, aspectos motivacionais;
- Características demográficas: setor de atuação, endereço, região;
- Local da sede: parques tecnológicos, incubadoras, espaços de coworking, escritórios próprios;
- Estratégia corporativa: plano estratégico, plano de negócios, público-alvo;

- Características dos produtos e serviços: tecnologia, inovação, marcas e patentes;
- Marketing: plano de marketing, design, propaganda;
- *Spin-offs*: acadêmicas, corporativas, outras;
- Fontes de capital: dos sócios, de familiares ou amigos, crowdfunding, investidores *angel*, aceleradoras, capital semente, fundos de *venture capital* ou *private equity*, empréstimos bancários;
- Instituições de fomento: aportes públicos, aportes privados;
- Parcerias: parceiros, acordos de colaboração;
- Apoios recebidos: doações, premiações.

A maioria de artigos publicados sobre o assunto foca determinados aspectos da lista acima, buscando correlacionar métricas de sucesso com algumas dessas variáveis. Van Gelderen et al. (2006), por exemplo, utilizou como variáveis explicativas em seu artigo sobre sucesso de novas empresas diversas variáveis da lista acima, inclusive motivação dos empreendedores e tempo parcial ou integral dedicado ao trabalho. Em outro exemplo, Andersson e Klepper (2013) utilizaram como uma das variáveis preditoras, a condição das empresas terem sido formadas a partir de outras, ou seja, se eram *spin-offs* de outras.

Os pesquisadores do MIT Fazio, Guzman, Murray e Stern (2016) publicaram uma pesquisa de certo modo semelhante a presente tese, com o objetivo de identificar, no momento em que nasce uma nova empresa, quais seriam as características que poderiam ser associados à maior probabilidade de seu crescimento acelerado; denominaram tais características de “*digital signatures*”, que embora não representam *drivers* causais de crescimento, estão associadas à qualidade dos empreendedores; usando o modelo de Regressão Logística e bases de dados secundários, chegaram à conclusão que as características que indicam a ambição dos empreendedores no momento da fundação da empresa, avaliada através de como a empresa é denominada, a escolha do tipo de entidade legal e se faz aplicações de patentes e marcas, estão associadas à maior probabilidade de crescimento acelerado.

Em estudo brasileiro, Arruda et al. (2014) estimou o impacto do número de sócios que trabalha em tempo integral em uma empresa, no momento em ela é constituída; concluíram que a chances de descontinuidade da empresa aumentava em 1,24 vezes a cada sócio a mais; esses autores afirmaram também que as novas empresas apresentavam chance de descontinuidade menor, em cerca de 3,45 vezes, quando instaladas em parques tecnológicos,

incubadoras ou aceleradoras, em comparação a escritórios próprios; dentre as percepções dos empreendedores para a sobrevivência das empresas, a aceitação do produto ou serviço pelo mercado e a disponibilidade de capital para investir foram consideradas importantes (Arruda et al., 2014).

A partir da profusão de métricas preditoras da literatura, mais de 30 variáveis, categóricas ou métricas, foram incluídas no formulário da presente pesquisa, conforme detalhado na seção 3.5 e formuladas as hipóteses a seguir elencadas.

2.5 Hipóteses

Todas as hipóteses desta tese partiram de uma mesma base comum: avaliar se determinadas variáveis ajudam a explicar o desempenho de empresas, de forma a caracterizá-las como empresas de alto crescimento (EACs) ou não.

No entanto, as variáveis que levam as empresas a alto crescimento são bastante fragmentadas e desprovidas de abordagem sistemática (Coad et al., 2014); assim, para formular as hipóteses da presente tese, procurou-se embasamento não apenas na literatura de EACs, mas também em alguns artigos publicados na área de empreendedorismo.

A primeira variável considerada está relacionada com a localização das empresas em parques tecnológicos, incubadoras ou espaços de coworking; a literatura sobre a influência do local onde a empresa está sediada no desempenho das firmas é relativamente vasta, desde estudos de casos únicos ou múltiplos a pesquisas empíricas de natureza quantitativa; alguns desses artigos foram citados a seguir, para dar suporte à primeira hipótese.

Motohashi (2013) avaliou o papel do parque tecnológico ligado à Universidade de Tsinghua em Beijing sobre o desempenho de inovação de startups nele sediadas; mais de metade das 68 empresas de survey conduzido pelo autor eram da indústria de tecnologia da informação; como conclusão, a pesquisa indicou que as empresas mais inovativas tiveram limitada colaboração formal de pesquisa e desenvolvimento originada na universidade; no entanto, houve uma benéfica relação informal das empresas residentes com os professores e estudantes, bem como valioso apoio para a gestão de recursos humanos, oferecido pelo parque; a rede de relacionamentos entre as empresas sediadas no parque revelou-se de limitada importância.

Em estudo focalizado em incubadoras, Barbero, Casillas, Ramos e Guitart (2012) estudaram o papel dessas instituições no desempenho de empresas incubadas na região de Andaluzia na Espanha; através de survey direcionado a 90 incubadoras, obtiveram 70 respostas que permitiram concluir que três em cinco arquétipos de incubadoras – de universidade, de pesquisa básica e privada – apresentaram desempenho superior das empresas incubadas em comparação a não incubadas; as incubadoras associadas ao desenvolvimento econômico e regional não apresentaram resultados comparativos diferentes.

Ainda sobre incubadoras, Bruneel, Ratinho, Clarysse e Groen (2012) publicaram um estudo de casos múltiplos sobre a evolução dessas instituições na Alemanha e Holanda, através da comparação da demanda e da oferta de serviços ao longo do tempo; pesquisando sete incubadoras e ouvindo tanto gestores como empresas incubadas, avaliaram os serviços oferecidos e utilizados nessas instituições; concluíram que embora os serviços básicos oferecidos pelas incubadoras de várias gerações sejam similares, as empresas incubadas mais antigas fazem menor uso do portfolio ofertado; por isso, recomendaram critérios mais restritos e claros de seleção e de saída de empresas incubadas, sobretudo para as incubadoras de geração mais antiga. Embora o estudo não tenha acessado diretamente o desempenho das empresas incubadas, percebe-se que a contribuição de incubadoras apresenta heterogeneidade. Os citados estudos sobre parques tecnológicos e incubadoras serviram para dar suporte à primeira hipótese da presente tese:

Hipótese 1: As empresas localizadas em parques tecnológicos, incubadoras ou espaços de coworking têm maior probabilidade de apresentar alto crescimento.

Sabe-se que um dos aspectos fundamentais que impactam o desempenho de empresas é a disponibilidade de capital financeiro; qualquer empresa para ser constituída precisa de investimentos iniciais; após sua fundação, a grande maioria das empresas vai continuar necessitando de financiamentos para crescer; uma comprovação empírica da necessidade de capital das empresas foi bem documentada por Mateev e Anastasov (2011), quando pesquisaram os determinantes de crescimento de pequenas e médias empresas, em sete países do centro e leste da Europa.

Os autores utilizaram o critério da OECD (2007) para identificar as empresas de alto crescimento, o mesmo utilizado nesta tese, e dados secundários de um painel formado por

4.561 firmas; dentre as conclusões, identificaram que não apenas as variáveis que consideraram tradicionais, como tamanho e idade da empresa, impactaram no seu crescimento, mas que aspectos financeiros, tais como endividamento e financiamento próprio, eram de grande relevância; segundo o artigo, as EACs foram mais sensíveis a problemas de fluxo de caixa e confiaram mais em fontes de financiamento externas que as firmas de baixo crescimento (Mateev e Anastasov, 2011).

Mais especificamente sobre o papel dos investidores *Angels*, uma revisão da literatura elaborada por Politis (2008), discutiu o valor agregado de tais atores, associado não apenas ao aporte de capital, mas também a outras contribuições para o desempenho das empresas com as quais se relacionavam. O autor identificou 14 estudos empíricos publicados entre 1992 e 2005, sobre o valor adicionado por *Angels* às empresas; como conclusões, defendeu que os papéis relacionados à orientação estratégica, supervisão e monitoramento de resultados, aquisição de recursos com apoio financeiro e *mentoring* são muito relevantes para as empresas nas quais os *Angels* fazem investimentos.

Em complemento aos diversos estudos da literatura sobre a importância de aportes financeiros para o crescimento das empresas, deve ser mencionado que o custo de capital no Brasil é elevado, o que desestimula o empreendedorismo; assim, com base nos artigos acadêmicos e na atual conjuntura brasileira, foi formulada a segunda hipótese:

Hipótese 2: As empresas que recebem aporte de capital de investidores privados (*angel*, capital semente, *venture capital* e outros) têm maior probabilidade de apresentar alto crescimento.

Outra variável muito estudada na literatura de empreendedorismo é a contribuição das aceleradoras para o sucesso das empresas aceleradas. Um desses estudos, feito por Wise e Valliere (2014), analisou as taxas de risco das empresas aceleradas, no processo de saída das aceleradoras; antes de apresentarem os resultados empíricos, os autores comentaram com propriedade os três elementos principais de valor agregado pelas aceleradoras para as empresas aceleradas: *mentoring* e habilidade de aprender com profissionais mais experientes, acesso à poderosa rede de relacionamento que pode alavancar a nova empresa e reforço de marca e legitimidade; de forma diferente das incubadoras, citaram que as aceleradoras podem

contribuir com capital para as empresas aceleradas e costumam limitar o tempo de permanência.

O foco da pesquisa conduzida por Wise e Valliere (2014) foi na experiência direta dos fundadores da aceleradora em novas empresas e no respectivo grau de relacionamento com o ecossistema de empreendedorismo; concluíram que quanto mais experientes os gestores da aceleradora, menor a probabilidade de fracasso no processo de saída da acelerada; o relacionamento dos gestores da aceleradora com o ecossistema de empreendedorismo não demonstrou resultado significativo; dessa forma, a pesquisa mostrou que aceleradoras com certas características podem contribuir para o sucesso de empresas aceleradas, dando suporte à terceira hipótese da presente tese:

Hipótese 3: As empresas que se relacionam com aceleradoras têm maior probabilidade de apresentar alto crescimento.

Inovação em produtos, serviços ou em modelo de negócios são áreas de grande relevância na literatura de empreendedorismo, razão pela qual receberam especial ênfase nesta tese; alguns estudos mostram que as empresas que inovam apresentam maior probabilidade de sucesso ou de apresentarem alto crescimento. Em um estudo de casos múltiplos, Bertels, Koen e Elsum (2015) analisaram o impacto de inovações que se distanciavam do negócio previamente existente de seis empresas americanas; tais inovações que os autores denominaram “*outside the core*” focaram novos clientes ou novos mercados; dos casos estudados, quatro empresas conseguiram inovar com sucesso, enquanto que as demais duas não tiveram êxito; a partir da verificação do modelo de negócios dessas empresas os autores concluíram que, ao contrário do esperado, a probabilidade de fracasso não era relacionada ao grau de distanciamento da inovação do negócio existente; tal risco tampouco era influenciado pela falta de familiaridade com o mercado, mas por falsas premissas sobre os canais de distribuição, estrutura de custos, margens unitárias e velocidade de implementação da inovação (Bertels, Koen e Elsum, 2015).

Outro estudo feito na Espanha por Segarra e Teruel (2014) comprovou o impacto positivo de inovação no desempenho de empresas; utilizaram dados secundários do Painel de Inovação Tecnológica Espanhola, conhecida por PITEC (análogo à PINTEC, Pesquisa de Inovação elaborada no Brasil pelo IBGE), o critério da OECD (2007) para identificar as EACs e o modelo estatístico Probit, similar à Regressão Logística; a partir de uma amostra com 3.807

firmas, os autores concluíram que os investimentos em Pesquisa e Desenvolvimento afetam positivamente a probabilidade de uma empresa conseguir alto crescimento. Esses estudos serviram para definir a quarta hipótese da presente tese:

Hipótese 4: As empresas que inovam em produtos e serviços ou em modelo de negócios têm maior probabilidade de apresentar alto crescimento.

A quinta e última hipótese teve como base de partida o estudo de Andersson e Klepper (2013), citado na seção 2.2 da presente tese, sobre o desempenho de empresas originadas através de *spin-offs* de outras; atestaram os autores que o desempenho de *spin-offs* foram superiores, quando comparados a outros tipos de novas empresas.

Corroborando o artigo anterior, Feeser e Willard (1989) compararam o desempenho de novas empresas, utilizando uma amostra com 39 de crescimento rápido e outras 39 de crescimento lento; utilizou variáveis independentes relacionadas com a firma na qual o empreendedor da nova empresa trabalhava antes de iniciar seu novo empreendimento; concluíram que as novas empresas de crescimento rápido foram caracterizadas por: i) competir em mercados com tecnologias muito próximas das firmas anteriores, das quais os empreendedores eram egressos; ii) tamanho maior das firmas anteriores; iii) natureza pública das firmas anteriores. Embora os autores não tenham utilizado a expressão *spin-offs* para as novas empresas, pode-se inferir que o construto utilizado, de relacionamento dos empreendedores com as firmas onde trabalhavam anteriormente, é de certo modo análogo. Assim, os estudos de Andersson e Klepper (2013) e de Feeser e Willard (1989) servem como sustentação para a quinta hipótese:

Hipótese 5: As empresas formadas a partir de outras – *spin-offs* – têm maior probabilidade de apresentar alto crescimento.

Esse conjunto de hipóteses constituiu a base principal de variáveis que foram testadas nesta pesquisa; o modelo Regressão Logística permitiu, no entanto, testar um conjunto bem mais abrangente de variáveis, detalhadas na seção 3.5 desta tese, além das especificadas nas cinco hipóteses.

3 MÉTODO DE PESQUISA

Este capítulo descreve o método de pesquisa desta tese e está dividido em cinco seções: a primeira apresenta um desenho metodológico da pesquisa e discute as principais variáveis de estudo; a segunda seção detalha como foi obtida a amostra de empresas e os procedimentos adotados para garantir sua representatividade; a seção 3 detalha o modelo Regressão Logística Binária, que é a principal ferramenta estatística a dar suporte às análises; na quarta seção é explicado o modelo Análise Fatorial, que embora de uso secundário, foi utilizado para tratar parte dos dados e fazer análises complementares; o capítulo se encerra com a seção 5, onde a lista completa de variáveis utilizadas na coleta de dados primários é detalhada.

3.1 Desenho da Pesquisa e Principais Variáveis

Como detalhado nos capítulos 1 e 2, a presente tese visa identificar as variáveis relevantes que ajudam a explicar porque algumas empresas apresentam alto crescimento, enquanto outras crescem de forma lenta e gradual. O conceito utilizado de alto crescimento é o definido pela OECD (2007, p.61), medido através do crescimento de faturamento anual ou de número de colaboradores durante um período de três anos; a firma que atingir crescimento maior que 20% ao ano em pelo menos uma das duas variáveis deve ser considerada uma empresa de alto crescimento (EAC).

Dessa forma, para que uma firma seja classificada como uma EAC é utilizada a métrica conhecida como CAGR, sigla utilizada para *Compound Annual Growth Rate*; o CAGR é calculado pelas equações matemáticas (1) e (2) a seguir, para um período de três anos, conotados como ano atual xx e ano base $xx-3$:

CAGR de faturamento anual da empresa para 3 anos

$$\sqrt[3]{\frac{\text{faturamento}_{xx}}{\text{faturamento}_{xx-3}}} - 1 > 20\% \quad (1)$$

CAGR de número de colaboradores da empresa para 3 anos

$$\sqrt[3]{\frac{\text{número de colaboradores}_{xx}}{\text{número de colaboradores}_{xx-3}}} - 1 > 20\% \quad (2)$$

Note-se que um CAGR de 20% em três anos equivale à taxa de crescimento acumulada de 72,8% durante o período definido entre os anos xx e $xx-3$. O CAGR é uma taxa composta e não significa que a empresa tenha que crescer 20% a cada ano do período, podendo naturalmente o crescimento de um ano ficar abaixo de 20%, desde que compensado em outro ano por crescimento acima desse limiar.

A OECD (2007, p. 61) recomenda um cuidado adicional seja utilizado; empresas pequenas, com reduzido faturamento e poucos colaboradores, de forma relativamente fácil, poderiam atingir o indicador de 72,8% de crescimento no período de três anos; como exemplo, bastaria passar de um faturamento de R\$ 10.000 por ano para R\$ 18.000 ou então de 4 para 7 colaboradores, em três anos, para que as taxas de crescimento fossem superiores a 72,8% e, conseqüentemente, a firma fosse considerada uma EAC. Assim, a entidade recomenda um limite inferior para o número de colaboradores no ano base $xx-3$, para evitar distorcer as métricas; esse limite não pode ser grande, para evitar a exclusão de muitas empresas, e foi sugerido 10 colaboradores no ano base do período de aferição do crescimento, ou seja, no ano $xx-3$.

Esse limite mínimo inicial de 10 colaboradores foi mantido como condição necessária para identificar uma EAC para as duas métricas de aferição, crescimento de faturamento ou crescimento do número de colaboradores. A OECD (2007, p. 61) decidiu não adotar um limite mínimo de faturamento no ano base, uma vez que entende que um mesmo valor seria difícil de aplicar a todos os países.

Assim, o critério de inclusão da OECD (2007) para uma empresa ser considerada uma EAC é atingir CAGR maior que 20% entre os anos xx e $xx-3$, medido em termos de faturamento ou número de colaboradores, desde que a empresa tenha no ano $xx-3$ pelo menos 10 colaboradores.

Definido o critério de inclusão para as EACs, o primeiro passo para lograr o objetivo da presente pesquisa foi obter uma amostra de empresas representativa e de tamanho adequado,

para utilização de modelos estatísticos, com dados de faturamento anual e do número de colaboradores ao final de cada ano; com base nesses dados, torna-se viável calcular os respectivos CAGRs e consequentemente separar as empresas em dois grupos: EACs e não EACs.

No entanto, não basta coletar os dados de faturamento e número de colaboradores das empresas; é preciso também coletar diversas outras características das firmas, uma vez que o passo seguinte é identificar quais são as variáveis que ajudam a explicar porque uma parcela delas atinge o desempenho que as classificam como EACs enquanto outras não.

Dessa maneira, foram coletadas diversas variáveis, detalhadas na seção 3.5, que podem ser agrupadas em cinco blocos: i) variáveis que permitam estimar os CAGRs das empresas em termos de faturamento e do número de colaboradores; ii) variáveis demográficas, tais como tempo de existência da empresa, localização ou número de sócios; iii) variáveis de percepção da importância atribuída pelos empreendedores a diversas ferramentas de gestão, tais como inovação em produtos e serviços, investimentos em marketing, marcas e patentes; iv) variáveis indicativas do relacionamento e aporte de capital de variadas fontes, tais como investidores *angel* ou bancos; v) variáveis sobre o perfil do empreendedor.

Em resumo, a presente tese partiu de uma amostra de empresas para as quais foram coletados dados de faturamento, número de colaboradores e diversas variáveis que as caracterizam, de modo a permitir entender como as firmas foram constituídas, como são gerenciadas e como se relacionam com os principais *stakeholders* de empreendedorismo. A partir dessa base de dados foram feitas as análises estatísticas para identificar as variáveis que ajudam a explicar porque algumas empresas apresentam elevado crescimento, em comparação com as demais.

3.2 Amostragem

Conforme detalhado na seção anterior, para concretizar esta tese foi necessário coletar uma base de dados primários através de *survey* com empresas em operação no Brasil. A opção por criar uma base de dados primários não é a forma recomendada pela literatura acadêmica como primeira escolha; o uso de bases de dados secundários apresenta muitas vantagens em comparação ao uso de dados primários: os dados secundários são obtidos de forma mais rápida, menos trabalhosa e mais econômica; como contrapartida, alguns cuidados são

necessários ao se utilizar dados secundários: não serem atualizados, não terem a acurácia necessária e não atenderem às necessidades da pesquisa (Malhotra, 2012, Aaker et al., 2013).

Após analisar as limitadas bases de dados secundários sobre empresas disponíveis no Brasil, concluiu-se que esta tese deveria contar com uma base de dados primários própria, desenvolvida para atender às suas necessidades específicas. Isso porque não foi encontrada nenhuma base de dados secundários que atendesse aos seus requisitos metodológicos, conforme justificado a seguir.

É sempre um desafio obter uma amostra representativa da população de empresas, com qualidade e em número suficiente para rodar modelos estatísticos. Com efeito, para que a presente tese pudesse aportar contribuições acadêmicas e gerenciais relevantes: i) a amostragem deve atender aos preceitos estatísticos; ii) a base de dados deve ter qualidade e acurácia nas respostas; iii) a base de dados deve ter uma quantidade de respondentes suficiente para rodar os modelos estatísticos; iv) o conjunto de variáveis a serem coletadas deve apresentar abrangência e rigor. Os quatro pontos indicaram a direção de buscar uma coleta de dados primários, o que foi então feito nesta pesquisa.

Em relação ao plano de amostragem, a presente tese seguiu as técnicas de Levantamento Amostral, na qual a amostra é obtida a partir de uma população bem definida, através de protocolos controlados pelo pesquisador (Bussab & Morettin, 2010); esse levantamento foi não probabilístico, pois não há como forçar uma empresa a responder a um questionário, dependendo assim de participação intencional ou voluntária. Como recomendado em processos de amostragem não probabilístico, foi implementado um procedimento de controle de não respondentes para validar a seleção de empresas que formou a amostra principal, consultando inclusive a opinião de especialistas na área de empreendedorismo (Bussab e Morettin, 2010, p. 261).

Para fazer a pesquisa de campo, foram acionadas duas entidades de grande relevância no universo de empreendedorismo no Brasil, com o intuito de convidar as empresas a participarem da pesquisa: Endeavor e Startup Farm.

A Endeavor¹ é uma organização não governamental que atua como uma instituição de fomento ao empreendedorismo. Busca estimular o crescimento das empresas através de

¹ Mais informações sobre a Endeavor disponíveis em <https://endeavor.org.br/>

programas de transferência de conhecimento e experiência de mentores a empreendedores. Os mentores são executivos de grandes empresas, investidores ou empreendedores de sucesso. A Endeavor Brasil, que faz parte da Endeavor global, instituição presente em 19 países, é uma das principais organizações de apoio a empreendedores no Brasil (Rosa, 2017).

A Endeavor oferece, de forma gratuita, constante apoio aos empreendedores, através de seu portal, *newsletters*, palestras e cursos; quando realiza eventos, além da participação presencial, a instituição também disponibiliza a transmissão via internet (*webinars*), permitindo que empreendedores de todo o Brasil possam se beneficiar. Por esse atuante protagonismo na área de empreendedorismo, a Endeavor dispõe de um *mailing* muito grande de empreendedores ou interessados na área, que se cadastram de forma espontânea.

A Startup Farm² é uma aceleradora privada com foco em startups inovadoras, que podem provocar disrupção em seus respectivos mercados; oferece apoio de uma rede de mentores e parceiros experientes às empresas, com o intuito de acelerar os resultados e reduzir riscos.

Segundo dados coletados em seu site, possui portfólio com mais de 250 empresas e já captou mais de US\$ 100 milhões de investimentos por meio da *Farm Venture Capital*, ligada à entidade. Para fazer parte dos programas de aceleração da Startup Farm, as empresas passam por um processo de seleção; é muito conhecido no setor o Startup Farm Day, evento onde as empresas apreendem a refinar e conferir agilidade a seus planos de negócios.

Embora Endeavor e Startup Farm sejam entidades que atuem na área de empreendedorismo, as duas têm perfis diferentes em relação às empresas com as quais se relacionam; esse fato contribuiu para diversificar e enriquecer o processo de amostragem, mantendo o foco em empreendedorismo.

Definidas as entidades para o processo de coleta de dados, o plano de amostragem seguiu um rígido planejamento, tendo sido realizadas as seguintes etapas:

- i) envio inicial do questionário via internet, através das entidades Endeavor e Startup Farm, para milhares de empresas, no início de setembro/2017;
- ii) concessão de prazo limite de 15 dias para obtenção das respostas;

² Mais informações sobre a Startup Farm disponíveis em <http://www.startupfarm.com.br/>

- iii) durante esse período, foi enviado um segundo e-mail como *reminder* do prazo limite e estimular as empresas a participarem da pesquisa;
- iv) encerramento do período de coleta de dados no final de setembro/2017;
- v) separação das empresas respondentes de cada uma das duas entidades, de modo a avaliar se havia consistência nos padrões de dados das duas fontes;
- vi) criação da amostra principal com as empresas respondentes da Endeavor e da amostra de controle com as empresas respondentes da Startup Farm;
- vii) contatos com algumas empresas não respondentes, de forma aleatória, para checar se aplicado o mesmo questionário, os padrões de dados dessas firmas seria equivalente ao das amostras principal e de controle;
- viii) análise estatística dos dados das amostras principal e de controle para avaliar se os padrões de dados eram equivalentes;
- ix) validação da amostra principal, sem integrar os dados da amostra de controle.

De acordo com o plano de amostragem, o primeiro passo foi o envio dos milhares de e-mails para as empresas cadastradas nas entidades Endeavor e Startup Farm. Foi feita uma segmentação dos respectivos *mailings* das duas entidades, visando obter uma amostra mais homogênea e representativa de empresas.

No caso da Endeavor, foram selecionados cerca de 11.000 empreendedores cadastrados na entidade, responsáveis por gerenciar empresas com faturamento maior que R\$ 360 mil no ano de 2016; essa segmentação foi viável graças ao processo de captura cadastral, através de *quiz*³ em seu portal, onde é obrigatório informar o porte das empresas; vale ressaltar que esse cadastro deixa como opcional a informação do nome da empresa, por motivos de confidencialidade, mas exige a informação sobre o faturamento bruto em 2015 e 2016. Em relação à Startup Farm, o *mailing* selecionado para envio foi de 800 empreendedores que tiveram algum contato direto com a entidade.

Os e-mails com os respectivos links de acesso aos formulários da pesquisa foram endereçados diretamente para os empreendedores, convidando-os a participarem da pesquisa; os textos dos e-mails foram propositadamente similares, para garantir homogeneidade ao processo de coleta de dados das duas fontes.

³ Quiz disponível em <https://checkup.endeavor.org.br/#/landing>

Em nenhuma das duas entidades foi oferecido um incentivo material para os respondentes, tal como a possibilidade de participar de sorteio ou o compromisso dos pesquisadores em fazer alguma doação para entidades sociais por resposta recebida. Tais incentivos são muito utilizados em *surveys*, dadas a grande dificuldade de sensibilizar empreendedores muito atarefados e bastante assediados a responderem “mais uma pesquisa”. Endeavor e Startup Farm entendem que o relacionamento com os empreendedores cadastrados em seus respectivos *mailings* deve seguir o padrão usual, sem oferecer vantagens que configurariam uma conduta diferenciada para a presente pesquisa, criando um indesejado precedente; dessa forma, futuras pesquisas que Endeavor e Startup Farm venham a fazer continuarão seguindo a mesma diretriz de não oferecer incentivos materiais aos respondentes.

Para aumentar a taxa de respostas, Endeavor e Startup Farm enfatizaram no e-mail de encaminhamento que a pesquisa estava sendo conduzida em parceria com a FEA / USP, que o tempo para responder era cerca de 10 minutos e, sobretudo, que a participação de cada empresa seria muito importante para aprimorar os serviços de apoio das próprias entidades ao empreendedorismo no Brasil. Fato é que o prestígio e a reputação positiva das entidades Endeavor e Startup Farm, somadas ao aval acadêmico da FEA / USP, funcionaram de maneira efetiva na obtenção de uma adequada taxa de respostas. A Tabela 3 mostra a quantidade de empresas que receberam e responderam aos e-mails e as respectivas taxas.

Tabela 3 – Amostragem de Empresas por Entidade

	Empresas Convidadas	Empresas Respondentes	Taxas de Respostas
Endeavor	11.000	593	5,4%
Startup Farm	800	30	3,8%
Total	11.800	623	5,3%

Fonte: elaborado pelo autor.

Os períodos amostrais para a Endeavor e Startup Farm foram similares, com o início e término da coleta de dados das duas entidades praticamente simultâneos. O processo de coleta de dados teve início nos primeiros dias de setembro/2017 e foi encerrado na penúltima semana do próprio mês para as duas entidades. Essa janela amostral relativamente curta, obtida graças a uma efetiva coordenação do processo amostral, foi muito importante para obter uma base de dados mais homogênea e sem viés metodológico temporal; esse problema

amostral ocorre com frequência quando o processo de coleta de dados primários é longo e não simultâneo, ou ainda quando são usados dados secundários de diversas fontes.

Como mencionado anteriormente nesta seção, foi adotado um procedimento de controle de não respondentes, conforme preconiza Bussab e Morettin (2010, p. 261). Consistiu na consulta a algumas empresas não respondentes, de forma aleatória, onde foram checadas como teriam respondido ao mesmo formulário da pesquisa; constatou-se que os padrões de dados das empresas respondentes e não respondentes eram similares.

Como último passo, para checar a consistência dos padrões de dados originados pelas duas fontes, Endeavor e Startup Farm, as duas amostras foram separadas e analisadas. As 593 empresas respondentes através da Endeavor formaram a amostra principal, enquanto que as 30 empresas respondentes através da Startup Farm constituíram a amostra de controle. As análises estatísticas das duas amostras comprovaram que os padrões de dados das empresas de ambas as fontes eram equivalentes, exceto pelo fato que todas as 30 empresas tinham tido contato com a própria aceleradora Startup Farm. Essa constatação de similaridade foi confirmada em entrevistas com gerentes da Endeavor e Startup Farm, que atuaram como especialistas para dar consistência ao plano de amostragem.

Dessa forma, a amostra principal de 593 empresas foi considerada validada; conforme estabelecido no último passo do plano de amostragem, a amostra de controle com as 30 empresas foi descartada, dado que não é recomendável que a amostra principal inclua os respondentes da amostra de controle; portanto, a amostra principal da presente tese ficou constituída pelas 593 empresas respondentes através da Endeavor.

3.3 Regressão Logística Binária

Foram diversas as técnicas estatísticas utilizadas na análise de dados nesta pesquisa, desde as mais simples, como média, mediana, desvio-padrão, teste de comparação de médias (ANOVA) até outras de maior complexidade: Regressão Linear, Regressão Logística e Análise Fatorial. Em função da importância desses dois últimos modelos estatísticos para as análises de dados da presente tese, esta seção é dedicada a detalhar os procedimentos metodológicos inerentes à Regressão Logística Binária, enquanto que a seção 3.4 tem como foco a Análise Fatorial.

O modelo Regressão Logística, também chamado Logit, é indicado quando o fenômeno a ser observado apresenta-se de forma qualitativa, caracterizado pela natureza categórica da variável resposta (Fávero, 2015); no caso dessa variável resposta ou dependente apresentar duas categorias, o modelo é chamado Regressão Logística Binária ou Logit Binário; caso a variável resposta possa assumir três ou mais respostas categóricas, tem-se a Regressão Logística Multinomial ou Logit Multinomial. Trata-se de um modelo de dependência de natureza preditiva.

O objetivo do modelo Regressão Logística Binária é estimar a probabilidade de ocorrência de um fenômeno, em função de diversas variáveis preditoras, que podem ser tanto métricas quanto categóricas. Segundo Malhotra (2012), o modelo Regressão Logística Binária trata de estimar a probabilidade que uma observação pertença a cada um dos dois grupos. Segundo Hair et al. (2010), é uma forma de regressão especializada que é formulada para prever e explicar uma variável categórica de dois grupos, em vez de uma variável dependente de natureza métrica.

Com exemplo, o modelo Regressão Logística Binária foi utilizado na pesquisa realizada por Van Gelderen et al. (2006) para identificar fatores de sucesso e riscos de startups em sua fase bem inicial; os autores estudaram empresas holandesas e concluíram que algumas variáveis foram relevantes para explicar o sucesso, avaliado pela sobrevivência versus não sobrevivência – variável resposta categórica binária – após três anos.

Especificamente na presente tese, dado que a variável resposta advinda da separação das empresas foi em dois grupos – EACs e não EACs – conforme detalhado na seção 3.1, configurou-se a indicação do modelo Regressão Logística Binária. Buscou-se estimar a probabilidade de uma empresa ser classificada como de alto crescimento ou não, em função das variáveis explicativas tais como: localização da empresa, número de sócios, participação de investidores, grau de inovação, dentre várias outras. A lista completa de variáveis está descrita na seção 3.5 Lista de Variáveis.

A técnica Regressão Logística Binária é elaborada com base na estimação por máxima verossimilhança, e passa pela modelagem estatística descrita nos parágrafos seguintes. A equação linear (3) que introduz o modelo é:

$$Z_i = \alpha + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \dots + \beta_k * X_{ki} \quad (3)$$

Nessa equação, Z_i é chamado *logito*, α representa a constante, β_j ($j = 1, 2, \dots, k$) são os parâmetros que serão estimados para as variáveis explicativas, e X_{ji} são as variáveis explicativas, que podem ser métricas ou dummies (Fávero, 2015). Diferentemente do modelo Regressão Linear Múltipla, o logito Z não é diretamente a variável dependente Y , mas a base para se chegar a ela; para tanto, é necessário definir a expressão de probabilidade p_i , de ocorrência do evento considerado sucesso, em função do logito Z_i , a ser estimado em função das variáveis explicativas. O modelo trabalha com o conceito de chance, conhecido internacionalmente por *odds* (Fávero, 2015), expressão adotada na presente tese, e segue a equação (4):

$$odds_{Y_i=1} = \frac{p_i}{1-p_i} \quad (4)$$

Embora odds e probabilidade sejam conceitos interligados, são calculados de forma diferente; probabilidade é calculada dividindo-se o número de eventos de sucesso pelo número total de eventos possíveis; odds é calculado dividindo-se a probabilidade de sucesso pela probabilidade de não sucesso; quando existem apenas dois resultados possíveis, sucesso e não sucesso, a probabilidade de não sucesso é o complemento para 1 ou 100% da probabilidade de sucesso, resultando então na equação (4) anterior.

Na Regressão Logística Binária é definido o logito Z como o logaritmo natural ou neperiano do odds; o logaritmo natural utiliza como base o número irracional e , conhecido como número de Euler ou de Napier, e é igual a 2.718281828...; assim, tem-se a equação (5):

$$\ln(odds_{Y_i=1}) = Z_i \quad (5)$$

Com base na equação (5) anterior, podem ser escritas as equações (6) e (7) a seguir:

$$\ln\left(\frac{p_i}{1-p_i}\right) = Z_i \quad (6)$$

$$\frac{p_i}{1-p_i} = e^{Z_i} \quad (7)$$

A partir da equação (7), obtém-se a equação da probabilidade p_i , conforme equação (8):

$$p_i = \frac{e^{Z_i}}{1 + e^{Z_i}} = \frac{1}{1 + e^{-Z_i}} \quad (8)$$

Dessa forma, através de uma modelagem relativamente simples, análoga à Regressão Linear Múltipla, são estimados os coeficientes do logito Z , que são os expoentes do e , gerando a estimativa de probabilidades a partir do odds de uma observação pertencer ou não a cada um dos dois grupos. Por se tratar de função exponencial, a curva de probabilidades tem a forma de S, varia entre 0 e 1 como esperado, enquanto o odds, que é o exponencial de e pelo logito Z pode variar entre 0 e $+\infty$. A curva que o modelo teórico gera aparece ilustrada na Figura 3.

Figura 3 – Curva de Probabilidade do Modelo Regressão Logística Binária

Fonte: adaptado de Fávero, 2015.

Na Regressão Logística Binária a variável dependente segue uma Distribuição de Bernoulli; essa distribuição é caracterizada quando determinada observação i ter sido classificada como evento de interesse, tem a probabilidade de ocorrência p_i e a probabilidade de ocorrência do não evento é igual a $(1 - p_i)$. A estimação dos parâmetros β_j é feita pelo método da máxima verossimilhança, razão pela qual o único pressuposto importante a ser observado no modelo Regressão Logística Binária é a ausência de multicolinearidade entre as variáveis explicativas, de forma oposta à técnica de Análise Fatorial, descrita na seção 3.4.

Para rodar o modelo Regressão Logística Binária foi utilizado o software estatístico STATA®, da empresa StataCorp LP (College Station, Texas), em sua versão número 15. Esse é um software de análise de dados e de estatística completo e rápido, que permite fazer não somente a aplicação requerida, mas produzir relatórios e gráficos que serão muito valiosos na interpretação dos achados.

O uso de STATA® foi ainda mais adequado para esta tese, pois além das análises básicas dispõe do comando *stepwise* para depurar o modelo final, retirando as variáveis sem significância estatística com grande precisão e rigor estatístico. A partir do modelo final é possível gerar a variável probabilidade para cada observação da amostra, usando o comando *predict* do STATA® e na sequência uma ampla variedade de gráficos com os resultados da modelagem. A visualização gráfica do modelo Regressão Logística, que mostra a curva de probabilidade em função de variáveis preditoras, é muito simples de entender e interpretar, sendo de grande valor acadêmico e gerencial.

A Regressão Logística Binária estimada nesta pesquisa se valeu de extensa lista de variáveis explicativas que foram detalhadas na seção 3.5; a variável resposta – EACs e não EACs – é de natureza categórica nominal, enquanto que as variáveis explicativas podem ser tanto métricas quanto categóricas; neste último caso, as variáveis categóricas são transformadas em variáveis do tipo dummy (Fávero, 2015).

A Regressão Logística Binária aplicada às variáveis desta pesquisa passou nos testes estatísticos para aferir se o modelo é ou não apropriado a um determinado nível de significância. O primeiro deles, o teste do χ^2 (qui-quadrado), indica que pelo menos um dos coeficientes β_j das variáveis explicativas é diferente de zero. Na sequência foram identificados quais são os parâmetros β_j que passaram no teste individual de cada variável, através da estatística z de Wald, aos níveis de significância de 5% ou 10%, que foram adotados na presente tese.

Além dos testes acima, outros três parâmetros foram analisados para verificar se o modelo Regressão Logística Binária era adequado para explicar o problema de pesquisa desta tese. O primeiro foi o chamado Pseudo R^2 de McFadden, similar ao R^2 do modelo Regressão Linear, porém com fortes diferenças e limitações; no modelo logístico, dado que a variável resposta é categórica, não faz muito sentido medir o percentual de sua variância que pode ser explicado

pelas variáveis preditoras; o Pseudo R^2 de McFadden pode servir apenas para comparar dois modelos alternativos (Fávero, 2015).

O segundo, bem mais robusto para avaliar a qualidade do ajuste do modelo, foi a elaboração e análise da Matriz de Classificação, que é a aplicação do próprio modelo preditivo aos dados da amostra com o intuito de apurar a percentagem de observações que foram classificadas corretamente pelo próprio modelo; foram calculados os dois percentuais de acerto, para evento e não evento, denominados respectivamente Sensitividade e Especificidade. A classificação do modelo é feita a partir de uma probabilidade limite, denominada *cutoff*, que é escolhida pelo pesquisador, para classificar como evento e não evento as observações da amostra, a partir das estimativas de probabilidades geradas pelo próprio modelo. Para ilustrar com um exemplo, se o *cutoff* escolhido for 40%, o modelo vai estimar a probabilidade de cada observação, com base nos coeficientes das variáveis, e classificar como evento quando a estimativa for igual ou superior a 40%; após essa classificação do modelo aplicado às observações da amostra, é feita uma comparação com a incidência real, ou seja, com os eventos da própria amostra, identificando os acertos e erros da previsão.

O terceiro instrumento de aferição da qualidade do modelo utilizado na presente tese foi a análise do gráfico conhecido como curva *ROC*, sigla de *Receiver Operating Characteristics*; esse gráfico apresenta a variação de sensibilidade no eixo das ordenadas, em função de (1 – especificidade) no eixo das abscissas, ou seja, ilustra de forma gráfica uma espécie de *trade off* entre sensibilidade e especificidade, quando da alteração de *cutoff*, apresentando formato convexo em relação ao ponto (0, 1). Portanto, quanto maior a área abaixo da curva *ROC*, que significa mais próxima de 1, maior a eficiência global da previsão; esse indicador também serve para comparar vários modelos (Fávero, 2015).

O modelo Regressão Logística Binária permitiu fazer diversas análises e simulações, após a definição dos vários modelos finais, usando o comando *predict* em STATA®, que estimou a probabilidade de cada empresa pertencer ao grupo de empresas de alto crescimento ou não. Diversos gráficos puderam ser feitos em função das principais variáveis explicativas que se mantiveram significantes no modelo final. Algumas análises complementares com o subgrupo de empresas conhecidas como gazelas, que são as EACs com até cinco anos de idade, também foram executadas. Nesse caso, a variável resposta continua sendo categórica, com duas possíveis categorias, gazelas ou não gazelas.

Como detalhado no capítulo 4, diversos resultados foram obtidos a partir da aplicação do modelo Regressão Logística Binária, com a variável resposta EACs ou não EACs; a variável resposta gazelas ou não gazelas gerou alguns resultados adicionais. As análises do banco de dados da presente pesquisa, elaboradas com a descrita modelagem estatística, revelaram achados acadêmicos e gerenciais significantes, na avaliação dos *drivers* de crescimento das empresas.

3.4 Análise Fatorial

Embora a Regressão Logística Binária, detalhada na seção anterior, seja o modelo estatístico de análise de dados mais relevante utilizado na presente tese, a técnica chamada Análise Fatorial ajudou a interpretar a percepção dos empreendedores em relação à importância de determinadas ferramentas de gestão, tais como inovação em produtos e serviços, investimentos em marketing, marcas e patentes, obtenção de capital de investidores, dentre outras; por esse motivo a Análise Fatorial está detalhada na presente seção.

Análise Fatorial é um modelo estatístico que faz parte do campo de técnicas multivariadas exploratórias, também chamadas técnicas de interdependência, que visam entender a relação entre as variáveis de um determinado banco de dados, sem a intenção de buscar um modelo confirmatório ou preditivo para outras observações que não as que formam a amostra. Assim, diferente de técnicas de dependência, como Regressão Linear ou Regressão Logística, onde existe uma variável resposta, e busca-se um modelo ou equação que permita fazer inferências estatísticas sobre o fenômeno em estudo, as técnicas de interdependência visam simplesmente entender como as variáveis se comportam e se existe correlação entre elas, de modo a permitir uma eventual redução ou simplificação estrutural dos dados (Fávero & Belfiore, 2015).

Assim, Análise Fatorial é o nome de uma classe de procedimentos utilizados essencialmente para a redução do número de variáveis, sempre que houver elevado coeficiente de correlação entre elas, característica conhecida como multicolinearidade; em pesquisas de marketing é frequente haver inúmeras variáveis que podem ser correlacionadas ou estejam avaliando construtos similares, que poderiam ser agrupadas em certos fatores subjacentes (Malhotra, 2012). O objetivo primário da Análise Fatorial é definir a estrutura básica entre as variáveis em análise, identificando a estrutura de interdependência entre um grande número de variáveis (ex. respostas de questionários), estabelecendo conjuntos de variáveis que são altamente correlacionadas, denominadas fatores (Hair et al., 2010).

Existem vários métodos para a determinação de fatores e nesta tese foi utilizada a Análise Fatorial por Componentes Principais, que tem como característica extrair fatores não correlacionados e, portanto, ortogonais entre si a partir de combinações lineares das variáveis originais (Fávero & Belfiore, 2015). Outra vantagem desse método é que leva em consideração a variação total dos dados; modelos alternativos, tais como Análise Fatorial por Fator Comum ou Análise Fatorial por Máxima Verossimilhança, não são indicados para o problema de pesquisa desta tese (Malhotra, 2012).

São quatro os objetivos da Análise Fatorial por Componentes Principais: i) a redução estrutural das variáveis, pela identificação de correlações entre elas; ii) verificação da validade dos construtos previamente estabelecidos; iii) elaboração de rankings por meio da criação de indicadores de desempenho econômico e financeiro a partir dos fatores; iv) estimar fatores ortogonais, eliminando assim a multicolinearidade (Fávero & Belfiore, 2015).

A utilização de Análise Fatorial por Componentes Principais na presente tese foi embasada no primeiro e no quarto objetivos do parágrafo anterior – redução estrutural das variáveis e estimar fatores ortogonais – aplicada às variáveis que mediram a percepção dos empreendedores sobre diversas ferramentas de gestão de empresas. Estabelecido que o método Análise Fatorial por Componentes Principais é o mais indicado para a redução estrutural de variáveis e estimação de fatores ortogonais, o mesmo passou a ser referido apenas como Análise Fatorial na presente tese.

As 16 variáveis que foram utilizadas na Análise Fatorial estão detalhadas na seção 3.5 do presente capítulo. Pode-se antecipar que algumas dessas variáveis apresentaram características de multicolinearidade entre elas. Todas essas variáveis são métricas do tipo *ratio* (razão), para as quais uma análise baseada nos coeficientes de correlação de Pearson pode ser corretamente empregada. Nenhuma das variáveis tem as características de intervalar, como as variáveis em Escala Likert, muito frequentemente utilizadas em pesquisas quantitativas. Não se deve fazer uma Análise Fatorial utilizando variáveis medidas em Escala Likert, pois isso, segundo Fávero e Belfiore (2015), é uma “ponderação arbitrária que levaria a erro grave!”.

Embora esse tema seja objeto de controvérsias e inúmeros autores utilizam o procedimento de Análise Fatorial com um banco de dados formado por variáveis em Escala Likert (ex. Malhotra, 2012), o uso de variáveis em escala de 0 a 10 pontos conferiu maior consistência

estatística para esta pesquisa; esse desejável rigor metodológico contribuiu para a opção por uma base de dados primários.

A robustez estatística de uma Análise Fatorial deve ser avaliada a partir dos métodos de esfericidade de Bartlett e da estatística Kaiser-Meyer-Olkin (KMO); o teste de esfericidade de Bartlett compara a matriz de correlações obtida na análise com a matriz identidade (matriz que tem 1 na diagonal e zero nas demais células) de mesma dimensão; se as correlações obtidas na Análise Fatorial são diferentes de zero, significa que a extração de fatores é adequada. A estatística KMO indica a proporção de variância considerada comum às variáveis da amostra, atribuível a um fator comum; varia entre 0 e 1; as variáveis que compartilham um percentual elevado, ou seja, que apresentam alto coeficiente de correlação de Pearson, apresentam KMO próximo de 1, o que indica boa adequação para a Análise Fatorial. A Tabela 4 mostra a relação entre a estatística KMO e a adequação global da Análise Fatorial, conforme critério amplamente aceito na literatura (Fávero & Belfiore, 2015).

Tabela 4 – Relação entre KMO e Adequação da Análise Fatorial

Estatística KMO	Adequação Global
Entre 1,00 e 0,90	Muito boa
Entre 0,90 e 0,80	Boa
Entre 0,80 e 0,70	Média
Entre 0,70 e 0,60	Razoável
Entre 0,60 e 0,50	Má
Menor que 0,50	Inaceitável

Adaptado de Fávero e Belfiore, 2015.

Para rodar o modelo Análise Fatorial foi utilizado o software estatístico STATA®, da empresa StataCorp LP (College Station, Texas), versão 15; o software é adequado para elaborar esse tipo de análise, gerando as tabelas com as variáveis que podem ser agrupadas; escolhido o número de fatores a serem retidos, as novas variáveis métricas – fatores ortogonais – são estimadas para todas as observações do banco de dados.

Após passar nos testes estatísticos que comprovem a adequação da Análise Fatorial, os fatores ortogonais retidos, com Eigenvalues superiores a 1, conforme critério de Kaiser, podem ser gerados; tais fatores são novas variáveis métricas, que podem substituir as variáveis originais do banco de dados. Como cada fator representa uma combinação linear das variáveis originais do banco de dados, o modelo gera os escores fatoriais, que são os coeficientes de uma

Regressão Linear Múltipla para cada variável (Fávero e Belfiore, 2015); a título de exemplo, supondo um modelo com k variáveis, a Análise Fatorial estima os seguintes k coeficientes s_i da equação linear múltipla (9) para o Fator 1:

$$F_1 = s_1 * x_1 + s_2 * x_2 + \dots + s_k * x_k \quad (9)$$

A partir dos escores fatoriais são gerados os fatores, que são novas variáveis métricas, para cada observação da amostra; esses fatores são ortogonais, portanto, com coeficiente de correlação de Pearson igual a zero, eliminando a multicolinearidade das variáveis originais; são também variáveis padronizadas com média igual a zero e desvio-padrão igual a 1. Na presente tese, as 16 variáveis de percepção dos empreendedores puderam ser agrupadas em uma quantidade reduzida de fatores, a partir dos quais foram feitas análises complementares; essas análises a partir dos fatores gerados serviram para comparações com as anteriores, feitas diretamente com as variáveis originais, de modo a confirmar a consistência dos achados e dar maior validade estatística à pesquisa. A própria interpretação dos fatores e das variáveis que cada um deles representa, identificando as principais dimensões dos dados originais, ajuda a entender como os empreendedores percebem a importância das diversas ferramentas de gestão.

3.5 Lista de Variáveis

O conjunto completo de variáveis coletadas nesta pesquisa está detalhado nesta seção; de um modo geral, as variáveis do banco de dados primários da pesquisa podem ser agrupadas em cinco blocos:

- i) variáveis que permitam estimar os CAGRs das empresas em termos de faturamento e de número de colaboradores;
- ii) variáveis demográficas, tais como tempo de existência da empresa, localização ou número de sócios;
- iii) variáveis de percepção da importância atribuída pelos empreendedores às diversas ferramentas de gestão, tais como inovação em produtos e serviços, investimentos em marketing, marcas e patentes;

- iv) variáveis indicativas do relacionamento e aporte de capital de variadas fontes, tais como investidores *angel* ou bancos;
- v) variáveis sobre o perfil do empreendedor.

As variáveis utilizadas para estimar os CAGRs, utilizados na classificação das empresas em dois grupos, as de alto crescimento (EACs) e as demais (não EACs), conforme explicado na seção 3.1, não devem fazer parte do conjunto de variáveis do modelo preditivo Regressão Logística Binária; são variáveis exclusivamente utilizadas na criação da nova variável categórica que classifica as empresas respondentes da amostra em EACs ou não EACs, ou seja, da variável dependente ou resposta para o modelo Regressão Logística Binária.

O formulário completo utilizado na pesquisa, com todas as variáveis coletadas, está detalhado no Apêndice 1. Trata-se de um formulário construído com base na revisão bibliográfica da presente tese e posteriormente revisto por especialistas da Endeavor; antes de ser enviado, o formulário passou ainda por pré-testes com algumas empresas, que forneceram subsídios para que fosse aprimorado. No processo de revisão foram identificados eventuais problemas de entendimento das questões, dificuldades para o preenchimento dos dados solicitados, omissões e aspectos que puderam ser suprimidos ou acrescentados.

A Tabela 5 a seguir detalha o conjunto de variáveis do primeiro grupo, utilizadas para estimar os CAGRs das empresas.

Tabela 5 – Variáveis de Crescimento de Empresas

Nome	Variáveis	Tipo
cfat_14	crescimento do faturamento em 2014 em comparação com 2013 (em %)	métrica
cfat_15	crescimento do faturamento em 2015 em comparação com 2014 (em %)	métrica
cfat_16	crescimento do faturamento em 2016 em comparação com 2015 (em %)	métrica
cfat_17	crescimento do faturamento previsto em 2017 em comparação com 2016 (em %)	métrica
col_13	número de colaboradores em 31/dez/2013	métrica
col_14	número de colaboradores em 31/dez/2014	métrica
col_15	número de colaboradores em 31/dez/2015	métrica
col_16	número de colaboradores em 31/dez/2016	métrica
col_17	número de colaboradores previsto em 31/dez/2016	métrica

Fonte: elaborado pelo autor.

É importante salientar que a coleta de dados das variáveis para estimar os CAGRs das empresas não ofereceu aos respondentes faixas de faturamento ou faixas de número de colaboradores, predefinidas no formulário de *survey*; o uso de faixas transforma uma variável

métrica do tipo *ratio*, em uma variável categórica, limitando sua utilização. É frequente pesquisas incorrerem em erros metodológicos ao coletarem dados quantitativos através de faixas predefinidas, e posteriormente transformarem as respostas em variáveis métricas, utilizando o valor médio de cada faixa; embora esses erros de arredondamento possam ter impacto limitado, é recomendável que a coleta de dados reflita números realistas, permitindo inclusive a identificação e eventual eliminação de *outliers*.

Deve-se salientar ainda que a forma de coleta de dados de faturamento foi diferente da aplicada ao número de colaboradores. Os dados de faturamento foram coletados através das respectivas taxas de crescimento anuais, enquanto que os dados sobre número de colaboradores foram coletados de forma direta, com a fotografia da variável ao final de cada ano.

Essa conduta foi indicada por especialistas em empreendedorismo, inclusive por membros da banca de qualificação da presente tese. Por trás dessa recomendação está a grande dificuldade em obter dados de faturamento real de empresas por meio de *surveys*; este fato está embasado em razões concretas e objetivas, como a exigência contratual de sócios investidores (*angel*, *venture capital*, *private equity*, etc...) em vetar que as empresas revelem valores faturados, e razões culturais ou psicológicas, como receio em divulgar dados considerados confidenciais. Deve-se enfatizar que, para aumentar a acurácia dos dados coletados, os nomes das empresas e dos respondentes foram solicitados, o que reduz ainda mais as chances em obter os respectivos faturamentos anuais.

A dificuldade em fornecer dados de faturamento das empresas não se repete em relação ao número de colaboradores, informado sem maiores restrições. Essa postura mais transparente dos empreendedores na divulgação do número de colaboradores deve ocorrer por aspectos culturais, tais como: é comum comunicar o dado, interna e externamente, de forma ampla, inclusive em site da empresa, está diretamente associado ao sucesso, é valorizado pela sociedade, não costuma ser considerado confidencial, dentre outros.

Com base nessa forma de reação dos empreendedores às solicitações de dados sobre faturamento e número de colaboradores, o formulário de pesquisa foi concebido solicitando as taxas de crescimento anuais do faturamento e do número absoluto de colaboradores ao final de cada ano. A coleta de dados mostrou que a solução encontrada funcionou adequadamente, seja porque houve grande adesão dos respondentes, seja porque para estimar os CAGRs das

empresas bastariam as taxas de crescimento anuais de faturamento, prescindindo dos dados absolutos.

A mesma abordagem não pode ser estendida para a coleta de dados do número de colaboradores uma vez que o critério de inclusão de EACs, como detalhado na seção 3.1, exige que a empresa atinja CAGR maior que 20% entre os anos xx e $xx-3$, medido em termos de faturamento ou número de colaboradores, desde que a empresa tenha no ano $xx-3$ pelo menos 10 colaboradores; ou seja, é necessário saber o número absoluto de colaboradores no ano base $xx-3$ e não apenas as taxas de crescimento do número de colaboradores.

A escolha de coletar dados apenas dos últimos cinco anos, dentre os quais um deles é 2017, o próprio ano em curso, com dados previstos para o encerramento do ano, foi discutida e aprovada por especialistas da Endeavor; dentre os principais motivos, pode-se citar: maior atualidade dos dados coletados, maior acurácia dos dados em função da recência, maior facilidade para os respondentes; esse recorte temporal restritivo aos últimos cinco anos, aliado ao fato que o Brasil vem passando no período por uma crise econômica, foi determinante para aumentar o rigor metodológico na identificação de EACs.

Com base nas duas formas de coleta de dados de faturamento e número de colaboradores e nos dois períodos de três anos disponíveis para estimar os quatro CAGRs, as equações finais para identificar as EACs, com base nos nomes das variáveis da Tabela 5 da presente tese, estão detalhadas nas equações (10) a (13) a seguir:

CAGR de faturamento da empresa para o período 2013 – 2016

$$\sqrt[3]{(1 + cfat_{14}) * (1 + cfat_{15}) * (1 + cfat_{16})} - 1 > 20\% \quad (10)$$

CAGR de faturamento da empresa para o período 2014 – 2017 (estimado)

$$\sqrt[3]{(1 + cfat_{15}) * (1 + cfat_{16}) * (1 + cfat_{17})} - 1 > 20\% \quad (11)$$

CAGR de número de colaboradores da empresa para o período 2013 – 2016

$$\sqrt[3]{\frac{col_{16}}{col_{13}}} - 1 > 20\% \quad (12)$$

CAGR de número de colaboradores da empresa para o período 2014 – 2017 (estimado)

$$\sqrt[3]{\frac{col_{17}}{col_{14}}} - 1 > 20\% \quad (13)$$

Deve-se levar em consideração que para uma firma ser considerada uma EAC, pelo critério da OECD (2007), além de apresentar pelo menos um dos quatro CAGRs maior que 20%, o número mínimo de colaboradores no ano $xx-3$, que no caso da presente pesquisa são os anos 2013 ou 2014, dependendo da estimativa utilizada, precisa ser igual ou superior a 10, conforme explicado na seção 3.1 deste capítulo.

A definição final de EACs passou por mais um filtro de consistência relacionado com a data de início de suas operações. As empresas com data do primeiro faturamento posterior ao primeiro dia útil de cada ano, tecnicamente não poderiam estimar a taxa de crescimento do ano seguinte em comparação com o ano anterior; a título de exemplo, se o primeiro faturamento de uma empresa ocorreu em novembro/2015, como apurar a taxa de crescimento de um período de 12 meses em 2016 em comparação com o faturamento de apenas dois meses de 2015? Assim, apenas as empresas que tenham faturamento durante o ano completo, ou seja, desde janeiro, podem estimar corretamente a taxa de crescimento do ano seguinte em relação ao anterior.

Dessa maneira, foi aplicado esse filtro, recomendado pela OECD (2007, p. 62), para identificar empresas a serem excluídas. Seguindo rigorosamente tal diretriz, a presente tese descartou da amostra principal todas as empresas que tivessem a data do primeiro faturamento posterior a 01/março/2014. A escolha da data de corte em março e não janeiro/2014 foi em função da grande sazonalidade que ocorre no Brasil nos meses de janeiro e fevereiro, atribuída a: alto verão, concentração de férias escolares e coletivas de empresas, redução de dias úteis em função do calendário de fevereiro e grande número de feriados. Dado que as

operações nos dois meses são limitadas e de pequeno impacto, considerou-se que somente as empresas com pelo menos 10 meses de faturamento no ano de 2014 fariam parte da amostra; assim, as empresas da pesquisa tiveram pelo menos quatro anos completos de operações.

Da amostra principal, formada por 593 respondentes, 123 empresas foram eliminadas, em função da data de corte de início de faturamento. A amostra com as empresas que passaram pelo filtro temporal – *amostra final* – ficou com 470 empresas, formando a base de dados utilizada nas análises estatísticas na presente pesquisa.

Não foi necessário fazer nenhum filtro de exclusão adicional para a métrica crescimento do número de colaboradores, uma vez que todas as empresas que apresentaram crescimento de faturamento em 2014, em comparação com 2013, também possuíam uma estrutura organizacional na data de 31/dezembro/2013. Assim, as 470 empresas formaram a *amostra final*, que passou a ser designada apenas como *amostra* nesta pesquisa.

A partir das nove variáveis constantes na Tabela 4, foi possível estimar os dois CAGRs de faturamento e os dois CAGRs de número de colaboradores para as 470 empresas da amostra, conforme as diretrizes da OECD (2007). Uma firma foi considerada uma EAC somente quando pelo menos um dos quatro CAGRs fosse superior a 20% e a empresa apresentasse 10 ou mais colaboradores no ano inicial *xx-3* do período apurado.

Foram muitas as variáveis preditoras do modelo Regressão Logística Binária, conforme detalhado no início desta seção, podendo ser agregadas como: demográficas, percepção da importância atribuída pelos empreendedores a ferramentas de gestão, relacionamentos, aportes de capital e perfis.

A Tabela 6 a seguir apresenta as 20 variáveis preditoras demográficas coletadas na pesquisa com as empresas. A maioria dessas variáveis é categórica nominal, com respostas na forma de textos descritivos, tais como nome da empresa, CNPJ ou breve descrição dos produtos e serviços comercializados da empresa. O formulário de pesquisa oferecia, quando conveniente e possível, algumas respostas pré-definidas, sobretudo do tipo sim ou não, que foram posteriormente transformadas em variáveis dummies.

Tabela 6 – Variáveis Predictoras Demográficas

Nome	Variáveis	Tipo
nom_res	Nome e sobrenome do respondente da pesquisa	categórica
cargo	Cargo do respondente da pesquisa na empresa	categórica
e_mail	e-mail do respondente da pesquisa (opcional)	categórica
nom_emp	Nome fantasia da sua empresa	categórica
cnpj	CNPJ da sua empresa	categórica
cidade	Cidade onde fica a sede da sua empresa	categórica
sede_at	Local onde fica a sede atual da sua empresa	categórica
instit	Nome da instituição onde fica a sede da sua empresa	categórica
mudou	Sua empresa mudou alguma vez de endereço?	categórica
sede_an	Local onde ficava anteriormente a sede da sua empresa	categórica
produto	Produtos e serviços da sua empresa	categórica
mod_neg	Modelo de negócios atual da sua empresa	categórica
pub_alv	Públicos-alvo da sua empresa	categórica
dat_fun	Data de fundação da sua empresa	métrica
pri_fat	Data do primeiro faturamento da sua empresa	métrica
homem	Quantidade e gênero dos sócios da sua empresa [Homens]	métrica
mulher	Quantidade e gênero dos sócios da sua empresa [Mulheres]	métrica
pos_gra	Formação acadêmica dos sócios da sua empresa [pós-graduação]	métrica
grad	Formação acadêmica dos sócios da sua empresa [graduação]	métrica
n_grad	Formação acadêmica dos sócios da sua empresa [sem graduação]	métrica

Fonte: elaborado pelo autor.

As variáveis sobre a localização da sede atual, bem como da sede anterior da empresa oferecia aos respondentes a possibilidade de selecionar algumas alternativas pré-determinadas. Conforme detalhado no formulário de pesquisa completo no Apêndice 1, as alternativas pré-determinadas eram:

- i) Parque tecnológico privado
- ii) Parque tecnológico público
- iii) Incubadora privada
- iv) Incubadora pública
- v) Espaço de coworking
- vi) Escritório próprio
- vii) Outros (detalhar)

Muitas empresas optaram por preencher a alternativa “Outros”, quando então era solicitado que complementassem a resposta com breve descrição de onde estão sediadas. A coleta de dados sobre a formação acadêmica dos sócios limitou-se a levantar a quantidade de sócios por

gênero e em cada categoria de formação acadêmica; para ambos os detalhes do número de sócios, o formulário de pesquisa oferecia tabelas simplificadas indicando as alternativas: 1, 2, 3, 4 e 5 ou mais. A informação sobre a data do primeiro faturamento é uma das mais importantes desse bloco, pois a partir dela foi possível estabelecer o critério de exclusão de forma rigorosa, e a consequente estimativa de CAGRs adequados.

As 16 variáveis com a percepção dos empreendedores sobre a importância de determinadas ferramentas de gestão de empresas estão detalhadas na Tabela 7. Deve-se salientar que para cada uma delas era oferecida uma escala que variava entre 0 e 10, sendo 0 pouco importante e 10 muito importante; os nomes das variáveis começam todos por “l_”, de forma a ajudar na interpretação das análises feitas nesta pesquisa.

Tabela 7 – Variáveis Preditores de Percepção dos Empreendedores

Nome	Variáveis	Tipo
l_prod	Inovar ou diferenciar produtos e serviços	métrica
l_desig	Inovar ou diferenciar design de produtos e serviços	métrica
l_mod	Inovar ou diferenciar modelo de negócios	métrica
l_p_neg	Elaborar e implementar Plano de Negócios	métrica
l_p_mkt	Elaborar e implementar Plano de Marketing	métrica
l_m_tra	Fazer propaganda em mídia tradicional (TV, rádio, revistas...)	métrica
l_m_dig	Fazer propaganda em mídia digital (internet, site, smartphone, aplicativos...)	métrica
l_soc_e	Ter sócios e diretores experientes em gestão de empresas	métrica
l_parce	Montar uma rede de parceiros	métrica
l_gd_cl	Conquistar grandes empresas como clientes	métrica
l_ecoss	Sediar a empresa em local com ecossistema favorável	métrica
l_marca	Registrar as marcas comerciais de sua empresa	métrica
l_pat	Patentear os produtos e serviços de sua empresa	métrica
l_inves	Obter capital de investidores	métrica
l_e_ban	Obter empréstimos bancários	métrica
l_d_fom	Obter doações de instituições de fomento ou de outras empresas	métrica

Fonte: elaborado pelo autor.

Conforme detalhado na seção 3.4 sobre Análise Fatorial, as 16 variáveis de percepção sobre a importância atribuída pelos empreendedores às diversas ferramentas de gestão, frequentemente utilizadas nas empresas, deve apresentar multicolinearidade; por essa razão, foi objeto de redução estrutural e criação de fatores ortogonais, através de Análise Fatorial. É de certo modo intuitivo, mesmo antes de apresentar os resultados da Análise Fatorial, que algumas dessas variáveis propositadamente medem construtos ou dimensões similares. Entender como se agrupam tais variáveis e a correlação delas com as empresas de alto

crescimento, através de um melhor conhecimento sobre a percepção dos empreendedores, é uma das contribuições relevantes da presente tese.

Entretanto, diversas análises foram também feitas com as 16 variáveis originais, uma vez que tal conduta é importante não só para identificar quais delas ajudaram diretamente a explicar as EACs, mas também explorar e comprovar se há consistência entre as variáveis originais separadas e o conjunto de fatores gerados a partir da Análise Fatorial. A mera redução do número de variáveis nas principais dimensões que as agrupam ajuda a interpretar o comportamento do empreendedor diante das diversas ferramentas de gestão.

As 15 variáveis indicativas do relacionamento das empresas com os diversos *stakeholders* que atuam no campo do empreendedorismo compõem a Tabela 8.

Tabela 8 – Variáveis Preditoras de Relacionamento das Empresas

Nome	Variáveis	Tipo
s_spin	Sua empresa é spin-off de alguma instituição ou de outra empresa?	categórica
spin_de	Nome da instituição ou da outra empresa da qual sua empresa é uma spin-off	categórica
s_r_ace	Sua empresa mantém ou manteve relacionamento com alguma aceleradora?	categórica
cap_soc	Sua empresa recebeu aporte de capital de: Sócios?	categórica
cap_fam	Sua empresa recebeu aporte de capital de: Familiares ou amigos?	categórica
cap_crow	Sua empresa recebeu aporte de capital de: Crowdfunding?	categórica
cap_ang	Sua empresa recebeu aporte de capital de: Investidores Angel?	categórica
cap_ace	Sua empresa recebeu aporte de capital de: Aceleradoras?	categórica
cap_sem	Sua empresa recebeu aporte de capital de: Fundos de Capital Semente?	categórica
cap_ven	Sua empresa recebeu aporte de capital de: Fundos de Venture Capital?	categórica
cap_pri	Sua empresa recebeu aporte de capital de: Fundos de Private Equity?	categórica
e_ban	Sua empresa recebeu aporte de capital de: Empréstimos Bancários?	categórica
s_doac	Sua empresa recebeu doação de alguma instituição ou empresa?	categórica
doa_de	Nomes das instituições ou empresas que fizeram doação para sua empresa	categórica
s_premi	Sua empresa recebeu alguma premiação?	categórica

Fonte: elaborado pelo autor.

Por último, a Tabela 9 trás as variáveis que buscaram avaliar o perfil dos empreendedores, de forma direta e objetiva. São quatro variáveis, que foram recomendadas pela Endeavor para compor o formulário de pesquisa, com base em *surveys* anteriores feitos pela instituição. Algumas análises complementares puderam ser feitas com essas variáveis, conforme descrito no capítulo 4 desta tese.

As duas variáveis métricas sobre a avaliação da vida foram medidas em uma escala de 0 a 10, sendo 0 a avaliação para “pior vida” e 10 para “melhor vida”; a variável métrica da auto avaliação do perfil empreendedor também utilizou escala de 0 a 10, onde 0 era “pouco

empreendedor” e 10 “muito empreendedor”. Já a variável categórica porque o empreendedor começou seu negócio próprio permitia três tipos de respostas fechadas, a saber: i) Achou que tinha uma oportunidade; ii) Teve Necessidade; iii) Ambas as alternativas.

Tabela 9 – Variáveis Predictoras do Perfil dos Empreendedores

vida_hj	Como você avalia sua vida hoje?	métrica
vida_5a	Como você acha que sua vida estará daqui a 5 anos?	métrica
empreen	Você se considera uma pessoa empreendedora?	métrica
ini_neg	De modo geral, você diria que começou seu negócio próprio porque:	categórica

Fonte: elaborado pelo autor.

No Apêndice 1 está apresentado o formulário completo com todas as variáveis utilizadas. O número total de variáveis predictoras no formulário de pesquisas foi 55; juntando-se com as nove variáveis de crescimento, chega-se ao total de 64; embora essa quantidade seja elevada, o tempo total de resposta do formulário de pesquisa era relativamente curto, ficando em torno de 10 minutos, como comentado anteriormente. Isso porque para muitas variáveis bastava indicar sim ou não, marcar um número ou atribuir uma nota de 0 a 10.

Por outro lado, essa quantidade de variáveis predictoras foi necessária e adequada para que análises estatísticas mais abrangentes fossem elaboradas, na tentativa de identificar os drivers que ajudam a explicar as empresas de alto crescimento, como detalhado no capítulo 4 a seguir.

4 RESULTADOS

Este capítulo apresenta os principais resultados obtidos a partir da análise de dados da amostra da presente tese e está dividido em cinco seções: a primeira apresenta a análise descritiva da amostra, através de tabelas e gráficos que ajudam a entender o perfil das empresas que participaram da pesquisa; a seção 2 trás os resultados do modelo Análise Fatorial aplicado à percepção dos empreendedores sobre as ferramentas de gestão das empresas, que embora secundário, mostrou alguns resultados dignos de comentários; a seção 3 é dedicada às análises do modelo Regressão Logística Binária, onde são detalhadas as variáveis relevantes que ajudam a explicar porque algumas empresas foram classificadas como de alto crescimento (EAC) e outras não; na quarta seção, resultados adicionais são apresentados e discutidos; embora tais resultados não sejam diretamente relacionados ao problema de pesquisa, podem contribuir para entender melhor a área de empreendedorismo no Brasil; o capítulo termina com a seção 5, onde são comentados os resultados obtidos em relação à questão de pesquisa e hipóteses formuladas na seção 2.

4.1 Análises Descritivas

As análises descritivas de uma pesquisa são sempre importantes para conhecer com mais detalhes o perfil de respondentes que fizeram parte do banco de dados, a partir do qual resultados e conclusões são inferidos. Assim, diversas variáveis que fizeram parte do formulário de pesquisa, tiveram como objetivo simplesmente descrever as empresas respondentes, de modo a permitir uma melhor interpretação das análises e resultados obtidos, a partir da aplicação de modelos estatísticos mais complexos.

Com base nas estimativas dos quatro CAGRs, que refletem o crescimento de faturamento e de número de colaboradores medidos para os dois períodos de três anos, ou seja, entre 2013 e 2016 e entre 2014 e 2017, foi possível identificar as empresas de alto crescimento (EACs) da amostra; conforme critério da OECD (2007), as firmas que apresentaram pelo menos um dos quatro CAGRS superior a 20% foram consideradas EACs; na amostra de 470 firmas, foram classificadas 91 EACs e 379 não EACs, com as respectivas taxas de participação de 19,4% e 80,6%.

A taxa de participação de EACs de 19,4% é formada pelas firmas com pelo menos um dos quatro CAGRs estimados superior a 20%; como esperado, algumas firmas apresentaram mais

de um CAGR atendendo ao critério de corte; independentemente das sobreposições, os dois CAGRs de faturamento identificaram 12,6% e 13,8% de EACs para os períodos 2013 a 2016 e 2014 a 2017 respectivamente, enquanto que os dois CAGRs de número de colaboradores resultaram em 8,5% e 7,7%.

A título de comparação, a publicação do IBGE b (2017), Estatísticas de Empreendedorismo 2015, estimou em 1,0% a taxa de participação de empresas de alto crescimento no Brasil, entre os anos de 2012 e 2015, com base apenas no critério da OECD (2007) relacionado ao número de colaboradores; as estimativas para períodos anteriores foram 1,4% entre o período 2010 a 2013 e 1,3% entre 2011 e 2014, mostrando trajetória de queda (IBGE b, 2017, p. 30). No entanto, devem ser consideradas algumas características das taxas estimadas pelo IBGE, que diferem da presente tese: i) referem-se a períodos isolados de quatro anos e não cumulativas; ii) baseiam-se apenas no critério número de colaboradores da OECD (2007), sem incluir as EACs com base em faturamento; iii) usa como denominador o total de empresas existentes no Cadastro Central de Empresas (CEMPRE) do IBGE, com pelo menos uma pessoa assalariada; a quantidade de empresas cadastradas é de grande magnitude, o que reduz a taxa estimada.

Outra comparação relevante pode ser feita com a pesquisa “As PMEs que Mais Crescem no Brasil”, elaborado pela empresa de consultoria Deloitte (2017), em colaboração com a Revista Exame; o ranking, já em sua 12ª edição, vem utilizando dados de *survey* com empresas de pequeno e médio porte; com base nas informações financeiras fornecidas pelas empresas foi estimado o CAGR de faturamento do período 2014 a 2016, portanto, de três anos; de um total de 291 firmas que responderam à pesquisa, 54 apresentaram CAGR superior a 20%, o que indica taxa de EACs igual a 18,6%. Fica a ressalva que não foi utilizado o critério de número de colaboradores e que o CAGR superior a 20% é mais fácil de ser atingido durante apenas três anos.

Considerando as mencionadas características da pesquisa do IBGE b (2017) e os resultados da Deloitte (2017), pode-se concluir que a taxa de participação de EACs de 19,4% obtida na presente tese ficou dentro do esperado. A taxa de EACs é relativamente baixa em função da grande dificuldade das empresas em atingirem o crescimento acumulado de 72,8% entre o ano *xx* e o ano base *xx-3*, seja em faturamento, seja em número de colaboradores. Os resultados das firmas brasileiras foram impactados pelos elevados custos indiretos na contratação de

recursos humanos e pela mencionada crise econômica durante os anos 2015 e 2016. Dessa forma, a taxa de participação de EACs de 19,4% pode ser considerada adequada.

Em complemento, foram estimados os CAGRs para os períodos compreendidos entre os anos de 2013 e 2017, referentes a cinco anos de faturamento e de número de colaboradores, refletindo quatro períodos de crescimento médio, com base nas equações (14) e (15) a seguir:

CAGR de faturamento da empresa para o período 2013 – 2017 (estimado)

$$\sqrt[4]{(1 + cfat_{14}) * (1 + cfat_{15}) * (1 + cfat_{16}) * (1 + cfat_{17})} - 1 \quad (14)$$

CAGR de número de colaboradores da empresa para o período 2013 – 2017 (estimado)

$$\sqrt[4]{\frac{col_{17}}{col_{13}}} - 1 \quad (15)$$

A Tabela 10 mostra as primeiras estatísticas descritivas das EACs e não EACs, com as médias dos CAGRs para os períodos de três e quatro anos; nota-se que as taxas de crescimento das EACs são maiores que as não EACs em todas as colunas, indicando consistência entre o crescimento de faturamento e de número de colaboradores.

Tabela 10 – CAGRs das EACs e Não EACs

	Número de Empresas	Faturamento			Número de Colaboradores		
		CAGR 2013/2016	CAGR 2014/2017	CAGR 2013/2017	CAGR 2013/2016	CAGR 2014/2017	CAGR 2013/2017
Não EACs	379	17,9%	14,8%	16,6%	7,2%	3,9%	6,2%
EACs	91	42,7%	33,4%	37,9%	27,6%	15,5%	22,0%
Total	470	22,7%	18,4%	20,8%	11,1%	6,2%	9,3%

Fonte: elaborado pelo autor.

Outra forma de confirmar a diferença entre as EACs e as não EACs é através da curva de distribuição de densidade da variável “CAGR 2013/2017 de faturamento”. A Figura 4 mostra

as 91 EACs à direita, com distribuição menos concentrada e cauda mais longa no histograma e na curva normalizada da variável, em comparação com as 379 não EACs, à esquerda.

Figura 4 – Distribuição de Densidade do CAGR 2013/2017 de Faturamento das Empresas

Fonte: elaborado pelo autor.

Em relação ao número de colaboradores, a Tabela 11 mostra outra importante característica de EACs, em comparação com as demais. As EACS apresentaram uma média de 28,5 colaboradores por empresa em 31/dezembro/2013, versus 31,9 colaboradores das não EACs; passados cinco anos, em 31/dezembro/2017, a média de EACs foi de 75,9 contra apenas 29,0 da não EACs.

Conforme enfatizado no capítulo 1, o Brasil vem passando por uma crise econômica nos últimos anos, que tem reduzido o número de empregos; esse cenário foi refletido no número médio de colaboradores das não EACs, que apresenta quedas em 2015 e 2016; tal fato valoriza ainda mais a importância das EACs na geração de empregos, dado que mesmo em período recessivo da economia brasileira, lograram mais que dobrar o número de colaboradores médio.

Tabela 11 – Número de Colaboradores Médio das EACs e Não EACs

	31/dez/2013	31/dez/2014	31/dez/2015	31/dez/2016	31/dez/2017 Previsto	Crescimento Acumulado 2013 - 2017
Não EACs	31,9	32,0	30,7	27,3	29,0	-9,1%
EACs	28,5	41,3	53,1	65,7	75,9	166,4%
Total	31,3	33,8	35,0	34,8	38,1	21,9%

Foram feitas outras análises descritivas da amostra com o intuito de descrever as 470 empresas em relação às principais variáveis coletadas. A Tabela 12 mostra as variáveis demográficas relativas às cidades onde as empresas estão sediadas, separando-as em três categorias: São Paulo Capital com 24,7%, São Paulo Interior com 21,7%, resultando um total de 46,4% no Estado de São Paulo. Parece haver maior concentração de EACs em São Paulo Capital em comparação com São Paulo Interior.

Essa tabela indica ainda a quantidade de empresas com sede atual ou anterior em locais rotulados como ecossistema; essa categoria criada para identificar locais favoráveis ao empreendedorismo, dado que o relacionamento entre empresas (*networking*) pode ser facilitado e fortalecido, agrupou parques tecnológicos, incubadoras ou espaços de coworking. Conforme mostra a tabela, 6,6% das firmas da amostra estão sediadas nesses ambientes, tanto para EACs como para não EACs.

Tabela 12 – Variáveis Predictoras Demográficas: Estatísticas Descritivas

	Cidade sede da empresa			Local: Ecosistema*
	São Paulo Capital	São Paulo Interior	São Paulo Estado	
Sim	116	102	218	31
Não	354	368	252	439
Total	470	470	470	470

* Local: Ecossistema inclui Parques Tecnológicos, Incubadoras ou Espaços de Coworking

Fonte: elaborado pelo autor.

A Tabela 13 descreve as variáveis predictoras de relacionamento das empresas, em relação aos aportes de capital das fontes consultadas na pesquisa; vale notar que a incidência de aportes de capital foi relativamente alta apenas para sócios, familiares ou amigos e empréstimos

bancários; os fundos de investimento tiveram pequena representatividade na amostra; deve-se destacar a maior incidência de aportes de investidores *angel*, aceleradoras, fundos de *private equity* e empréstimos bancários para as EACs.

Tabela 13 – Variáveis Predictoras de Aporte de Capital: Estatísticas Descritivas

	Sua empresa recebeu aporte de capital de:								
	Sócios	Família- res ou amigos	Crowd- funding	Angels	Acele- radoras	Capital Semen- te	Venture Capital	Private Equity	Emprést. Bancá- rios
Sim	392	115	1	13	4	2	3	2	245
Não	78	355	469	457	466	468	467	468	225
Total	470	470	470	470	470	470	470	470	470

Fonte: elaborado pelo autor.

Em relação às demais variáveis de relacionamento, a Tabela 14 detalha a quantidade de empresas que tiveram origem a partir de outras (*spin-offs*), aproximação com aceleradoras e recebimento de doações ou prêmios; para as quatro variáveis, as EACs apresentaram maiores taxas de incidência; merece ainda destaque, o número de empresas que receberam premiação em geral, que representa 35,1% do total.

Tabela 14 – Variáveis Predictoras de Relacionamento: Estatísticas Descritivas

	Spin-off	Relac. Acele- radoras	Recebeu:	
			Doa- ções	Premia- ção
Sim	48	19	10	165
Não	422	451	460	305
Total	470	470	470	470

Fonte: elaborado pelo autor.

Conforme comentado na seção 3.5 e detalhado na Tabela 7, foi coletada a percepção dos empreendedores sobre a importância de uma série de ferramentas de gestão, de utilização frequente nas empresas; no total foram 16 variáveis. É pertinente a dúvida: tais percepções seriam as mesmas para as empresas, independentemente se foram classificadas como de alto crescimento ou não? Para endereçar esse questionamento, uma das formas visuais mais didáticas de interpretar a percepção dos empreendedores é através do gráfico box plot.

Esse tipo de gráfico indica as medidas de dispersão dos respondentes de variáveis métricas, ficando na linha central a mediana, enquanto que o primeiro e o terceiro quartis compõem o retângulo de cada variável; o box plot ilustra a posição, dispersão, assimetria, caudas e dados discrepantes, conhecidos por *outliers* (Bussab & Morettin, 2010). A Figura 5 a seguir trás o mencionado gráfico box plot elaborado para as 470 empresas da amostra, separados em dois blocos; à esquerda figuram os resultados de percepção para os empreendedores das 379 empresas que não atingiram alto crescimento, enquanto que à direita estão as 91 empresas de alto crescimento; como todas as variáveis foram medidas em escala de 0 a 10, não houve identificação de *outliers*.

Figura 5 – Percepção dos Empreendedores sobre a Importância de Ferramentas de Gestão para EACs e Não EACs: Box Plot

Legenda na Tabela 7. Fonte: elaborado pelo autor.

A figura anterior dá a impressão que existe muito similaridade em todas as percepções dos empreendedores dos dois grupos para todas as variáveis; para confirmar essa aparente similaridade recomenda-se aplicar o teste estatístico de comparação de médias; esse teste

pode identificar as médias que são estatisticamente diferentes, a partir da análise da variância das duas distribuições em torno da média.

Por se tratar de apenas dois grupos, o teste t de Student para comparar as médias tem sido utilizado; quando se tem mais de dois grupos, com três ou mais médias, o teste ANOVA (Analysis of Variance), que se baseia na estatística F de Snedecor, é indicado; entretanto, as curvas de distribuição dos dois testes t de Student e F de Snedecor apresentam resultados idênticos para testes de dois grupos. Assim, foi utilizado o teste ANOVA para cada uma das 16 variáveis, de forma a identificar as variáveis nas quais a percepção dos empreendedores eram diferentes para os dois grupos, EACs e não EACs; os resultados dos testes estão detalhados na Tabela 15.

Tabela 15 – ANOVA: Comparação de Médias de Percepção dos Empreendedores de EACs e Não EACs

Nome	Variáveis	F	Prob > F
l_prod	Inovar ou diferenciar produtos e serviços	0,00	0,9593
l_desig	Inovar ou diferenciar design de produtos e serviços	0,02	0,8944
l_mod	Inovar ou diferenciar modelo de negócios	0,71	0,3992
l_p_neg	Elaborar e implementar Plano de Negócios	0,05	0,8181
l_p_mkt	Elaborar e implementar Plano de Marketing	0,09	0,7650
l_m_tra	Fazer propaganda em mídia tradicional (TV, rádio, revistas...)	0,01	0,9266
l_m_dig	Fazer propaganda em mídia digital (internet, site, smartphone, aplicativos...)	0,68	0,4084
l_soc_e	Ter sócios e diretores experientes em gestão de empresas	1,19	0,2764
l_parce	Montar uma rede de parceiros	0,53	0,4681
l_gd_cl	Conquistar grandes empresas como clientes	0,25	0,6160
l_ecoss	Sediar a empresa em local com ecossistema favorável	0,00	0,9986
l_marca	Registrar as marcas comerciais de sua empresa	2,98	0,0850
l_pat	Patentear os produtos e serviços de sua empresa	1,29	0,2558
l_inves	Obter capital de investidores	2,62	0,1059
l_e_ban	Obter empréstimos bancários	6,59	0,0105
l_d_fom	Obter doações de instituições de fomento ou de outras empresas	0,40	0,5294

Número de observações = 470 EAC 91 empresas Não EAC: 379 empresas

Fonte: elaborado pelo autor.

Conforme destacado na tabela anterior, somente três variáveis mostraram evidências de diferenças entre as médias; ao nível de significância de 5% apenas a variável “Obter empréstimos bancários” (l_e_ban) apresenta diferença entre as duas médias; se for considerado o nível de significância de 10%, a variável “Registrar as marcas comerciais de sua empresa” (l_marca) também mostra diferença estatisticamente relevante; a variável “Obter capital de investidores” (l_inves) está também muito próxima desse limite com 10,59%. Assim, pode-se concluir que para as 16 variáveis apenas três delas mostraram

indícios de percepção diferente entre os empreendedores de EACs e não EACs. A Figura 6 mostra essas três variáveis em box plot para uma melhor visualização dos achados estatísticos.

Figura 6 – Percepção dos Empreendedores sobre a Importância de Marcas, Capital de Investidores e Empréstimos Bancários para EACs e Não EACs: Box Plot

Fonte: elaborado pelo autor.

Percebe-se pela Figura 6 que os empreendedores de EACs valorizam mais o registro de marcas comerciais, a obtenção de capital de investidores e a obtenção de empréstimos bancários; ou seja, aparentemente as EACs estão mais preocupadas com proteger suas marcas e ter acesso a capital para alavancar seu desenvolvimento; esse achado pode ser confirmado pela Tabela 16. É importante antecipar que essa diferença de percepção dos empreendedores das EACs versus as não EACs, em relação a essas três ferramentas de gestão teve impacto nas análises do modelo Regressão Logística, detalhadas na seção 4.3.

Tabela 16 – Percepção dos Empreendedores sobre a Importância de Marcas, Capital de Investidores e Empréstimos Bancários para EACs e Não EACs: Médias e Desvios

Nome	Variáveis	não EACs		EACs	
		Média	D. Padrão	Média	D. Padrão
l_marca	Registrar as marcas comerciais de sua empresa	6,9103	3,3495	7,5714	3,4468
l_inves	Obter capital de investidores	5,6913	3,4468	6,3407	3,3804
l_e_ban	Obter empréstimos bancários	4,6939	3,4251	5,7033	3,1109

Fonte: elaborado pelo autor.

4.2 Resultados da Análise Fatorial

Nesta seção são apresentados os principais resultados do modelo Análise Fatorial, aplicado às variáveis de percepção dos empreendedores sobre a importância de ferramentas de gestão, detalhadas na Tabela 7. A rigor, essa seção trás resultados menos importantes para a questão de pesquisa da presente tese que a seção 4.3 seguinte, que contém os resultados de Regressão Logística Binária. É na próxima seção que são feitas as análises dos dados das 470 empresas da amostra, separados em dois grupos, de acordo com as taxas de crescimento em faturamento ou número de colaboradores, que identifica as empresas de alto crescimento (EACs) e as demais (não EACs). Entretanto, algumas análises de Regressão Logística Binária foram feitas a partir dos fatores, e não das variáveis originais do banco de dados; assim, foi necessária a inversão da sequencia de apresentação dos resultados dos dois modelos, para que os fatores fossem analisados nesta seção, para uso posterior.

Antes de rodar o modelo Análise Fatorial é sempre recomendável identificar se há indícios de multicolinearidade entre as variáveis, estimando os coeficientes de correlação de Pearson entre elas, bem como os respectivos testes de significância estatística ou p-value. A Tabela 17 trás a Matriz de Correlações de Pearson para as 16 variáveis bem como os testes p-value, indicados por *, quando inferiores a 5%. Os coeficientes superiores a 0,40 foram destacados na tabela para facilitar a visualização; note-se que alguns são relativamente elevados e estatisticamente significantes, o que comprova bons indícios para se realizar uma Análise Fatorial.

A título de exemplo, destaca-se a alta correlação das três últimas variáveis da matriz, relacionadas a aspectos financeiros: “Obter capital de investidores” (l_inves), “Obter empréstimos bancários” (l_e_ban) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom). Outro exemplo é a correlação de 0,6029 entre a importância atribuída a

“Elaborar e implementar Plano de Negócios” (l_p_neg) e “Elaborar e implementar Plano de Marketing” (l_p_mkt). Isso indica que provavelmente tais variáveis poderão se juntar em fatores comuns na Análise Fatorial, agrupando a percepção dos empreendedores sobre a importância de determinadas ferramentas de gestão.

Tabela 17 – Percepção dos Empreendedores: Matriz de Correlação de Pearson

	l_prod							
l_desig	0,3734*	l_desig						
l_mod	0,4379*	0,2607*	l_mod					
l_p_neg	0,2585*	0,2002*	0,3448*	l_p_neg				
l_p_mkt	0,2104*	0,2060*	0,2750*	0,6029*	l_p_mkt			
l_m_tra	0,1130*	0,2208*	0,2144*	0,3346*	0,3458*	l_m_tra		
l_m_dig	0,2579*	0,3076*	0,2766*	0,3001*	0,5225*	0,3367*	l_m_dig	
l_soc_e	0,1072*	0,1203*	0,1590*	0,2438*	0,2678*	0,1610*	0,2276*	
l_parce	0,1961*	0,1978*	0,1644*	0,2172*	0,2602*	0,2465*	0,2670*	
l_gd_cl	0,1451*	0,1562*	0,1441*	0,1258*	0,1605*	0,1231*	0,1324*	
l_ecoss	0,1930*	0,2894*	0,2078*	0,2714*	0,2861*	0,2369*	0,2351*	
l_marca	0,2907*	0,3565*	0,2502*	0,3088*	0,3656*	0,3014*	0,3598*	
l_pat	0,1802*	0,3429*	0,2310*	0,2906*	0,2872*	0,2707*	0,2488*	
l_inves	0,1347*	0,1846*	0,2033*	0,2607*	0,3096*	0,2422*	0,2697*	
l_e_ban	0,0496	0,1241*	0,0994*	0,1682*	0,1721*	0,2966*	0,2264*	
l_d_fom	0,1557*	0,2076*	0,2155*	0,2375*	0,2800*	0,2258*	0,2221*	

	l_soc_e							
l_parce	0,3279*	l_parce						
l_gd_cl	0,1966*	0,2194*	l_gd_cl					
l_ecoss	0,2090*	0,1714*	0,4160*	l_ecoss				
l_marca	0,2205*	0,2650*	0,2051*	0,3272*	l_marca			
l_pat	0,2257*	0,1545*	0,3070*	0,4293*	0,5915*	l_pat		
l_inves	0,2969*	0,2010*	0,2774*	0,3652*	0,3646*	0,3568*	l_inves	
l_e_ban	0,1394*	0,2427*	0,1462*	0,1258*	0,2410*	0,2230*	0,4988*	l_e_ban
l_d_fom	0,2026*	0,2150*	0,3220*	0,4123*	0,2741*	0,3440*	0,6167*	0,4603*

* indica p-value < 0,05

Fonte: elaborado pelo autor.

O próximo passo foi rodar o modelo Análise Fatorial descrito na seção 3.4 da presente tese utilizando o mencionado software estatístico STATA®. A primeira tabela de resultados do modelo mostra um número de fatores idêntico ao número de variáveis, dado que é feita uma transformação matricial, partindo de 16 variáveis e criando igual quantidade de fatores.

Os resultados da Análise Fatorial estão na Tabela 18, onde na coluna Fator aparecem os 16 fatores gerados e na coluna *Eigenvalue* (Autovalor) o quanto cada fator explica das 16 variáveis; a soma dos 16 Eigenvalues é exatamente 16; a coluna Diferença mostra quanto os Eigenvalues estão variando a cada fator; da divisão dos respectivos Eigenvalues de cada fator por 16, resulta a coluna Proporção de variáveis explicadas, que se acumula na última coluna Acumulada, chegando a 100% com os 16 fatores.

Tabela 18 – Percepção dos Empreendedores: Análise Fatorial – Eigenvalues

Fator	Eigenvalue	Diferença	Proporção	Acumulada
Fator 1	4,947690	3,4155	0,3092	0,3092
Fator 2	1,532240	0,2741	0,0958	0,4050
Fator 3	1,258100	0,2236	0,0786	0,4836
Fator 4	1,034460	0,0334	0,0647	0,5483
Fator 5	1,001020	0,0586	0,0626	0,6108
Fator 6	0,942380	0,1475	0,0589	0,6697
Fator 7	0,794840	0,0750	0,0497	0,7194
Fator 8	0,719800	0,0677	0,0450	0,7644
Fator 9	0,652130	0,0257	0,0408	0,8052
Fator 10	0,626450	0,0905	0,0392	0,8443
Fator 11	0,535980	0,0281	0,0335	0,8778
Fator 12	0,507830	0,0783	0,0317	0,9096
Fator 13	0,429590	0,0459	0,0268	0,9364
Fator 14	0,383740	0,0559	0,0240	0,9604
Fator 15	0,327850	0,0220	0,0205	0,9809
Fator 16	0,305900	.	0,0191	1,0000

LR teste: independente vs. saturado: $\chi^2(120) = 2145.91$ Prob> $\chi^2 = 0.0000$

Método dos Componentes Principais (pcf); 470 observações; 70 parâmetros

Fonte: elaborado pelo autor.

O teste LR que aparece na parte inferior da Tabela 18 é o teste de esfericidade de Bartlett, comentado na seção 3.4, que mostra que a matriz de correlações obtida na análise é diferente da matriz identidade de mesma dimensão; portanto, as correlações obtidas na Análise Fatorial são diferentes de zero e a extração de fatores foi adequada (Fávero & Belfiore, 2015). Dessa forma, pode-se evoluir na análise e decidir o número de fatores a reter.

Com base nos resultados da Tabela 18, os Fatores 1 a 5 devem ser retidos para representar as 16 variáveis, uma vez que são os fatores mais importantes, medidos pelos respectivos Eigenvalues; a decisão por reter cinco fatores, todos com Eigenvalue maior que 1, segue o critério de Kaiser, que recomenda manter no modelo os fatores que expliquem uma ou mais

variável. Juntos os cinco fatores resumem as 16 variáveis e são capazes de explicar 61,08% da variabilidade total dos dados (Fávero & Belfiore, 2015).

A segunda tabela de resultados do modelo Análise Fatorial é uma matriz de correlações entre os cinco fatores retidos e as 16 variáveis originais, que no caso é chamada Matriz de Cargas Fatoriais (*Factor Loading*). Na Tabela 19 estão os resultados obtidos para as 16 variáveis de percepção dos empreendedores. Além das cargas fatoriais, a tabela também reporta os valores de *Uniqueness* (Singularidade) e *Communality* (Comunalidade) para cada variável.

As cargas fatoriais da Tabela 19 indicam o coeficiente de correlação de Pearson entre cada um dos cinco fatores retidos no modelo e as 16 variáveis originais; assim, se elevarmos ao quadrado todas as cargas fatoriais e somarmos as células de uma mesma coluna ou linha separadamente, constata-se que: a soma das cargas fatoriais ao quadrado de cada coluna resulta o respectivo *Eigenvalue* de cada um dos cinco fatores; ou seja, a porcentagem de explicação da variabilidade dos dados das 16 variáveis que é explicada pelo fator; a soma das cargas fatoriais ao quadrado de cada linha resulta o total de explicação da respectiva variável pelos cinco fatores retidos, ou seja, sua *Communality*; a diferença entre essa soma e o total de 100% é o valor do *Uniqueness*, que representa a porcentagem da variabilidade dos dados da variável não explicado pelos fatores retidos. O software STATA® não calcula diretamente a *Communality* de cada variável, mas reporta o *Uniqueness*, que no modelo Análise Fatorial por Componentes Principais, detalhado na seção 3.4, pode ser facilmente calculado pela equação (16) a seguir:

$$Communality = 1 - Uniqueness \quad (16)$$

Quanto maior a *Communality* de cada uma das 16 variáveis, melhor ela estará representada pelos cinco fatores retidos; ao contrário, quanto maior o *Uniqueness*, pior representada ela estará. Assim, um bom modelo de Análise Fatorial deveria apresentar baixas comunalidades para todas as variáveis; analisando a tabela anterior, percebe-se que as variáveis destacadas na tabela que estão ligeiramente abaixo de 50%, foram “Fazer propaganda em mídia tradicional” (l_m_tra) e “Fazer propaganda em mídia digital” (l_m_dig), com respectivamente 44,80% e 49,84%. O significado desse achado é que tais variáveis não estão bem representadas pelos cinco fatores retidos, em relação às demais, embora em níveis aceitáveis.

Tabela 19 – Percepção dos Empreendedores: Análise Fatorial – Matriz de Cargas Fatoriais

Variável	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Uniqueness	Communality
l_prod	0,4415	0,4759	0,3276	-0,1439	0,4303	0,2655	0,7345
l_desig	0,5058	0,2722	0,3628	-0,2105	0,0774	0,4882	0,5118
l_mod	0,4934	0,3981	0,1355	-0,1383	0,3157	0,4609	0,5391
l_p_neg	0,5941	0,3052	-0,2913	0,0688	-0,1999	0,4244	0,5756
l_p_mkt	0,6459	0,2675	-0,3788	0,1074	-0,2538	0,2918	0,7082
l_m_tra	0,5278	0,0648	-0,3250	-0,1582	-0,1859	0,5520	0,4480
l_m_dig	0,5937	0,2688	-0,2622	-0,0607	-0,0359	0,5016	0,4984
l_soc_e	0,4471	-0,0595	-0,1939	0,5779	0,1767	0,3937	0,6063
l_parce	0,4695	0,0235	-0,2192	0,3690	0,4170	0,4209	0,5791
l_gd_cl	0,4448	-0,2915	0,3867	0,4173	0,0382	0,3920	0,6080
l_ecoss	0,5948	-0,1674	0,3705	0,2061	-0,2320	0,3846	0,6154
l_marca	0,6666	0,0792	0,1561	-0,1145	-0,2519	0,4485	0,5515
l_pat	0,6396	-0,0932	0,3205	-0,0413	-0,3885	0,3268	0,6732
l_inves	0,6468	-0,4789	-0,0684	-0,1597	0,1057	0,3110	0,6890
l_e_ban	0,4721	-0,4825	-0,3035	-0,3883	0,2455	0,2412	0,7588
l_d_fom	0,6211	-0,4876	0,0413	-0,1659	0,1540	0,3235	0,6765

Fonte: elaborado pelo autor.

Além do teste de esfericidade de Bartlett, recomenda-se para comprovar se uma Análise Fatorial é adequada, ou se alguma das 16 variáveis deveria ser retirada do modelo rodar a estatística Kaiser-Meyer-Olkin, conhecida como KMO, conforme detalhado na seção 3.4. A Tabela 20 trás os valores de KMO das variáveis de percepção quando reduzidas a cinco fatores.

Tabela 20 – Percepção dos Empreendedores: Análise Fatorial – KMO

Variável	KMO
l_prod	0,7827
l_desig	0,8811
l_mod	0,8553
l_p_neg	0,8094
l_p_mkt	0,7950
l_m_tra	0,9069
l_m_dig	0,8480
l_soc_e	0,8640
l_parce	0,8454
l_gd_cl	0,8560
l_ecoss	0,8709
l_marca	0,8492
l_pat	0,8323
l_inves	0,8395
l_e_ban	0,7889
l_d_fom	0,8411
Global	0,8389

Fonte: elaborado pelo autor.

A estatística KMO indica que a Análise Fatorial é adequada, no global e para cada uma das variáveis, dado que a nota de corte para inaceitável é abaixo de 0,5. Com o KMO global de 0,8389 a adequação pode ser considerada boa, conforme indicado a Tabela 4; a variável que se revelou menos adequada, ou seja, que resultou o menor KMO foi “Inovar produtos e serviços” (l_prod) com 0,7827 indicando adequação média, mas muito próximo da faixa boa que se inicia em 80%; portanto, a Análise Fatorial foi adequada pelos dois critérios – Bartlett e KMO – e nenhuma variável deve ser retirada do modelo (Fávero & Belfiore, 2015).

Embora a matriz de cargas fatoriais dê uma boa indicação das variáveis que estão correlacionadas com cada um dos fatores, sempre é recomendável proceder à rotação dos fatores; essa operação justifica-se, pois algumas variáveis apresentam cargas fatoriais em mais de um fator, dificultando o processo de interpretação dos fatores (Malhotra, 2012). Assim, optou-se por fazer a rotação ortogonal Varimax®, onde os eixos são mantidos em ângulo reto, para minimizar o número de variáveis com cargas fatoriais elevadas em mais de um fator; busca-se com a rotação que cada variável apresente carga fatorial alta com apenas um fator, e baixa com todos os demais. Após passar pela operação Varimax®, os valores dos Eigenvalues de cada um dos cinco fatores são alterados, mas a representatividade acumulada permanece inalterada, conforme Tabela 21.

Tabela 21 – Percepção dos Empreendedores: Análise Fatorial – Eigenvalues Rotacionados

Fator	Eigenvalue	Diferença	Proporção	Acumulada
Fator 1	2,36935	0,2335	0,1481	0,1481
Fator 2	2,13584	0,0048	0,1335	0,2816
Fator 3	2,13101	0,3241	0,1332	0,4148
Fator 4	1,80688	0,4764	0,1129	0,5277
Fator 5	1,33044	.	0,0832	0,6108

LR teste: independente vs. saturado: $\chi^2(120) = 2145.91$ Prob> $\chi^2 = 0.0000$

Método dos Componentes Principais (pcf); 470 observações; 70 parâmetros; Rotação Varimax®

Fonte: elaborado pelo autor.

Quanto à Matriz de Cargas Fatoriais Rotacionadas, constante da Tabela 22, percebe-se que cargas fatoriais foram alteradas, em comparação com a Tabela 19, antes da rotação, mantendo inalterados os valores de *Uniqueness* e *Communality*; a rotação facilitou a interpretação dos fatores, conforme as células destacadas na tabela.

Tabela 22 – Percepção dos Empreendedores: Análise Fatorial – Matriz de Cargas Fatoriais Rotacionadas

Variável	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Uniqueness	Communality
l_prod	0,0736	0,0160	0,0820	0,8441	0,0984	0,2655	0,7345
l_desig	0,1578	0,0959	0,3368	0,5966	-0,0918	0,4882	0,5118
l_mod	0,2286	0,0999	0,0578	0,6796	0,1081	0,4609	0,5391
l_p_neg	0,7078	0,0612	0,1441	0,1611	0,1556	0,4244	0,5756
l_p_mkt	0,7927	0,0973	0,1568	0,0849	0,1964	0,2918	0,7082
l_m_tra	0,5882	0,3015	0,0893	0,0546	-0,0135	0,5520	0,4480
l_m_dig	0,6072	0,1855	0,0650	0,2752	0,1238	0,5016	0,4984
l_soc_e	0,2385	0,0867	0,1627	-0,0053	0,7179	0,3937	0,6063
l_parce	0,2011	0,2158	-0,0232	0,2266	0,6634	0,4209	0,5791
l_gd_cl	-0,1286	0,1312	0,6215	0,0651	0,4286	0,3920	0,6080
l_ecoss	0,1392	0,1409	0,7341	0,1018	0,1643	0,3846	0,6154
l_marca	0,4232	0,2048	0,5071	0,2626	-0,0660	0,4485	0,5515
l_pat	0,2988	0,1914	0,7227	0,1266	-0,0946	0,3268	0,6732
l_inves	0,1482	0,7388	0,3104	0,0482	0,1501	0,3110	0,6890
l_e_ban	0,1301	0,8594	-0,0460	0,0220	0,0274	0,2412	0,7588
l_d_fom	0,0467	0,7237	0,3465	0,1050	0,1398	0,3235	0,6765

Método dos Componentes Principais (pcf); 470 observações; 70 parâmetros; Rotação Varimax®

Fonte: elaborado pelo autor.

Os fatores devem ser interpretados através das variáveis que possuem cargas fatoriais elevadas, ou seja, altas correlações com cada um deles; assim, o Fator 1 representa e pode substituir as 4 variáveis de importância atribuída pelos empreendedores à: “Elaborar e implementar Plano de Negócios” (l_p_neg), “Elaborar e implementar Plano de Marketing” (l_p_mkt), “Fazer propaganda em mídia tradicional” (l_m_tra) e “Fazer propaganda em mídia digital” (l_m_dig), tendo sido denominado “Fator 1 Marketing”. Embora não obrigatório, é recomendável atribuir um nome a cada fator, de forma a identificá-lo com mais facilidade. O nome deve remeter para o conjunto que variáveis por ele representadas, razão pela qual o nome “Marketing” foi escolhido.

O Fator 2 com as três variáveis “Obter capital de investidores” (l_inves), “Obter empréstimos bancários” (l_e_ban) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom) foi denominado “Fator 2 Finanças”.

O Fator 3, que representa as variáveis “Conquistar grandes empresas como clientes” (l_gd_cl), “Sediar a empresa em local com ecossistema favorável” (l_ecoss), “Registrar as marcas comerciais de sua empresa” (l_marca) e “Patentear os produtos e serviços de sua empresa” (l_pat), foi designado com o nome “Fator 3 Ecossistema”.

O Fator 4, com as variáveis “Inovar ou diferenciar produtos e serviços” (l_prod), “Inovar ou diferenciar design de produtos e serviços” (l_desig) e “Inovar ou diferenciar modelo de negócios” (l_mod), foi chamado de “Fator 4 Inovação”.

O último fator, Fator 5, que substitui as 2 variáveis “Ter sócios e diretores experientes em gestão de empresas” (l_soc_e) e “Montar uma rede de parceiros” (l_parce) foi denominado “Fator 5 Networking”.

Com a interpretação dos principais fatores e conhecendo as variáveis que cada um representa e substitui, o conjunto de 16 variáveis de percepção dos empreendedores sobre a importância de ferramentas de gestão ficou reduzido a cinco grandes dimensões: Marketing, Finanças, Ecossistema, Inovação e Networking. Esse achado tem certa relevância uma vez que indica a forma como os empreendedores esquematizam mentalmente um amplo conjunto de percepções.

O último passo da Análise Fatorial é a geração das novas variáveis métricas com os cinco fatores retidos; assim, estimou-se o valor de cada um dos cinco fatores para cada uma das 470 empresas da amostra. Isso foi feito por meio da criação dos escores fatoriais, que são os coeficientes de um modelo linear, que calcula o valor de cada fator, para as 16 variáveis de percepção de cada empresa, conforme detalhado na seção 3.4. O banco de dados passou a ter após a Análise Fatorial as cinco novas variáveis, denominadas “Fator 1 Marketing”, “Fator 2 Finanças”, “Fator 3 Ecossistema”, “Fator 4 Inovação” e “Fator 5 Networking”, que foram usadas em análises posteriores.

4.3 Resultados da Regressão Logística Binária

Nesta seção são apresentados os resultados obtidos através de Regressão Logística Binária, usando sempre como variável resposta categórica a classificação das 470 empresas da amostra, em empresas de alto crescimento (EACs) e as demais (não EACs); quanto às variáveis preditoras, foram utilizados diversos conjuntos de variáveis, combinando a ampla base de dados disponíveis das empresas. Como frequentemente ocorre em pesquisas, alguns conjuntos de variáveis preditoras revelaram significância estatística e são detalhados nesta seção; entretanto, as variáveis que não se mostraram relevantes para explicar as EACs

também são comentadas, com o intuito de contribuir para futuras pesquisas; o alinhamento dos achados com a literatura, quando identificado, também é discutido nesta seção.

Modelagem 1: Regressão Logística EAC e Percepção dos Empreendedores

A primeira modelagem testada foi Regressão Logística Binária tendo como variável resposta EACs e variáveis preditoras as 16 variáveis de percepção dos empreendedores sobre a importância das ferramentas de gestão de suas empresas, detalhadas na Tabela 7. Utilizando o comando *stepwise* do software estatístico STATA®, detalhado na seção 3.3, o modelo final manteve apenas quatro das 16 variáveis originais, ao nível de significância de 10%, conforme mostra a Tabela 23.

Tabela 23 – EACs e Percepção dos Empreendedores: Regressão Logística Binária – Coeficientes

Variável	Coeficiente	Erro Padrão	z de Wald	P > z	Intervalo de Confiança 95%	
<i>l_m_dig</i>	-0,0892	0,05041	-1,77	0,077	-0,1881	0,0096
<i>l_marca</i>	0,0861	0,04261	2,02	0,043	0,0026	0,1696
<i>l_d_fom</i>	-0,0732	0,03489	-2,10	0,036	-0,1416	-0,0048
<i>l_e_ban</i>	0,1281	0,04114	3,11	0,002	0,0475	0,2087
constante	-1,6826	0,41992	-4,01	0,000	-2,5056	-0,8596

Número de obs = 470; LR qui2(4) = 16,46; Prob > chi2 = 0,0025; Pseudo R2 = 0.0356; Log likelihood = -222,74139

Fonte: elaborado pelo autor.

Pode-se afirmar que o modelo final é adequado, pois o teste do χ^2 (qui-quadrado) indica uma probabilidade inferior a 0,25% que nenhum coeficiente seja diferente de zero; ou seja, podemos inferir que pelo menos um dos coeficientes é diferente de zero. O Pseudo R² de McFaden de 3,56% foi bastante baixo, porém, como discutido na seção 3.3, trata-se de uma métrica secundária de aferição da qualidade de uma Regressão Logística, ficando a qualidade do modelo assegurada pelo teste do χ^2 (qui-quadrado) e Matriz de Classificação que foi detalhada na Tabela 24.

Tabela 24 – EACs e Percepção dos Empreendedores: Regressão Logística Binária – Matriz de Classificação

Classificado como	Incidência na Amostra		Total
	EAC	não EAC	
EAC	60	174	234
não EAC	31	205	236
Total	91	379	470
Eficiência global do modelo			56,38%
Sensitividade			65,93%
Especificidade			54,09%

Classificado como EAC se previsão de Prob (EAC) $\geq 0,19$

Fonte: elaborado pelo autor.

A Matriz de Classificação foi construída a partir do modelo final da Tabela 23, utilizando o comando *predict* em STATA®, que estimou a probabilidade de uma empresa ser de alto crescimento (EAC), com base nas 4 variáveis preditoras que passaram no teste do p-value, dado pela coluna *z de Wald*. Essa matriz serve para validar a qualidade do modelo, para um determinado *cutoff*, que representa o limiar para o modelo classificar uma empresa como EAC; assim, escolhido o *cutoff* de 19%, se uma empresa tivesse probabilidade maior ou igual a 19% de ser uma EAC, o modelo classificava como EAC; se a probabilidade fosse inferior a 19%, o modelo classificava como não EAC; a escolha do *cutoff* igual a 19% foi decorrência da incidência aproximada de EACs da amostra, conforme do cálculo da equação (17) a seguir:

$$cutoff = \frac{\text{número de EACs}}{\text{número de empresas}} = \frac{91}{470} = 19,36\% \quad (17)$$

A Matriz de Classificação da Tabela 24 trás 3 métricas da qualidade do modelo, conforme detalhado na seção 3.3: o índice global de acertos do modelo, a sensibilidade e a especificidade; o primeiro indicador, de eficiência global é dado pela soma dos acertos de EACs e não EACs, ou seja, pela soma dos valores nas diagonais $60 + 205 = 265$, dividido pelo total de empresas da amostra que é 470, resultando em 56,38%; por ser pouco superior a uma classificação aleatória para duas possibilidades, que é 50%, pode-se afirmar que a qualidade do modelo é razoável. No entanto, a sensibilidade de 65,93%, calculada pela divisão do número de acertos de EACs igual a 60 em um total de 91 empresas, revela uma qualidade aceitável. A especificidade que resultou em 54,09% foi calculada dividindo-se o

número 205 de acertos de não EACs pelo total de 379 não EACs da amostra e pode ser considerada razoável.

Conforme detalhado na seção 3.3, a terceira métrica de avaliar a adequação do modelo é através da curva *ROC*; esse gráfico apresenta a variação de sensibilidade no eixo das ordenadas e $(1 - \text{especificidade})$ no eixo das abscissas, mostrando de forma gráfica o *trade off* entre sensibilidade e especificidade, quando da alteração de *cutoff*. Portanto, quanto maior e mais próxima de 1 for a área abaixo da curva *ROC*, maior a eficiência global da previsão (Fávero, 2015). A Figura 7 mostra a curva *ROC* para o modelo EACs em função da percepção dos empreendedores, com uma área embaixo da curva estimada em 0,6358.

Figura 7 – Curva *ROC* de EACs em função da Percepção dos Empreendedores

Fonte: elaborado pelo autor.

Assim, embora o primeiro modelo, com base nas 16 variáveis de percepção dos empreendedores sobre a importância das ferramentas de gestão, tenha apresentado uma Matriz de Classificação apenas razoável, os resultados do teste χ^2 (qui-quadrado) e da curva *ROC* indicam uma boa adequação, o que avaliza sua interpretação.

Analisando-se em detalhes a Tabela 23, como conclusões do primeiro modelo, as variáveis “Fazer propaganda em mídia digital” (l_m_dig), “Registrar as marcas comerciais” (l-marca), “Obter empréstimos bancários” (l_e_ban) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom) foram relevantes para explicar porque uma empresa foi classificada como EAC em comparação com as demais, ou seja, tais variáveis afetam a probabilidade de uma empresa conseguir atingir a condição de EAC.

O sinal matemático dos coeficientes indica a direção em que está crescendo a probabilidade de EACs; a cada unidade a mais da variável preditora, se o sinal for positivo, aumenta a probabilidade de EACs; se o sinal for negativo, a probabilidade de EACs é reduzida. Assim, se a importância atribuída à marca e a empréstimos bancários for alta, a probabilidade de EACs é maior; ao contrário, elevada importância à mídia digital e obter doações indicam menor probabilidade de EACs.

Existe uma consistência parcial com os achados nos testes de comparação de médias ANOVA, detalhados na seção 4.1 desse capítulo, onde as variáveis de importância atribuída à marca e a empréstimos bancários já mostravam diferenças estatisticamente relevantes. As variáveis negativas, sobre a importância em fazer propaganda em mídia digital e doações, não estavam previstas e revelam um resultado relevante.

Uma forma simples de interpretar os achados estatísticos é através de gráficos que mostram a curva de probabilidade de EACs em função de alguma variável métrica *Ceteris Paribus*, ou seja, mantidas as demais variáveis constantes. Seguem as Figuras 8, 9 e 10 com três exemplos de curvas de probabilidade de EACs em função da percepção dos empreendedores sobre as variáveis “Fazer propaganda em mídia digital” (l_m_dig), “Registrar as marcas comerciais” (l-marca) e “Obter empréstimos bancários” (l_e_ban).

Figura 8 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Fazer Propaganda em Mídia Digital

Fonte: elaborado pelo autor.

Figura 9 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Registrar Marcas Comerciais

Fonte: elaborado pelo autor.

Figura 10 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Obter Empréstimos Bancários

Fonte: elaborado pelo autor.

Como se pode notar, na Figura 8 quanto maior a percepção da importância atribuída pelos empreendedores a fazer propaganda em mídia digital, menor a probabilidade de EACs; ao contrário, nas Figuras 9 e 10, um aumento na percepção da importância em registrar marcas comerciais e obter empréstimos bancários leva a uma maior probabilidade de EACs.

Uma variante do primeiro modelo consistiu em utilizar o conjunto de cinco fatores gerados pelas 16 variáveis de percepção dos empreendedores para estimar a probabilidade de EACs; buscou-se com esse modelo identificar se havia similaridade entre os achados com as variáveis originais em comparação com os fatores delas originados. Entretanto, a qualidade do modelo gerado não passou no teste de χ^2 (qui-quadrado) e nenhum dos fatores ficou no modelo com coeficiente diferente de zero. O Fator 2 Finanças mostrou o melhor teste de p-value dentre os cinco fatores com 11,77%.

Esses achados podem ser explicados pela complexidade do problema e pelo fato dos fatores terem sido gerados para as 470 empresas da amostra, de certo modo perdendo variâncias de percepções entre empreendedores. Com efeito, gerado os mesmos cinco fatores com base nas 16 variáveis, mas apenas para as 91 empresas classificadas como EACs, a Regressão

Logística Binária passou no teste de χ^2 (qui-quadrado) com p-value igual a 4,14% e o Fator 2 Finanças mostrou-se significativo ao nível de 4,40%, com coeficiente positivo de 0,2423.

Dado que o Fator 2 Finanças representa as três variáveis “Obter capital de investidores” (l_inves), “Obter empréstimos bancários” (l_e_ban) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom) o modelo a partir dos fatores confirma que uma percepção elevada dos empreendedores sobre essas variáveis ajuda a explicar as EACs. Isso comprova certa consistência com o primeiro modelo preditivo em relação à obtenção de empréstimos bancários.

Modelagem 2: Regressão Logística EAC e Variáveis Categóricas

A segunda modelagem, com resultados válidos do ponto de vista estatístico e da literatura, utilizou como preditoras um conjunto de variáveis dummies, tipo sim ou não; a maior parte dessas variáveis categóricas está descrita na Tabela 8 e foram coletadas diretamente na pesquisa; outras cinco variáveis foram originadas a partir dos dados primários e referem-se a:

- i) cidade onde fica a sede da empresa: três novas variáveis dummies, classificando as cidades em São Paulo Capital, São Paulo Interior e São Paulo Estado, ficando as demais cidades como base; essas variáveis foram detalhadas na Tabela 12;
- ii) sede onde fica ou ficava a empresa: uma nova variável dummy, identificando as sedes atual ou anterior em Parques Tecnológicos, Incubadores ou Espaços de Coworking como ecossistema favorável, ficando as demais sedes como base; da amostra; conforme Tabela 12, foram 31 empresas classificadas como “ecossistema”, representando 6,6% da amostra;
- iii) aporte de capital de terceiros: uma nova variável dummy, consolidando aporte de capital de seis fontes de investimentos: crowdfunding, investidores *angel*, aceleradoras, capital semente, *venture capital* e *private equity*; ou seja, todas as fontes financeiras da pesquisa exceto aporte de capital dos próprios sócios, de familiares ou de amigos e empréstimos bancários; foram 22 empresas classificadas com aporte de capital de terceiros ou 4,7% da amostra, detalhadas na Tabela 13.

Essas novas variáveis estão detalhadas na Tabela 25 a seguir; dessa maneira, esse modelo começou com 18 variáveis categóricas, sendo 13 da Tabela 7 mais as cinco novas da Tabela 25, para testar quais eram relevantes para as EACs, abrangendo os seguintes aspectos: cidades onde ficam as empresas, locais onde estão ou estavam sediadas, se a empresa era uma *spin-off*, se a empresa manteve relacionamento com aceleradoras, se a empresa recebeu aportes de capital, empréstimos bancários, doações ou premiações.

Tabela 25 – Novas Variáveis Preditoras Originadas dos Dados Primários

cid_spc	Cidade onde fica a sede da empresa: São Paulo Capital?	categórica
cid_spi	Cidade onde fica a sede da empresa: São Paulo Interior?	categórica
cid_spt	Cidade onde fica a sede da empresa: São Paulo Estado?	categórica
s_ecoss	Local onde fica a sede atual (ou anterior): Ecossistema? *	categórica
cap_ter	Sua empresa recebeu aporte de capital de: Terceiros? **	categórica

* Local: Ecossistema inclui Parques Tecnológicos, Incubadoras ou Espaços de Coworking

** Aporte de Capital de Terceiros inclui: Crowdfunding, Angels, Aceleradoras, Fundos Semente, Venture Capital ou Private Equity

Fonte: elaborado pelo autor.

A partir desse conjunto de 18 variáveis preditoras categóricas, foram iniciadas as rodadas de Regressão Logística Binária e análise dos coeficientes obtidos com o uso do comando *stepwise* do STATA®, ao nível de significância de 10%. Logo foram verificados os usuais problemas de modelagem, que tiveram que ser sanados um a um, como a seguir explicado.

O primeiro problema foi a multicolinearidade entre as três variáveis relativas à cidade onde fica a sede da empresa; assim, foram testadas duas combinações, uma com São Paulo Capital e São Paulo Interior e a outra com São Paulo Capital e São Paulo Estado; a primeira combinação, São Paulo Capital e São Paulo Interior, revelou-se mais indicado para permanecer na modelagem, que passou de 18 para 17 variáveis.

O seguinte problema de estimação ocorreu devido à baixa taxa de incidência de aportes de capital nas empresas, por parte dos fundos de investimento, tais como capital semente ou *private equity*; a Tabela 13 detalhou que apenas entre um e quatro empresas receberam aportes de capital de cinco tipos de fundos; a única taxa de incidência ligeiramente maior foi de investidores *angel*, com 13 em 470 empresas, representando 2,8%. Quando uma variável apresenta um número de observações muito pequeno na amostra, a Regressão Logística não pode estimar os coeficientes e, conseqüentemente, é forçada a eliminar os respectivos respondentes. Dado que a modelagem com as 17 variáveis preditoras não poderia estimar

adequadamente os coeficientes das seis variáveis relacionadas a aportes de capital, inclusive de investires *angel*, tais variáveis originais foram eliminadas; em substituição às mesmas foi mantida a nova variável “Sua empresa recebeu aporte de capital de: Terceiros?”, que representava as seis fontes financeiras excluídas; essa nova variável detalhada na Tabela 25 ocorria com 22 empresas e taxa de incidência de 4,7%; a modelagem ficou então reduzida a 11 variáveis categóricas preditoras.

Os modelos inicialmente testados com as 11 variáveis remanescentes revelaram que seria necessário eliminar duas outras variáveis categóricas: aporte de capital de sócios e aporte de capital de familiares ou amigos; isso ocorreu em função de três motivos: primeiro, pela relativa subjetividade das variáveis, dado que a maior parte das empresas precisaria ter recebido algum aporte de capital de sócios ou de familiares ou amigos para ser criada; segundo, em decorrência das elevadas taxas de incidência desses dois tipos de aporte de capital na amostra, o que pode ser comprovado na Tabela 13, com 83,4% e 24,5% respectivamente; terceiro, com a introdução dessas duas variáveis na modelagem, outras eram eliminadas, impedindo uma análise mais adequada do conjunto de variáveis preditoras categóricas; identificar que o aporte de capital de sócios ou familiares ou amigos contribuía para a probabilidade de EACs, à custa de perder outros achados mais relevantes, agregaria muito pouco a presente pesquisa, e foram eliminadas; dessa forma, a modelagem 1 foi executada com 9 variáveis preditoras categóricas, tendo sido obtidos os resultados detalhados na Tabela 26.

Tabela 26 – EACs e Variáveis Categóricas: Regressão Logística Binária – Coeficientes

Variável	Coeficiente	Erro Padrão	z de Wald	P > z	Intervalo de Confiança 95%	
s_premi	0,7886	0,2443	3,23	0,001	0,3099	1,2674
cid_spc	0,4478	0,2695	1,66	0,097	-0,0803	0,9760
e_ban	0,5074	0,2505	2,03	0,043	0,0163	0,9984
s_spin	0,6998	0,3600	1,94	0,052	-0,0058	1,4055
s_r_ace	1,5746	0,4907	3,21	0,001	0,6128	2,5364
_cons	-2,3287	0,2531	- 9,20	0,000	-2,8248	-1,8326

Número de obs = 470; LR qui2(5) = 31,15; Prob > chi2 = 0,0000; Pseudo R2 = 0.0674; Log likelihood = -215,39364

Fonte: elaborado pelo autor.

Esse modelo é aceitável, uma vez que passa pelo teste do χ^2 (qui-quadrado), com uma probabilidade inferior a 0,01%, ou seja, pelo menos um dos coeficientes é diferente de zero. O Pseudo R² de McFaden obtido de 6,74% continua relativamente baixo, mas ligeiramente superior à modelagem 1 de 3,56%; como já discutido na seção 3.3 e na modelagem 1, o Pseudo R² é uma métrica secundária de qualidade do modelo; a Matriz de Classificação da

Tabela 27 foi também ligeiramente melhor que na modelagem 1, com eficiência global de 62,98% versus 56,38%.

Tabela 27 – EACs e Variáveis Categóricas: Regressão Logística Binária – Matriz de Classificação

Classificado como	Incidência na Amostra		Total
	EAC	não EAC	
EAC	54	137	191
não EAC	37	242	279
Total	91	379	470
Eficiência global do modelo			62,98%
Sensitividade			59,34%
Especificidade			63,85%

Classificado como EAC se previsão de Prob (EAC) $\geq 0,19$

Fonte: elaborado pelo autor.

A curva *ROC*, discutida na seção 3.3 e na apresentação de resultados da modelagem 1, resultou em uma área abaixo da curva de 0,6735, mais próxima de 1 que o modelo anterior que apresentou 0,6358; portanto, todos os indicadores de aferição de qualidade, sinalizam que a modelagem 2 é relativamente melhor que a primeira, o que dá um aval para sua interpretação e discussão.

Figura 11 – Curva ROC de EACs em função de Variáveis Categóricas

Fonte: elaborado pelo autor.

Voltando à Tabela 26 para a discussão dos resultados da modelagem 2, pode-se inferir que cinco variáveis foram relevantes para as EACs, todas com coeficiente positivo; portanto, seguindo a ordem da tabela, as empresas que receberam premiação, estão localizadas em São Paulo Capital, receberam empréstimos bancários, são oriundas de outras empresas (*spin-offs*) ou tiveram relacionamento com aceleradoras são as mais propensas a atingirem a condição de EACs.

Deve-se salientar a magnitude dos coeficientes, com destaque para as empresas que tiveram relacionamento com aceleradoras de 1,5746 e p-value igual a 0,1%; como todas as variáveis são categóricas, podendo assumir os valores 0 ou 1, o maior coeficiente do modelo indica que sua ocorrência (evento) produz o maior aumento na probabilidade de EACs, *Ceteris Paribus*.

A segunda variável em importância na probabilidade de EACs é o recebimento de premiação, também com p-value de 0,1%; o reconhecimento público de uma empresa é sempre um bom indicador de competitividade, o que o resultado parece confirmar; na sequência aparece *spin-offs*, com p-value inferior a 5%; o achado faz sentido, dado que empresas *spin-offs* aproveitam a experiência de outras empresas das quais são oriundas e está em linha com Andersson e Klepper (2013); a variável relacionada à obtenção de empréstimos bancários também é positiva, com p-value inferior a 5%; esse resultado está em linha com a modelagem 1, onde a percepção de importância em obter empréstimos bancários também mostrou significância estatística; ou seja, ambos, percepção sobre a importância em obter empréstimos bancários e o efetivo aporte contribuíram para aumentar a probabilidades de EACs.

A localização da sede da empresa em São Paulo Capital também contribui positivamente para a probabilidade de EACs, porém deve-se ressaltar que apresentou o menor coeficiente e o maior p-value da modelagem, com 9,7%, quase sendo eliminada; a taxa de incidência de empresas sediadas na cidade de São Paulo era de 24,7% na amostra; trata-se de um achado complementar que dá uma boa informação sobre localização de empresas.

Registra-se como resultados relevantes que as variáveis categóricas relacionadas a aporte de capital de terceiros, recebimento de doações, localização em São Paulo Interior e sede em ecossistemas (parques tecnológicos, incubadoras e espaços de coworking) foram excluídas do modelo final; nem a variável que agregou aporte de capital de variados fundos, incluindo investidores *angel*, foi preservada; o fato dessas variáveis terem sido eliminadas indica que

elas não ajudavam a explicar a probabilidade de EACs na pesquisa, ao nível de significância de 10%; tais achados e suas limitações são discutidos na seção 4.5.

Como a modelagem 2 só contém variáveis preditoras categóricas fica sem sentido plotar gráficos com a curva de probabilidade; esse modelo trouxe achados que podem ser úteis para a modelagem 3 a seguir.

Modelagem 3: Regressão Logística EAC, Percepção dos Empreendedores e Variáveis Categóricas

Na modelagem 3 foram incluídas as 16 variáveis métricas de percepção dos empreendedores sobre a importância de ferramentas de gestão de empresas, utilizadas na modelagem 1, e simultaneamente as 9 variáveis categóricas da modelagem 2; assim, iniciou-se a pesquisa com um conjunto de 25 variáveis. Utilizando o comando *stepwise* do STATA®, com um nível de significância de 10%, foram eliminadas 20 variáveis e retidas as cinco que constam da Tabela 28; esse número de variáveis eliminadas era esperado, em função da quantidade relativamente grande de variáveis iniciais da modelagem e do tamanho da amostra.

Tabela 28 – EACs, Percepção dos Empreendedores e Variáveis Categóricas: Regressão Logística Binária – Coeficientes

Variável	Coeficiente	Erro Padrão	z de Wald	P > z	Intervalo de Confiança 95%	
l_d_fom	-0,08391	0,0358724	-2,34	0,019	-0,1542	-0,0136
l_e_ban	0,14829	0,0420641	3,53	0,000	0,0658	0,2307
s_premi	0,86318	0,2469247	3,50	0,000	0,3792	1,3471
s_spin	0,64903	0,3649565	1,78	0,075	-0,0663	1,3643
s_r_ace	1,74722	0,4977822	3,51	0,000	0,7716	2,7229
_cons	-2,34847	0,2873087	-8,17	0,000	-2,9116	-1,7854

Número de obs = 470; LR qui2(5) = 38,59; Prob > chi2 = 0,0000; Pseudo R2 = 0.0835; Log likelihood = -211,67602

Fonte: elaborado pelo autor.

A modelagem 3 é adequada, dado que passa pelo teste do χ^2 (qui-quadrado) com probabilidade menor que 0,01%, apresentando vários coeficientes diferentes de zero. O Pseudo R² de McFaden subiu para 8,35%, ainda relativamente baixo, mas melhor que os anteriores; como já discutido na seção 3.3 e nos dois modelos anteriores, o Pseudo R² é uma métrica secundária de qualidade do modelo; a Tabela 29 trás a Matriz de Classificação, também a melhor dos três modelos, com eficiência global de 66,17%.

Tabela 29 – EACs, Percepção dos Empreendedores e Variáveis Categóricas: Regressão Logística Binária – Matriz de Classificação

Classificado como	Incidência na Amostra		Total
	EAC	não EAC	
EAC	55	123	178
não EAC	36	256	292
Total	91	379	470
Eficiência global do modelo			66,17%
Sensitividade			60,44%
Especificidade			67,55%

Classificado como EAC se previsão de Prob (EAC) $\geq 0,19$

Fonte: elaborado pelo autor.

Foi feita a curva *ROC*, já discutida na seção 3.3 e para os modelos anteriores, que mostrou área abaixo da curva de 0,7038, a mais próxima de 1 dos três modelos; dessa maneira, todos os indicadores de aferição de qualidade, sinalizam que a modelagem 3 é melhor que as anteriores, razão pela qual sua interpretação e discussão serão mais detalhadas.

Figura 12 – Curva *ROC* de EACs em função da Percepção dos Empreendedores e de Variáveis Categóricas

Fonte: elaborado pelo autor.

A Tabela 28 mostra as cinco variáveis que ajudam a explicar a probabilidade de EACs; dentre elas, duas são variáveis de percepção dos empreendedores e as três restantes categóricas; a variável de percepção da importância em obter empréstimos bancários apresentou coeficiente positivo, enquanto que a percepção da importância em obter doações de instituições de fomento mostrou impacto negativo; ambas apresentam p-values inferiores a 5%; em relação às três variáveis categóricas, todos os coeficientes resultaram positivos e, pela ordem da tabela, referem-se a empresas que receberam premiações, são originadas a partir de outras empresas (*spin-offs*) e possuem relacionamento com aceleradoras; a variável *spin-offs* apresentou p-value elevado com 7,5%, e para as outras duas foram inferiores a 0,1%.

Esse modelo confirma os achados da modelagem 2 para a variável de relacionamento com aceleradoras, com a maior magnitude da tabela, porém é necessário fazer uma ressalva; não se deve comparar diretamente os coeficientes de variáveis métricas, medidas em escala de 0 a 10 com coeficientes de variáveis categóricas que assumem apenas os valores 0 ou 1.

As variáveis retidas na modelagem 3 já figuravam nos modelos anteriores, o que mostra uma desejável consistência dos resultados das três análises. Assim, pode-se inferir que as firmas que receberam premiação, relacionam-se com aceleradoras e foram *spin-offs* possuem maior probabilidade de serem EACs. O mesmo vale para as empresas que tem empreendedores que valorizam empréstimos bancários, mas atribuem pouca importância a doações de instituições de fomento ou de outras empresas; o significado desses achados está discutido na seção 4.5.

Dado que esse modelo contém variáveis categóricas e métricas, foram feitos três gráficos com as curvas de probabilidade, em função das variáveis preditoras mais relevantes do ponto de vista estatístico do modelo final, de modo a ilustrar os achados e ajudar a interpretá-los. A Figura 13 mostra a probabilidade de EACs em função da variável de percepção dos empreendedores sobre a importância em “Obter empréstimos bancários”, para empresas com e sem relacionamento com aceleradoras.

Figura 13 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Obter Empréstimos Bancários e Relacionamento com Aceleradoras

Fonte: elaborado pelo autor.

A Figura 13 mostra as duas curvas de probabilidade plotadas no mesmo gráfico, facilitando sua interpretação; quando uma empresa mantém ou manteve relacionamento com aceleradoras, a probabilidade de EACs é bem superior, independentemente do nível de percepção da importância em obter empréstimos bancários; por outro lado, quando essa percepção de importância aumenta, também aumenta a probabilidade de EACs.

O próximo exemplo, mostrado na Figura 14, trás a probabilidade de EACs em função da variável de percepção dos empreendedores sobre a importância em “Obter empréstimos bancários” e da variável que indicava se a empresa recebeu ou não alguma premiação.

Figura 14 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Obter Empréstimos Bancários e Recebimento de Premiação

Fonte: elaborado pelo autor.

Pode-se notar na Figura 14, de forma similar à anterior, que as empresas que receberam premiação tiveram maior probabilidade de EACs, para qualquer nível de percepção da importância em obter empréstimos bancários; essa variável de percepção continua tendo um impacto positivo similar ao do gráfico anterior, ou seja, quanto maior a percepção, maior a probabilidade de EACs.

O terceiro e último exemplo aparece na Figura 15, onde a probabilidade de EACs é declinante, dado que a variável preditora de percepção dos empreendedores sobre a importância em “Obter doações de instituições de fomento e outras empresas”, teve um coeficiente negativo no modelo; as curvas mostram de forma simultânea o impacto da variável dummy sobre recebimento de premiação por parte da empresa.

Figura 15 – Curva de Probabilidade de EACs em função da Percepção dos Empreendedores da Importância em Receber Doações de Instituições de Fomento e Recebimento de Premiação

Fonte: elaborado pelo autor.

Fica claro, mais uma vez, que o recebimento de premiação melhora a curva de probabilidade de EACs, para quaisquer valores da variável métrica; a diferença com os dois gráficos anteriores e que estas curvas se apresentam de forma declinante, em decorrência do coeficiente negativo da variável métrica na modelagem 3.

Como procedido na modelagem 1, foi testado o mesmo modelo utilizando os cinco fatores gerados na Análise Fatorial, em substituição as 16 variáveis originais, para estimar a probabilidade de EACs, com o conjunto de variáveis métricas e categóricas; a intenção era certificar se havia equivalência de resultados com as variáveis originais. A qualidade do modelo com os cinco fatores não passou no teste de χ^2 (qui-quadrado) e nenhum dos fatores permaneceu no modelo com coeficiente diferente de zero, continuando o Fator 2 Finanças, com o teste de p-value igual a 11,77%, o mais relevante.

A explicação para esse resultado continua sendo a complexidade do problema de pesquisa e pelo fato dos fatores terem sido gerados para as 470 empresas da amostra, que limitaram as variâncias de percepções entre empreendedores. Com um modelo alternativo, considerando os

cinco fatores gerados a partir apenas das 91 EACs, a Regressão Logística Binária passa no teste de χ^2 (qui-quadrado) com p-value inferior a 5%; nessa modelagem, o Fator 2 Finanças teve coeficiente positivo de 0,2423 e p-value igual a 4,14%.

Como mencionado, o Fator 2 Finanças substituiu as três variáveis “Obter capital de investidores” (l_inves), “Obter empréstimos bancários” (l_e_ban) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom). Dessa forma, o modelo a partir dos fatores confirmou mais uma vez que a percepção elevada dos empreendedores da importância em obter aportes financeiros aumenta a probabilidade de EACs. Tais resultados da modelagem 3 utilizando fatores são muito similares aos encontrados na modelagem 1, o que demonstra adequada consistência das análises.

Com essas três modelagens de Regressão Logística Binária utilizando a variável resposta EACs, essa seção se encerra; foram identificadas diversas variáveis preditoras que ajudaram a explicar porque algumas empresas atingem alto crescimento, em resposta à questão de pesquisa; a discussão das variáveis relevantes para as EACs dos modelos consta da seção 4.5.

4.4 Gazelas

Nesta seção foram detalhadas as análises sobre as gazelas identificadas nesta pesquisa. Segundo a definição da OECD (2007) adotada, as gazelas são as empresas de alto crescimento (EACs) com menos de cinco anos de idade; na base de dados de 470 empresas desta pesquisa foram identificadas apenas seis gazelas, representando 1,3% da amostra e 6,6% das EACs.

Era esperado que o número fosse pequeno, conforme explica Monteiro (2017), em função da curva de distribuição da taxa de crescimento das firmas em uma economia; no entanto, acabou ficando ainda menor, em decorrência dos filtros de exclusão recomendados pela OECD (2007), detalhados na seção 3.5; para serem EACs, as firmas precisavam apresentar crescimento de pelo menos um dos quatro CAGRs calculados, dois de faturamento e dois do número de colaboradores, superior a 20%, além do número mínimo de 10 colaboradores no ano $xx-3$; após os dois filtros, 91 firmas foram classificadas como EACs e 379 como não EACs.

Outro filtro adicional contribuiu para limitar a classificação de EACs como gazelas nesta pesquisa; sempre seguindo o critério da OECD (2007), as 123 empresas que não apresentaram

faturamento no ano completo de 2013 foram excluídas, ficando na amostra as empresas mais velhas, com três ou mais anos de idade; tal fato gerou um conflito de métricas, pois eram poucas as novas EACs, com menos de cinco anos, passíveis de serem gazelas, com uma base de dados centrada no período 2013 a 2017; assim, a janela para uma EAC ser classificada como gazela nesta pesquisa foi bastante restritiva.

Para avaliar a incidência de gazelas na presente pesquisa, foi feita uma comparação com os dados de novas empresas dos Estados Unidos, publicado por Morelix e Russell-Fritch (2017, p. 14), no relatório *The Kauffman Index Growth Entrepreneurship*, da *Ewing Marion Kauffman Foundation*. Embora utilize critérios diferentes para identificar as chamadas empresas *scaleups*, conceitualmente referem-se a empresas de crescimento acelerado em relação ao número de colaboradores. O relatório estimou que em 2016 a participação de *scaleups* foi 1,1%, com cerca 1.100 em um total de 100.000 novas empresas. Olhando o mesmo indicador em perspectiva no citado relatório, matéria da revista *The Economist* (2017) chegou a 30 mil *scaleups* em um total de 4,4 milhões de novas empresas criadas nos Estados Unidos nos últimos dez anos, com 0,68% de participação. Esses resultados estão relativamente alinhados com a participação de 1,3% de gazelas encontradas na presente tese.

Em relação ao cenário brasileiro, a participação obtida de 6,6% de gazelas dentre as EACs também se mostrou razoavelmente compatível com os dados do IBGE b (2017), publicados em Estatísticas de Empreendedorismo 2015; utilizando de forma parcial o critério da OECD (2007), dado que se baseia apenas no CAGR do número de colaboradores, aquela instituição estimou a participação de gazelas em relação às EACs em 13,6% em 2013, 13,5% em 2014 e 13,8% em 2015 (IBGE, 2017, p. 37).

As principais características demográficas das seis gazelas identificadas nesta pesquisa seguem padrão semelhante das EACs, em comparação com as não EACs, como detalhado na Tabela 30.

Do primeiro bloco da Tabela 30 deve-se destacar que para todas as variáveis de relacionamento – ecossistema, *spin-offs*, relacionamento com aceleradoras e recebimento de doações e prêmios, as taxas de incidências das gazelas foram maiores; em relação às variáveis do segundo bloco da tabela, os investidores *angel* e as aceleradoras parecem ter apostado de forma mais intensa nas gazelas, enquanto que a taxa de empréstimos bancários está em linha

com as não EACs; duas das seis gazelas são sediadas em São Paulo Capital, o que mostra taxa equivalente às EACs, conforme Tabela 12.

Tabela 30 – Gazelas: Estatísticas Descritivas

	Número de Empresas	Local: Ecosistema*	Spin-off	Relac. Aceleradoras	Recebeu:	
					Doações	Premiação
não EACs	379	6,6%	9,2%	2,4%	1,6%	31,1%
EACs	91	6,6%	14,3%	11,0%	4,4%	51,6%
Gazelas	6	16,7%	33,3%	33,3%	16,7%	66,7%

	Número de Empresas	Sua empresa recebeu aporte de capital de:				
		Sócios	Famíliares ou amigos	Angels	Aceleradoras	Emprést. Bancários
não EACs	379	81,8%	22,4%	1,6%	0,5%	49,9%
EACs	91	90,1%	33,0%	7,7%	2,2%	61,5%
Gazelas	6	100,0%	16,7%	16,7%	33,3%	50,0%

* Local: Ecosistema inclui Parques Tecnológicos, Incubadoras ou Espaços de Coworking
Fonte: elaborado pelo autor.

O número de colaboradores médio das seis gazelas apresentou forte crescimento no período da pesquisa, saindo de 23,0 em 31/dezembro/2013, para 67,0 em 2016 e 73,6 previstos para o final de 2017; o crescimento do número de colaboradores acumulado no período de cinco anos foi de 220,0%, superior aos 166,4% registrado pelas EACs na Tabela 11.

Foi rodado o modelo Regressão Logística Binária, com a variável resposta categórica gazela sim ou não; dada a taxa de incidência extremamente baixa de gazelas na amostra, a sensibilidade da modelagem não funcionou, gerando indicadores de adequação não satisfatórios.

Dada a confidencialidade inerente à ética de pesquisas, reafirmada no formulário de coleta de dados do Apêndice 1, as informações individuais de cada empresa não podem ser reveladas. Mesmo sem identificar as variáveis relevantes do ponto de vista estatístico, que ajudam a explicar o desempenho das gazelas, a caracterização descritiva dessas empresas indica que seguiram padrão similar ao das EACs, de forma ainda mais pronunciada.

4.5 Resultados Adicionais

Nesta seção foram detalhadas as análises e resultados considerados complementares à seção 4.3, obtidos a partir do modelo Regressão Linear, como alternativa à Regressão Logística e de variáveis não incluídas nas modelagens anteriores; tais variáveis adicionais analisadas nesta seção referem-se a: número de sócios, formação acadêmica dos sócios e indicadores de perfil dos empreendedores.

Modelagens com Regressão Linear

A Regressão Linear é, de certo modo, análoga à Regressão Logística, pois ambos os modelos são preditivos e estimam coeficientes de variáveis explicativas, que podem ser tanto métricas quanto categóricas transformadas em dummies. Entretanto, a grande diferença é a natureza da variável resposta dos dois modelos: na Regressão Linear é métrica e na Regressão Logística é categórica. Tal diferença faz com que a Regressão Linear estime os valores da variável resposta y , enquanto que a Regressão Logística a probabilidade de ocorrência de y .

Alguns pressupostos dos dois modelos são similares, como a ausência de multicolinearidade entre as variáveis preditoras. Outros pressupostos aplicam-se somente à Regressão Linear, como a necessidade de estimar e analisar os resíduos, para avaliar se apresentam distribuição normal, se estão correlacionados com qualquer variável preditora, ou seja, se existe heterocedasticidade em vez da recomendável homocedasticidade, e se são aleatórios e independentes (Fávero, 2015, p. 45). Além disso, na Regressão Linear a presença de *outliers* pode aportar distorções relevantes aos resultados, enquanto que em Regressão Logística isso não ocorre.

Para rodar o modelo Regressão Linear na presente pesquisa foi necessário substituir a variável categórica, que indicava se uma empresa era ou não de alto crescimento, por outra variável resposta do tipo métrica; foi escolhida a variável CAGR 2013/2017 de faturamento, descrita na Tabela 10 e na Figura 4, que representa o crescimento médio anual do faturamento das 470 empresas, durante os cinco anos coletados na amostra. Foi analisada como variável resposta alternativa o CAGR 2013/2017 de número de colaboradores; no entanto, o uso do CAGR 2013/2017 de faturamento revelou-se mais indicado por retratar com maior sensibilidade a métrica de crescimento das empresas.

Os mesmos modelos utilizados em Regressão Logística Binária e detalhados na seção 4.3 foram repetidos para Regressão Linear Múltipla, utilizando as 16 variáveis métricas de percepção dos empreendedores, os cinco fatores gerados na Análise Fatorial e as nove variáveis categóricas das modelagens anteriores.

Os principais resultados foram comentados nos próximos parágrafos de forma simplificada, dado que não se tratava do principal problema de pesquisa; os achados, no entanto, foram relevantes não só para descobrir novas variáveis que afetam o crescimento de faturamento das empresas, mas também para verificar se havia consistência nos resultados dos dois modelos preditivos.

Modelagem 1: Regressão Linear CAGR 2013/2017 de Faturamento e Percepção dos Empreendedores

A primeira Regressão Linear com a variável resposta métrica CAGR 2013/2017 de faturamento e as 16 variáveis preditoras métricas de percepção dos empreendedores sobre a importância das ferramentas de gestão foi rodada em STATA®. Antes de iniciar a modelagem, foi analisada a presença de eventuais *outliers* na amostra; em relação às variáveis preditoras, dado que todas usam a escala de 0 a 10, não há ocorrência de *outliers*; em relação à variável resposta CAGR 2013/2017 de faturamento, foi aplicado o algoritmo *BACON* em STATA®, sigla para *Blocked Adaptive Computationally Efficient Outlier Nominators*, proposto por Billor, Hadi e Velleman (2000). O comando identificou a presença de 14 *outliers* ao nível de significância de 5%, conforme distribuição χ^2 (qui-quadrado), utilizada pelo algoritmo; os *outliers*, empresas com elevados CAGRs 2013/2017 de faturamento, foram dessa forma excluídos da amostra para a modelagem.

A Regressão Linear feita com uso do comando *stepwise* e nível de significância de 10% indicou adequação com Prob > F = 0,0317, ou seja, pelo menos um dos coeficientes foi diferente de zero; o R^2 igual a 1,51% foi baixo, mas deve ser visto com reserva, dadas as limitações da estatística, análogas às discutidas para o Pseudo R^2 de McFaden na seção 3.3.

Das 16 variáveis originais, as duas que fizeram parte do modelo final referiam-se à percepção dos empreendedores sobre a importância em “Inovar ou diferenciar modelo de negócios” (l_mod) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom); os p-

values das variáveis foram 0,041 e 0,039 e os coeficientes 0,0124 e -0,0052 respectivamente; portanto, coeficiente positivo para inovação em modelo de negócios e negativo para obtenção de doações.

O aparecimento da nova variável significativa relativa ao modelo de negócios, em comparação com a modelagem 1 de Regressão Logística, da seção 4.3, era esperada e até mesmo desejável, constituindo-se em achado relevante. Embora os dois modelos preditivos apresentem certa similaridade, buscam coeficientes com critérios totalmente diferentes; a Regressão Linear estima o valor da variável resposta métrica, em função das variáveis explicativas, enquanto que a Regressão Logística busca identificar as variáveis explicativas que impactam na probabilidade de ocorrência da variável resposta categórica; os valores da variável resposta CAGR 2013/2017 de faturamento foram utilizados na Regressão Linear para testar a significância estatística de cada variável preditora e estimar os seus respectivos coeficientes; assim, apresenta limitada relação quando comparada a utilizar um conjunto de empresas classificadas como EACs, testar a significância estatística de cada variável preditora na probabilidade de ocorrência do evento e estimar os respectivos coeficientes para o cálculo de probabilidades.

Pode-se concluir que a percepção positiva dos empreendedores sobre inovar ou diferenciar o modelo de negócios ajuda a explicar as empresas de maior crescimento médio em faturamento. Ao contrário, quanto maior a importância atribuída à obtenção de doações de instituições e outras empresas, menor o resultado de crescimento médio em faturamento, resultado esse rigorosamente alinhado com a modelagem 1 de Regressão Logística, onde tal percepção apareceu com coeficiente negativo.

Foram verificados os principais pressupostos de Regressão Linear para a primeira modelagem, através da aplicação dos testes de: i) Shapiro-Francia, referente à normalidade dos resíduos e mais indicado para amostras maiores que 30 observações; ii) inexistência de multicolinearidade, medida pelas estatísticas *tolerance* e *vif* (*variance inflation factor*); iii) Breusch-Pagan / Cook-Weisberg, que avalia se não existe heterocedasticidade, assegurando a homocedasticidade (Fávero, 2015).

O modelo depurado resultante da regressão com a variável resposta CAGR 2013/2017 de faturamento e as variáveis preditoras de percepção dos empreendedores passou nos testes de inexistência de multicolinearidade e heterocedasticidade; no entanto, no teste de Shapiro-

Francia não pode ser comprovada a normalidade dos resíduos. Segundo Fávero (2015), a normalidade dos resíduos é necessária tão somente para validar os testes de hipóteses dos modelos de regressão, ou seja, os p-values dos testes t e do teste F; aquele autor citou Wooldridge (2012), para quem a violação desse pressuposto pode ser minimizada quando grandes amostras são utilizadas. Esse é o caso da presente tese com 470 empresas, quantidade bem superior a 30, razão pela qual o teste de Shapiro-Francia não foi comentado nas regressões lineares seguintes.

Repetindo a mesma modelagem para os cinco fatores, em substituição das 16 variáveis de percepção originais, os resultados foram surpreendentes; o modelo foi adequado com o teste de Prob > F igual a 0,0381 e R^2 de 0,94%; ficou no modelo Regressão Linear, ao nível de significância de 5%, o Fator 4 Inovação com coeficiente positivo de 0,0198 e p-value igual a 0,038; o modelo passou nos testes de inexistência de multicolinearidade e heterocedasticidade.

O Fator 4 Inovação representa as variáveis “Inovar ou diferenciar produtos e serviços” (l_prod), “Inovar ou diferenciar design de produtos e serviços” (l_desig) e “Inovar ou diferenciar modelo de negócios” (l_mod), ou seja, as três variáveis relacionadas à percepção dos empreendedores sobre inovação; portanto, as empresas com maior crescimento médio em faturamento podem ser explicadas por atribuírem maior importância à inovação.

Esse achado é relevante e consistente com a modelagem anterior, onde a variável “Inovar ou diferenciar modelo de negócios” teve influência positiva; é também mais representativa que a modelagem 1 da seção 4.3, que utilizou os fatores como variáveis preditoras, dado que nenhum fator foi significativo em Regressão Logística.

A Regressão Linear a partir dos cinco fatores gerados apenas a partir das 91 EACs da amostra resultou em teste de significância F igual a 0,0384 e R^2 de 0,94%; os testes de inexistência de multicolinearidade e heterocedasticidade deram resultados satisfatórios. Os fatores do modelo ao nível de significância de 10% foram: Fator 1 Marketing com coeficiente positivo e p-value de 0,038 e o Fator 2 Finanças com coeficiente negativo e p-value igual a 0,080; o Fator 4 Inovação foi excluído.

Conforme detalhado na seção 4.2, o Fator 1 Marketing representa as quatro variáveis de percepção dos empreendedores sobre a importância de: “Elaborar e implementar Plano de Negócios” (l_p_neg), “Elaborar e implementar Plano de Marketing” (l_p_mkt), “Fazer

propaganda em mídia tradicional” (l_m_tra) e “Fazer propaganda em mídia digital” (l_m_dig); analogamente, o Fator 2 Finanças representa as variáveis: “Obter capital de investidores” (l_inves), “Obter empréstimos bancários” (l_e_ban) e “Obter doações de instituições de fomento ou de outras empresas” (l_d_fom).

Comparando as duas modelagens a partir dos cinco fatores gerados com as percepções dos empreendedores das 91 EACs, aparentemente verificou-se certa inconsistência em relação ao Fator 2 Finanças. Na Regressão Logística o Fator 2 Finanças estava associado a maior probabilidade de EACs e na Regressão Linear a um menor crescimento em faturamento, dado seu coeficiente negativo. Esse problema pode ocorrer em modelos de regressão, quando o comando *stepwise* é utilizado; na presença de uma variável, o coeficiente de uma segunda variável pode mudar de sinal (Fávero, 2015). A técnica para esclarecer esse conflito é rodar o modelo apenas com uma variável por vez, para identificar o verdadeiro sinal positivo ou negativo que associa a variável resposta com a variável preditora.

Foram rodadas as modelagens utilizando apenas um fator gerado a partir das 91 EACs de cada vez, em Regressão Logística e Regressão Linear; por ter sido o Fator 2 Finanças o único a mostrar significância no modelo logístico, a influência positiva já estava confirmada, como documentado na seção 4.3 modelagem 1; para a Regressão Linear constatou-se que o Fator 2 Finanças não apresentou significância estatística, na ausência do F1 Marketing, o que pode explicar parcialmente a incongruência verificada; o Fator 1 Marketing, como variável preditora única, manteve a influência significativa e positiva, o que confirma a robustez do achado.

Em complemento, a interpretação dos fatores em uma Análise Fatorial é relativamente complexa, dado que um fator representa e substitui algumas variáveis que não necessariamente indicam a mesma direção de variação; com efeito, quando analisadas as três variáveis do Fator 2 Finanças, percebe-se que uma delas, a percepção da importância dos empreendedores em relação a receber doações de instituições de fomento ou de outras empresas apareceu na modelagem 1 de Regressão Logística, seção 4.3, com sinal negativo, enquanto que a percepção em relação à obtenção de empréstimos bancários teve sinal positivo; a terceira variável que compõe o Fator 2 Finanças, de percepção em relação a obter capital de investidores, embora tenha saído da modelagem 1, também teve sinal positivo.

Portanto, em um mesmo fator, ocorreu um efeito ambíguo: i) duas variáveis foram favoráveis, tipo quanto maior, melhor, tanto para a probabilidade de EACs quanto para o crescimento médio em faturamento; ii) uma variável foi desfavorável, tipo quanto menor, melhor para ambas as regressões. Os efeitos podem ter se contrabalançado; na Regressão Logística a estimativa de probabilidade de EACs aumentou com o Fator 2 Finanças, dando maior relevância às duas variáveis positivas, empréstimos bancários e aporte de capital de investidores terceiros; na Regressão Linear, ao contrário, em função dos valores da variável crescimento médio em faturamento da amostra, o Fator 2 Finanças acabou pendendo para a variável doações, resultando em impacto negativo. Assim, houve adequada consistência geral entre as duas modelagens, com Regressão Logística e Linear, dado que ambas indicaram que a variável preditora de doações de instituições de fomento tem impacto negativo. Os resultados estatisticamente significantes e coeficientes positivos de Regressão Linear para as variáveis relativas à inovação em modelo de negócios e ao Fator 4 Inovação complementam os achados relevantes da presente pesquisa.

Modelagem 2: Regressão Linear CAGR 2013/2017 de Faturamento e Variáveis Categóricas

Seguindo os mesmos passos da modelagem 2 da seção 4.3 de Regressão Logística, foi rodada uma Regressão Linear em STATA®, usando o comando *stepwise* e nível de significância de 5%, que teve como variáveis preditoras as 9 categóricas anteriores e a variável resposta CAGR 2013/2017 de faturamento.

Os resultados de Regressão Linear foram adequados, bem superiores à modelagem anterior, com o teste F inferior a 0,01% e R^2 igual a 9,34%; foram satisfatórios os testes de inexistência de multicolinearidade e heterocedasticidade; ficaram no modelo final quatro variáveis com impacto significativo, pela ordem de magnitude dos coeficientes: relacionamento com aceleradoras; obtenção de aporte de capital de terceiros, que consolidou os vários fundos de investimento coletados na pesquisa; se a empresa era uma *spin-off* de outra; e cidade da sede da empresa em São Paulo Capital; essas variáveis tiveram impacto positivo para o crescimento médio em faturamento e p-values menores ou igual a 1%.

Houve grande consistência das duas modelagens, logística e linear, em relação às variáveis de relacionamento com aceleradoras, empresas *spin-offs* e sede da empresa em São Paulo

Capital. A novidade dessa regressão foi a variável de aporte de capital de terceiros, que para essa modelagem teve impacto positivo para o crescimento médio em faturamento das empresas.

Modelagem 3: Regressão Linear CAGR 2013/2017 de Faturamento, Percepção dos Empreendedores e Variáveis Categóricas

Por último, foi feita a Regressão Linear com todas as variáveis juntas, ou seja, as 16 de percepções e as nove categóricas, e a variável resposta CAGR 2013/2017 de faturamento das empresas. A regressão foi rodada em STATA®, utilizando o comando *stepwise* e nível de significância 10%. Por se tratar da modelagem de Regressão Linear mais importante para a presente pesquisa, seja pelos achados, seja para verificar a consistência geral com a modelagem 3 de Regressão Logística, os resultados são apresentados de forma mais detalhada a seguir.

A modelagem 3 revelou-se adequada, dado que o teste F resultou em probabilidade inferior a 0,01% e o R^2 em 10,30%. Das 25 variáveis iniciais, foram eliminadas 19, resultando no modelo final as seis que formam a Tabela 31, com os respectivos coeficientes e significâncias estatísticas; foram adequados os testes de inexistência de multicolinearidade e heterocedasticidade.

Tabela 31 – CAGR 2013/2017 de Faturamento, Percepção dos Empreendedores e Variáveis Categóricas: Regressão Linear – Coeficientes

Variável	Coeficiente	Erro Padrão	t	P > t	Intervalo de Confiança 95%	
s_spin	0,0760	0,0304	2,50	0,013	0,0162	0,1358
cid_spc	0,0758	0,0214	3,54	0,000	0,0338	0,1179
l_mod	0,0133	0,0058	2,29	0,023	0,0019	0,0246
cap_ter	0,1290	0,0471	2,74	0,006	0,0365	0,2216
l_d_fom	-0,0048	0,0024	-1,99	0,048	-0,0095	0,0000
s_r_ace	0,1656	0,0511	3,24	0,001	0,0652	0,2660
_cons	0,0344	0,0515	0,67	0,504	-0,0668	0,1355

Número de obs = 456; F(6, 449) = 8,59; Prob > F = 0,0000 R² = 0,1030 R² Ajustado = 0,0910

Fonte: elaborado pelo autor.

As variáveis de percepção de empreendedores relacionadas a “Inovar ou diferenciar modelo de negócios” (l_mod) e “Obter doações de instituições de fomento ou de outras empresas”

(l_d_fom) apresentaram coeficientes significantes, o primeiro positivo e o segundo negativo. As variáveis categóricas significantes para o modelo final, pela ordem de magnitude dos coeficientes, foram: “Mantém ou manteve relacionamento com alguma aceleradora” (s_r_ace), “Recebeu aporte de capital de terceiros” (cap_ter), “É *spin-off* de alguma instituição ou de outra empresa” (s_spin) e “Cidade onde fica a sede de sua empresa em São Paulo Capital” (cid_spc); a variável relativa a aporte de capital de terceiros engloba os vários fundos de investimento coletados na pesquisa; as quatro variáveis categóricas tiveram impacto positivo.

Em relação à modelagem 3 de Regressão Logística, os resultados da Tabela 31 confirmaram a importância e a natureza do impacto de três variáveis: relacionamento com aceleradoras, empresas *spin-offs* e obtenção de doações de instituições de fomento ou de outras empresas; por outro lado, aportaram três variáveis novas, todas com impacto positivo no crescimento médio em faturamento das empresas: percepção da importância em inovar ou diferenciar modelo de negócios e as categóricas de aporte de capital de terceiros e localização da sede da empresa na cidade de São Paulo. De um modo geral, registrou-se adequada consistência dos resultados dessa modelagem com as anteriores.

Foi feita a substituição das variáveis originais de percepção dos empreendedores pelos cinco fatores, seguindo a mesma linha anterior, para verificar consistência dos dois tipos de regressão. Essa modelagem foi adequada com o teste F inferior a 0,01% e R^2 de 10,25%, passando nos testes de inexistência de multicolinearidade e heterocedasticidade.

Mais uma vez, quando utilizados os fatores gerados para as 470 empresas da amostra, fica na modelagem final o Fator 4 Inovação, representando as variáveis “Inovar ou diferenciar produtos e serviços” (l_prod), “Inovar ou diferenciar design de produtos e serviços” (l_desig) e “Inovar ou diferenciar modelo de negócios” (l_mod); essa variável apresentou coeficiente positivo e p-value de 3,40%.

Além do Fator 4 Inovação, o modelo também mostrou significância estatística das variáveis categóricas de relacionamento com aceleradoras, aporte de capital de terceiros, *spin-offs* de outras empresas e sede na cidade de São Paulo com coeficientes positivos e p-values inferiores a 5%; a variável relativa a empréstimos bancários apresentou coeficiente negativo e p-value de 6,30%. Portanto, foi comprovada adequada consistência com as modelagens anteriores.

Repetidas as análises utilizando os fatores gerados com as percepções apenas dos empreendedores das 91 EACs, nenhum dos cinco fatores foram estatisticamente significantes ao nível de 10% no modelo final, não agregando novos achados à pesquisa.

A consistência entre os modelos de Regressão Logística e Regressão Linear não foi perfeita e nem poderia ser em função de seus distintos objetivos e fundamentos estatísticos; no entanto, serviu para confirmar que algumas variáveis são relevantes tanto para aumentar a probabilidade de EACs quanto para influenciar positivamente o crescimento médio em faturamento das empresas; esse fato confere maior robustez aos achados da presente pesquisa, conforme discutido na seção 4.5.

Modelagens com Variáveis Adicionais

A base de dados da pesquisa incluiu diversas variáveis que não fizeram parte das três modelagens de Regressão Logística e Linear detalhadas anteriormente. Dentre elas, as principais foram: número de sócios, separados por gênero e formação acadêmica, e as características de perfil dos empreendedores. As variáveis relacionadas ao número de sócios foram detalhadas na Tabela 6 e perfil dos empreendedores na Tabela 9.

Houve um problema na coleta de dados relacionados ao número total de sócios das empresas, separados por gênero; esse total deveria ser rigorosamente idêntico à soma de sócios com pós-graduação, graduação ou sem graduação; isso não ocorreu em 131 das 470 empresas, ou seja, 27,9% dos respondentes. Percebeu-se que tais respondentes não entenderam corretamente o formulário de pesquisa e prejudicaram a coleta de tais informações de forma correta.

Restaram três possibilidades para analisar essas variáveis; i) utilizar os dados originais respondidos com precaução, dado que para 131 empresas havia erros; ii) fazer análises apenas para as 339 empresas que apresentaram dados consistentes; iii) abandonar as análises relacionadas ao número de sócios, gênero ou formação acadêmica. Entendeu-se que o terceiro caminho era o mais indicado, para manter o rigor metodológico desta tese. Assim, as análises dessas variáveis foram abandonadas, devendo ser objeto de futuros estudos.

Quanto ao perfil dos empreendedores, as quatro variáveis da Tabela 9, questionavam em uma escala de 0 a 10, como os respondentes avaliavam a vida hoje, daqui a cinco anos e se em uma autoavaliação ele se considerava uma pessoa empreendedora; uma quarta variável, de

natureza categórica, sobre os motivos que levaram o empreendedor a começar seu negócio próprio, permitia três respostas pré-definidas: i) achou que tinha uma oportunidade; ii) teve necessidade; iii) ambas as alternativas.

Feitas as análises com essas três variáveis utilizando Regressão Logística e a variável resposta EACs, identificou-se que apenas a variável “Como você avalia sua vida hoje?” mostrava significância estatística; as demais três variáveis foram eliminadas da modelagem com nível de significância de 10%. No entanto, percebeu-se que a variável sobre a vida hoje aporta um viés do tipo “referência circular”: os empreendedores das empresas de alto crescimento têm uma avaliação mais positiva da vida hoje; a recíproca também é verdadeira; o achado foi considerado de pouco interesse acadêmico e por essa razão é apenas citado, sem maiores detalhes.

Assim, os achados mais relevantes da pesquisa foram obtidos nas modelagens feitas com Regressão Logística, detalhadas na seção 4.3 e discutidos na seção 4.5 desse capítulo a seguir.

4.6 Comentários sobre a Questão de Pesquisa e Hipóteses

A seção final desse capítulo procurou fazer um resumo dos achados nas diversas análises de dados descritas nas seções 4.1 a 4.5 da presente tese e, principalmente, avaliar se as respostas encontradas para a questão de pesquisa e se as hipóteses formuladas na seção 2.5 foram ou não confirmadas.

Como resumo dos achados, foram encontradas variáveis relevantes que contribuem para aumentar a probabilidade das empresas atingirem alto crescimento (EACs), ou pelo menos, que ajudam a entender e explicar as EACs. Tais variáveis preditoras de EACs são de dois tipos básicos: variáveis métricas ligadas à percepção dos empreendedores sobre diversas ferramentas de gestão e variáveis categóricas identificando características demográficas ou de relacionamento das empresas.

No primeiro bloco de achados, os resultados detalhados na modelagem 1 de Regressão Logística, indicaram que a percepção dos empreendedores sobre a importância atribuída ao registro da marca comercial ou à obtenção de empréstimos bancários está positivamente relacionada à probabilidade de EACs; em outras palavras, quanto maior a percepção dos empreendedores sobre essas duas variáveis, maior a probabilidade de EACs. Ao contrário,

quanto maior a percepção sobre a importância em fazer propaganda em mídia digital ou em obter doações de instituições de fomento ou de outras empresas, menor a probabilidade de EACs.

Quanto às variáveis categóricas da modelagem 2 de Regressão Logística verificou-se que cinco características das firmas estão positivamente relacionadas às probabilidades de EACs: empresas que receberam premiação, estão localizadas em São Paulo Capital, receberam empréstimos bancários, são oriundas de outras empresas (*spin-offs*) e tiveram relacionamento com aceleradoras. Pela magnitude dos coeficientes destacou-se o relacionamento com aceleradoras.

Quando combinadas as variáveis métricas de percepção dos empreendedores com as variáveis categóricas de características das empresas, conforme detalhado na modelagem 3 de Regressão Logística, cinco impactaram a probabilidade de EACs: duas variáveis de percepção dos empreendedores e três categóricas. A percepção sobre a importância em obter empréstimos bancários aumenta a probabilidade de EACs, enquanto que a percepção sobre doações de instituições de fomento diminui; em complemento, as empresas que receberam premiações, são *spin-offs* de outras ou possuem relacionamento com aceleradoras tem maior probabilidade de atingirem alto crescimento.

Os resultados obtidos por meio de Regressão Logística, com as variáveis originais do banco de dados, foram validados pelo uso de fatores, gerados em Análise Fatorial, para as 16 variáveis de percepção dos empreendedores; a razoável consistência entre as duas abordagens conferiu maior robustez aos achados.

O uso de Regressão Linear como segundo modelo preditivo, usando a variável resposta métrica – CAGR 2013/2017 de faturamento – que representa o crescimento médio anual em faturamento das empresas durante cinco anos, conforme detalhado na seção 4.4, serviu para confirmar os resultados e testar a validade da pesquisa; dentre as confirmações, conforme detalhado na Tabela 31, a percepção dos empreendedores sobre a obtenção de doações de instituições de fomento ou de outras empresas impactou de forma negativa o crescimento médio em faturamento, enquanto que o relacionamento com aceleradoras e ser *spin-off* de outra empresa contribuíram positivamente; as modelagens com Regressão Linear ajudaram a aumentar a validade dos achados principais da tese com o modelo Regressão Logística.

As análises sobre as gazelas foram limitadas, conforme explicado na seção 4.4, mas trouxeram alguns esclarecimentos sobre esse raro tipo de novas empresas, de alto crescimento e com menos de cinco anos de idade. As gazelas parecem seguir os mesmos padrões das EACs, de forma ainda mais pronunciada. Embora apenas seis tenham sido identificadas na pesquisa, ajudaram a dar maior validade às conclusões relativas às EACs.

Com base nos resultados, pode-se afirmar que a resposta à questão de pesquisa:

Quais são as variáveis relevantes que ajudam a explicar o desempenho das empresas de alto crescimento (EACs) no Brasil?

foi parcialmente alcançada; a parcialidade é inerente à base empírica da pesquisa, fundamentada em uma determinada amostra de empresas brasileiras, retratadas em corte transversal em dado momento de suas trajetórias. Feita a ressalva anterior, como resposta à questão de pesquisa, pode-se inferir das análises estatísticas que:

- 1) as variáveis categóricas relevantes que ajudaram a explicar o desempenho das EACs foram: i) o relacionamento com aceleradoras; ii) o recebimento de premiações; iii) a obtenção de empréstimos bancários; iv) ser *spin-off* de outra empresa.
- 2) as variáveis métricas relevantes, de percepção dos empreendedores sobre a importância das ferramentas de gestão, que ajudaram a explicar o desempenho das EACs foram: i) obter empréstimos bancários; ii) registrar marcas comerciais;
- 3) a variável de percepção dos empreendedores sobre a importância em obter doações de instituições de fomento ou de outras empresas teve impacto negativo para explicar o desempenho das EACs.

Em relação às cinco hipóteses formuladas na seção 2.5, *sobre as possíveis variáveis relevantes que ajudam a explicar as EACs*, em decorrência dos achados já detalhados nas seções 4.1 a 4.4 e resumidos nesta seção, inferiu-se que:

Hipótese 1: As empresas localizadas em parques tecnológicos, incubadoras ou espaços de coworking têm maior probabilidade de apresentar alto crescimento.

Não Confirmada

Hipótese 2: As empresas que recebem aporte de capital de investidores privados (*angel*, capital semente, *venture capital* e outros) têm maior probabilidade de apresentar alto crescimento.

Não Confirmada

Hipótese 3: As empresas que se relacionam com aceleradoras têm maior probabilidade de apresentar alto crescimento.

Confirmada

Hipótese 4: As empresas que inovam em produtos e serviços ou em modelo de negócios têm maior probabilidade de apresentar alto crescimento.

Não Confirmada

Hipótese 5: As empresas formadas a partir de outras – *spin-offs* – têm maior probabilidade de apresentar alto crescimento.

Confirmada

A Hipótese 1 não foi confirmada, dado que as firmas sediadas em parques tecnológicos, incubadoras ou espaços de coworking presentes na amostra desta pesquisa não registraram maior probabilidade de alto crescimento que as demais. A não confirmação desta hipótese não implica que o local da sede de empresa não exerça influência no seu desempenho; de acordo com a literatura, parte da qual citada na seção 2.5 desta tese, ambientes favoráveis ao

relacionamento entre empresas e os serviços de apoio prestados por parques tecnológicos, incubadoras ou espaços de coworking são muito valorizados pelos empreendedores e devem contribuir para o desempenho das empresas.

A Hipótese 1 não se confirmou provavelmente porque a base de dados da presente pesquisa apresentou baixa incidência de empresas em ambientes favoráveis, como mostrado na Tabela 12; com efeito, apenas 31 empresas das 470 da amostra (6,6%) eram localizadas em parques tecnológicos, incubadoras ou espaços de coworking; dentro desse pequeno grupo, somente seis eram EACs, o que limitou a modelagem estatística, não permitindo a confirmação da hipótese.

A Hipótese 2 não foi confirmada uma vez que as empresas que receberam aporte de capital de investidores privados, excetuando-se empréstimos bancários, não apresentaram maior probabilidade de apresentar alto crescimento. De forma análoga aos comentários relativos à Hipótese 1, foi baixa a taxa de incidência de empresas com aporte de capital na amostra; conforme Tabela 13 e discutido na seção 4.3, modelagem 2, 22 empresas da amostra (4,7%) receberam aportes de capital de terceiros, das quais somente duas apresentaram alto crescimento.

Na modelagem 3 com Regressão Linear, detalhada na seção 4.5, registrou-se indícios positivos da variável aporte de capital de terceiros sobre o crescimento médio em faturamento das empresas; de forma diferente da Regressão Logística, o modelo considerou todas as 22 empresas da amostra para estimar seu efeito; esse resultado confirmou os estudos empíricos da literatura.

A Hipótese 3 foi confirmada nas modelagens 2 e 3 de Regressão Logística e na terceira de Regressão Linear; o relacionamento com aceleradoras contribuiu favoravelmente tanto para aumentar a probabilidades das EACs, quanto para o crescimento médio em faturamento das empresas. Esse resultado, alinhado com a maioria dos estudos empíricos sobre o papel das aceleradoras no desempenho das empresas, embora esperado e intuitivo, foi relevante para as implicações acadêmicas e gerenciais detalhadas no capítulo 5.

A Hipótese 4 não foi confirmada, pois as variáveis envolvidas não foram significantes nas várias modelagens de Regressão Logística; nas modelagens 1 e 3 de Regressão Linear, houve

indícios positivos de inovação em modelo de negócios para o crescimento médio em faturamento das empresas, em linha com Segarra e Teruel (2014).

Avaliar se de fato uma empresa está oferecendo produtos e serviços inovadores para o mercado, ou se possui um novo modelo de negócios que agregue valor para seus clientes, é uma tarefa de grande complexidade; demandaria aprofundar a análise das empresas da amostra, o que é praticamente inviável em estudos quantitativos; por essa razão, a presente tese baseou-se exclusivamente nas percepções dos próprios empreendedores sobre a importância da inovação para o desempenho das empresas.

A grande maioria dos empreendedores, tanto das EACs como das não EACs, manifestou percepção positiva sobre a importância atribuída à inovação, o que limitou os resultados das modelagens com Regressão Logística; quando utilizado o modelo Regressão Linear foi possível identificar de forma relevante o impacto positivo de inovação em modelo de negócios, para o crescimento em faturamento das empresas, como detalhado na seção 4.5.

Em relação à Hipótese 5, houve confirmação por ambos os modelos Regressão Logística e Regressão Linear; as empresas *spin-offs* foram mais propensas a obter resultados positivos, em linha com os estudos de Andersson e Klepper (2013) e Feeser e Willard (1989). Embora fosse um resultado esperado, a confirmação dessa hipótese, baseada em testes estatísticos, serviu para validar que o contexto das empresas brasileiras é similar a de países desenvolvidos, conforme comentado no capítulo 5.

Das cinco hipóteses iniciais, portanto, duas foram confirmadas; esse resultado, embora possa parecer limitado, foi fruto da complexidade do problema de pesquisa e das hipóteses relativamente ambiciosas da presente tese. Obter validação estatística, com nível de significância adequado, para a identificação das variáveis críticas de desempenho de empresas, não é tarefa trivial.

Assim, os achados obtidos na presente pesquisa devem ser avaliados com base nas contribuições que podem aportar ao conhecimento acadêmico, suas implicações gerenciais e sugestões de futuras pesquisas, como discutido no capítulo 5.

5 CONSIDERAÇÕES FINAIS

Este capítulo trata das conclusões e considerações finais da presente pesquisa, sendo constituído por quatro seções; a primeira detalha as conclusões principais; a segunda as implicações gerenciais dos achados de pesquisa; a seção 3 discute as limitações e precauções que devem ser tomadas ao interpretar os resultados desta pesquisa; finalmente, na seção 4, trás algumas sugestões para futuros estudos.

5.1 Conclusões Principais

A partir dos resultados estatísticos detalhados no capítulo 4 desta tese, as conclusões finais consideradas mais robustas e relevantes foram pinçadas e aqui reapresentadas, em função do problema de pesquisa:

Quais são as variáveis relevantes que ajudam a explicar o desempenho das empresas de alto crescimento (EACs) no Brasil?

A pesquisa empírica com 470 firmas representativas do empreendedorismo brasileiro identificou nove variáveis associadas às empresas de alto crescimento no país: i) cinco variáveis categóricas que descrevem as características das empresas; ii) quatro variáveis métricas de percepção dos empreendedores sobre a importância atribuída a ferramentas de gestão.

Pela ordem de impacto na probabilidade de EACs, as cinco variáveis categóricas identificadas foram o relacionamento com aceleradoras, o recebimento de premiações, *spin-offs* de outras empresas, obtenção de empréstimos bancários e localização da sede na cidade de São Paulo.

Foi confirmado o papel positivo das aceleradoras para o crescimento das empresas (Wise e Valliere, 2014), atestando a importância de tais entidades também para o empreendedorismo brasileiro. As aceleradoras são parte integrante do ecossistema de empreendedorismo e, na medida em que selecionam, orientam, prestam *mentoring* e aportam recursos financeiros, estão definitivamente contribuindo para o desenvolvimento das empresas e do país.

No Brasil existem cerca de 40 aceleradoras de startups em atividade de acordo com “O Panorama das Aceleradoras de Startups no Brasil”, publicado por Abreu e Campos (2016), com base em pesquisa conduzida pelo Centro de Estudos em *Private Equity* e *Venture Capital*

(GVcepe) em parceria com o Centro de Empreendedorismo e Novos Negócios (GVcenn), ambos da Escola de Administração de São Paulo (EAESP) da Fundação Getúlio Vargas (FGV). Segundo a referida fonte, a maior parte das aceleradoras de startups está localizada no Estado de São Paulo e até janeiro de 2016 tinham aportado cerca de R\$ 51 milhões em mais de 1.100 empresas que tinham sido aceleradas no Brasil. Dado que a existência de aceleradoras é um fenômeno relativamente novo no ecossistema de empreendedorismo no mundo e no Brasil, os resultados obtidos nesta pesquisa ajudam a comprovar a utilidade dessas instituições.

Em relação ao recebimento de premiações, embora tal variável não tivesse integrado as hipóteses desta tese, a comprovação de seu impacto positivo para as EACs foi um achado relevante.

A literatura indica que premiações estão associadas ao desempenho das empresas de forma positiva. Exemplo dessa associação foi relatado na pesquisa publicada por Gicheva e Link (2016), com 1.878 empresas que tinham recebido a premiação americana de Pesquisa e Desenvolvimento Fase II do programa *U.S. Small Business Innovation Research (SBIR)*; utilizando o modelo Probit, similar à Regressão Logística, concluíram que as empresas nascentes, baseadas no uso de novas tecnologias, tinham grande probabilidade de fracassarem em comparação com as empresas existentes; no entanto, apresentaram maior probabilidade de comercializarem suas tecnologias desenvolvidas; ou seja, a premiação foi associada a uma maior probabilidade de introdução de novas tecnologias no mercado.

Em outro estudo, Michelsen, Wolf e Schwartz (2013) focalizaram a fase de transição entre empreendedorismo nascente e efetivo, a partir de 103 novas empresas do setor de biotecnologia na Alemanha; dentre outros, coletaram dados dos empreendedores que tinham participado em competições de startups e recebido pelo menos uma premiação; comprovaram a hipótese que as oportunidades regionais e o ecossistema de empreendedorismo, associados a premiações de empreendedores, aumentaram a probabilidade de transição exitosa das empresas nascentes para startups reais.

Obviamente a variável recebimento de premiação não tem característica prescritiva, mas trata de uma sinalização da importância das premiações que o país disponibiliza; algo relativamente simples, mas que tudo indica estar sendo bem feito no Brasil; as premiações

aparentemente estão sendo atribuídas com critérios justos, dado que ajudam a explicar as EACs, e esse achado é estimulante.

Os resultados obtidos para as *spin-offs*, em linha com Andersson e Klepper (2013) e Feeser e Willard (1989), confirmaram que a experiência das empresas das quais são oriundas, contribuem para aumentar a probabilidade de apresentarem alto crescimento. Em um país complexo como o Brasil, onde abrir e operar uma empresa são atividades desafiadoras, o fato de uma nova empresa nascer a partir de outra já existente, faz uma grande diferença.

Era esperado que o papel de aportes de capital fosse relevante para o desempenho das empresas de alto crescimento, em linha com a literatura (Mateev e Anastasov, 2011; Politis 2008). No entanto, a Hipótese 2 não foi confirmada e, de certa maneira, substituída pela variável relacionada ao recebimento de empréstimos bancários.

Essa substituição pode ser considerada relativamente surpreendente, uma vez que as taxas de juros praticadas pelos bancos brasileiros estão entre as mais altas do mundo. Embora o setor bancário do país seja bem desenvolvido e competitivo, com diversos bancos de grande porte e bem estruturados, as taxas de juros reais são historicamente elevadas. Boa parte dos grandes bancos apoia explicitamente o empreendedorismo e alguns até oferecem futuristas espaços de coworking subsidiados para startups; no entanto, a oferta de empréstimos é restrita e, quando disponível, a taxas de juros pouco atrativas para pequenas e médias empresas. Talvez, por essa situação de assimetria, a presente pesquisa identificou essa variável como relevante para as EACs.

A localização da sede da empresa na cidade de São Paulo, embora com menor impacto e relevância estatística próxima a 10%, confirmou a percepção de se tratar da cidade que oferece melhores condições em geral para o desenvolvimento de empreendedorismo no Brasil. Vale ressaltar que se trata do município de São Paulo e não da área geográfica conhecida como Grande São Paulo, que reúne as cidades da região metropolitana, tais como Santo André, São Bernardo do Campo, São Caetano do Sul e Diadema.

Sabe-se que a cidade de São Paulo é o maior polo de negócios do Brasil, em função de sua importância econômica, concentração de boas universidades, tamanho de sua população e diversos outros aspectos; figura em alguns rankings globais entre as cidades com melhor ecossistema de empreendedorismo, o que foi de certa maneira confirmado nesta pesquisa.

Em relação ao bloco de quatro variáveis métricas, relacionadas à percepção dos empreendedores sobre a importância de algumas ferramentas de gestão, foram identificados os impactos positivos para as empresas de alto crescimento em obter empréstimos bancários e registrar marcas comerciais, e os impactos negativos em obter doações de instituições de fomento ou de outras empresas e fazer propaganda em mídia digital.

Segundo Fazio et al. (2016), em artigo comentado na seção 2.4, o registro de marca comercial de uma empresa americana estava associado a um relevante aumento da sua probabilidade de crescimento acelerado. Os resultados obtidos na presente tese confirmaram que também no Brasil parece existir a percepção da importância do registro da marca comercial por parte dos empreendedores de EACs. A utilização de marcas adequadas e que possam ser protegidas através de registro de propriedade intelectual é uma das principais ferramentas em estratégia de marketing de qualquer empresa; no entanto, essa conscientização aparentemente ainda não está permeada entre todos os empreendedores brasileiros. São poucas as marcas brasileiras conhecidas e fortes no exterior e, dessa forma, o resultado desta pesquisa causou certa surpresa.

A percepção dos empreendedores sobre a obtenção de doações de instituições de fomento ou de outras empresas, ao contrário, está associada de forma negativa às EACs. Tal fato, pouco citado na literatura, também não era esperado e parece indicar que os empreendedores entendem que o melhor caminho não é correr atrás de doações, mas de fontes de financiamento; essa última variável, em várias modelagens da presente tese, confirmou sua importância na visão dos empreendedores de EACs.

Por último, a variável métrica sobre a importância de fazer propaganda em mídia digital (internet, site, smartphone, aplicativos...) demonstrou ter impacto negativo para os empreendedores de EACs. Esse achado, de certo modo surpreendente ao senso comum, e pouco citado na literatura, demandaria uma pesquisa de campo mais aprofundada para ser corretamente interpretado; no entanto, deixa clara certa desilusão com a mídia digital.

Uma das possíveis explicações teóricas para o referido impacto negativo pode ser a grande pulverização dos meios digitais, que trás como consequência a necessidade de fazer elevados investimentos de marketing; outra justificativa pode ser a percepção que a propaganda digital é invasiva, pouco eficiente e cria limitado valor, dado que frequentemente é associada ao

conceito de *Outbound Marketing*, e não a *Inbound Marketing*⁴. Independentemente dos motivos, fica registrada certa frustração com a propaganda digital por parte dos empreendedores de EACs, enquanto que para os demais continua uma ferramenta valiosa.

As gazelas, empresas de alto crescimento com menos de cinco anos, identificadas nesta tese formaram um reduzido grupo; no entanto, a taxa de participação de gazelas obtida foi consistente com os resultados de grandes amostragens feitas nos Estados Unidos (Morelix e Russell-Fritch, 2017) e no Brasil (IBGE b, 2017), conforme comentado na seção 4.4.

Dada o reduzido tamanho da amostra de gazelas, não foi possível utilizar modelos estatísticos para o grupo; suas principais características descritivas foram consistentes com as demais EACs. As gazelas apresentaram de forma mais intensa taxas de incidência de relacionamento com aceleradoras, recebimento de premiações, serem spin-offs de outras empresas e obtenção de aportes de capital de aceleradoras e investidores *angel*; tais achados e sua consistência ajudaram a validar as conclusões da presente tese.

5.2 Implicações Gerenciais

Como implicações gerenciais, em primeiro lugar, pode-se concluir que o ecossistema brasileiro de empreendedorismo deve oferecer mais financiamento às empresas; esse achado acadêmico, alinhado com as percepções dos empreendedores, ficou evidenciado nas várias modelagens e testes estatísticos. Tais financiamentos podem ser através de bancos, instituições de fomento ou mesmo dos diversos atores do mundo do empreendedorismo, como investidores *angel* e fundos de capital semente, *venture capital* ou *private equity*. Quanto às taxas de juros, quanto mais competitivas comparadas com as taxas de países desenvolvidos, mais suporte estará sendo dado ao empreendedorismo.

Como segunda implicação gerencial, destaca-se a importância das aceleradoras, em atuação multifacetada, de apoio e financiamento inicial; conforme mencionado, existem cerca de 40 aceleradoras de startups no Brasil, que aportaram aproximadamente R\$ 51 milhões em 1,1 mil empresas, conforme pesquisa do GVcepe e do GVCenn da FGV (Abreu e Campos, 2016); são

⁴ *Inbound Marketing* refere-se a atrair e fidelizar clientes, de forma orgânica e espontânea, sem ir atrás deles, disponibilizando produtos, serviços e conteúdos que façam com que busquem a oferta da empresa; *Outbound Marketing* refere-se à busca de clientes de forma agressiva, sem se importar se eles estão interessados na oferta da empresa.

números extremamente relevantes, mas seguramente inferior ao potencial e a necessidade do ecossistema de empreendedorismo do país.

Outra implicação gerencial está relacionada ao registro de marcas comerciais e a limitada importância atribuída por parte dos empreendedores em geral, razão pela qual a variável foi relevante para as EACs; em recente visita ao país, Kotler (2013), um dos autores mais respeitados em marketing, afirmou que “o Brasil precisa de marcas globais” e parece que grande parte dos empreendedores brasileiros não percebeu isso.

Por último, mas não menos relevante, estimular os atores do ecossistema de empreendedorismo a continuarem, e se possível, aumentarem as premiações de empresas; mais que os valores pagos em premiações, servem para estimular os empreendedores e contribuir para o círculo virtuoso de crescimento das empresas. Em geral, as empresas que recebem alguma premiação, costumam divulgar tal reconhecimento público e parecem adquirir aumentada confiança para enfrentar os enormes desafios de empreender no Brasil.

5.3 Limitações

São várias as limitações da presente pesquisa, que buscou analisar dados de empresas e fazer inferências empíricas; as principais são discutidas nesta seção. O envio de formulários de pesquisas para respondentes – *surveys* – aporta diversas restrições intrínsecas ao processo de coleta de dados. Algumas delas, tais como o controle de não respondentes, foram endereçadas e parcialmente contornadas; embora certas precauções tenham sido tomadas, como amostra de controle, entrevistas pessoais com alguns empreendedores e outras citadas na seção 3.2, não se pode assegurar que uma amostragem não probabilística seja isenta de imperfeições. Outra limitação foi evitada ao serem descartadas da amostra final as empresas respondentes por intermédio da Startup Farm; dado que tais empresas tiveram algum tipo de relacionamento com a própria aceleradora, poderiam aportar um viés às análises; a utilização de respondentes exclusivamente da Endeavor favoreceu a independência dos dados coletados.

Devem ser mencionadas as limitações relacionadas com a qualidade e acurácia das respostas coletadas; não existe uma garantia que os dados fornecidos pelos respondentes sejam corretos e reflitam a realidade de cada uma das empresas; alguns cuidados foram tomados, tais como solicitar o nome da empresa e do respondente, dado que isso contribui para reduzir respostas

levianas. No entanto, mais efetivo controle de acurácia pode ser garantido pelo tamanho da amostragem e pelo uso de ferramentas estatísticas; são elas que conferem robustez aos achados.

Outra limitação vem do fato desta pesquisa representar um corte transversal, com dados coletados das empresas em um determinado momento, mais precisamente em setembro de 2017; portanto, representou uma fotografia, identificada em parte por respostas objetivas, em parte por um conjunto de percepções dos empreendedores; percepções podem mudar com o tempo e com as condições dos ambientes interno e externo das empresas. Como se tratou de uma pesquisa quantitativa com 470 empresas, adequada para fazer algumas inferências estatísticas, essa limitação foi parcialmente contornada.

Por último, vale lembrar que a complexidade do problema de pesquisa e a natureza do processo de amostragem impediram comprovar resultados esperados, tais como a importância da localização em ambientes favoráveis, inovação ou aporte de capital, conforme comentado na seção 4.6. Isso não significa que tais variáveis não sejam relevantes, mas simplesmente que as análises empíricas não puderam comprová-las. Esse conjunto de limitações principais é a base para futuros estudos que estão descritos na última seção desta tese.

5.4 Estudos Futuros

Temas complexos, como encontrar as variáveis relevantes para o desempenho de empresas de alto crescimento, objeto central desta tese, tem como vantagem identificar de forma simples e rápida avenidas para futuras pesquisas. A própria base de dados obtida com as 470 empresas pode conter outros achados relevantes, que não fizeram parte desta pesquisa; como exemplo, uma análise mais profunda das empresas de alto crescimento, identificando o setor de atuação ou o grau efetivo de inovação, por meio de consulta a sites e pesquisas com dados secundários, pode apresentar bom potencial para futuros estudos.

Complementar as análises com algumas outras variáveis, tais como o número de sócios e formação acadêmica dos mesmos, antes procedendo à depuração das inconsistências verificadas, é outra área de oportunidades.

Dado que as empresas de alto crescimento são temporárias, o acompanhamento longitudinal das EACs pode conduzir a achados relevantes; a Endeavor tem uma reputação tão forte com

as empresas cadastradas em seu *mailing*, que não seria difícil, com a colaboração da entidade, dar sequência a esta pesquisa com uma segunda rodada de coleta de dados.

Novos estudos centrados em gazelas são recomendados; a presente tese, de natureza quantitativa, mostrou-se inadequada para elaborar análises mais profundas sobre esse fenômeno da área de empreendedorismo; em função da baixa incidência e grande variabilidade de atuação gerencial das gazelas, o método de estudo de caso parecer ser o mais indicado para futuras pesquisas sobre tais empresas.

A busca do conhecimento acadêmico é um *never ending task* e a frase atribuída ao filósofo grego Sócrates – “Só sei que nada sei!” – ajuda a contextualizar esta seção final. Mesmo diante da impressão do limitado avanço desta tese em relação ao problema de pesquisa, espera-se que possa ter trazido um pouco mais de luz para o conhecimento acadêmico sobre empreendedorismo no Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aaker, D. A., Kumar, V., & Day, G. S. (2013). *Marketing research*. 11 ed. John Wiley & Sons.
- Abreu, P.R.M and Campos, N.M.C (2016). *O panorama das aceleradoras de startups no Brasil*. Centro de Estudos em Private Equity e Venture Capital (GVcepe). CreateSpace Independent Publishing Plataform, USA.
- ABVCAP & KPMG (2015). Associação Brasileira de Private Equity e Venture Capital e KPMG. Consolidação de Dados da Indústria de Private Equity e Venture Capital no Brasil 2011 2012 2013 2014. Disponível em <http://www.abvcap.com.br/Download/Estudos/3233.pdf>
- ABVCAP e KPMG (2016). Associação Brasileira de Private Equity e Venture Capital e KPMG. Consolidação de Dados da Indústria de Private Equity e Venture Capital no Brasil 2011 2012 2013 2014 2016. Disponível em <http://www.abvcap.com.br/Download/Estudos/3484.pdf>
- Andersson, M., & Klepper, S. (2013). Characteristics and performance of new firms and spinoffs in Sweden. *Industrial and Corporate Change*, 22(1), 245-280.
- Andreassi, T., & Sbragia, R. (2002). Relações entre indicadores de P&D e de resultado empresarial. *Revista de Administração da Universidade de São Paulo*, 37(1).
- Anprotec a (2016). Estudo de Impacto Econômico: segmento de incubadoras de empresas no Brasil. Associação Nacional de Entidades Promotoras de Empreendimentos Inovadores. Brasília. Anprotec, Sebrae.
- Anprotec b (2016). Conceito de Incubadoras e Parques Tecnológicos. Associação Nacional de Entidades Promotoras de Empreendimentos Inovadores. Disponível em <http://anprotec.org.br/site/menu/incubadoras-e-parques/>
- Arruda, C., Nogueira, V., Cozzi, A., & Costa, V. (2014). Causas da mortalidade de startups brasileiras: o que fazer para aumentar as chances de sobrevivência no mercado. *Núcleo de Inovação e Empreendedorismo: Fundação Dom Cabral*.
- Barbero, J. L., Casillas, J. C., Ramos, A., & Guitar, S. (2012). Revisiting incubation performance: How incubator typology affects results. *Technological Forecasting and Social Change*, 79(5), 888-902.
- Bell, C. G., & McNamara, J. E. (1991). *High-tech ventures: The guide for entrepreneurial success*. Basic Books.
- Bertels, H. M., Koen, P. A., & Elsum, I. (2015). Business models outside the core: Lessons learned from success and failure. *Research-Technology Management*, 58(2), 20-29.
- Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, 34(3), 279-298.
- Birch, D. L. (1981). Who creates jobs? *The Public Interest*, 65(Fall), 3–14.

- Blank, S. (2006). The four steps to the epiphany: successful strategies for products that win. Foster City, Calif. *Cafepress.com*. Disponível em http://web.stanford.edu/group/e145/cgi-bin/winter/drupal/upload/handouts/Four_Steps.pdf
- Bravo-Biosca, A. (2011, October). A look at business growth and contraction in Europe. In *3rd European Conference on Corporate R&D and Innovation CONCORD-2011*.
- Bruneel, J., Ratinho, T., Clarysse, B., & Groen, A. (2012). The Evolution of Business Incubators: Comparing demand and supply of business incubation services across different incubator generations. *Technovation*, 32(2), 110-121.
- Bussab, W. D. O., & Morettin, P. A. (2010). *Estatística básica*. Saraiva.
- Bygrave, W. D., & Timmons, J. (1992). *Venture capital at the crossroads*. Boston, Massachusetts: Harvard Business School Press.
- Cahen, F. R., Lahiri, S., & Borini, F. M. (2016). Managerial perceptions of barriers to internationalization: An examination of Brazil's new technology-based firms. *Journal of Business Research*, 69(6), 1973-1979.
- Coad, A., Daunfeldt, S. O., Hözl, W., Johansson, D., & Nightingale, P. (2014). High-growth firms: introduction to the special section. *Industrial and Corporate Change*, 23(1), 91-112.
- Coworking (2016). Como funciona o coworking. Coworking Brasil. Disponível em <https://coworkingbrasil.org/como-funciona-coworking/>
- Daunfeldt, S. O., & Halvarsson, D. (2015). Are high-growth firms one-hit wonders? Evidence from Sweden. *Small Business Economics*, 44(2), 361-383.
- Deloitte (2017). As PMEs que Mais Crescem no Brasil. Pesquisa 2017. Em parceria com a Revista Exame. 12ª. edição. Disponível em <https://www2.deloitte.com/br/pt/pages/strategy/articles/pequenas-medias-empresas-mais-crescem-Brasil.html>
- Demir, R., Wennberg, K., & McKelvie, A. (2017). The strategic management of high-growth firms: a review and theoretical conceptualization. *Long Range Planning*, 50(4), 431-456.
- Endeavor (2016). Venture Capital: o que é e como conseguir. Endeavor Brasil. Disponível em <https://endeavor.org.br/venture-capital/>
- Erken, H., Donselaar, P., & Thurik, R. (2016). Total factor productivity and the role of entrepreneurship. *The Journal of Technology Transfer*, 1-29.
- Ewing Marion Kauffman Foundation (2016). The Economic Impact of High-Growth Startups. *Entrepreneurship Policy Digest*, June, 07. Disponível em <http://www.kauffman.org/newsroom/2016/06/understanding-the-economic>
- Fávero, L. P. (2015). *Análise de dados: modelos de regressão com EXCEL®, STATA® e SPSS®*. 1ª. ed. – Rio de Janeiro: Elsevier.
- Fávero, L. P. e Belfiore, P. (2015). *Análise de dados. Técnicas Multivariadas Exploratórias*. 1ª. ed. – Rio de Janeiro: Elsevier.

- Fazio, C., Guzman, J., Murray, F., & Stern, S. (2016). A New View of the Skew: Quantitative Assessment of the Quality of American Entrepreneurship. Cambridge, MA: MIT Innovation Initiative. Disponível em <http://www.kauffman.org/neg/section-3#anewviewoftheskew>
- Feeser, H. R., & Willard, G. E. (1989). Incubators and performance: A comparison of high- and low-growth high-tech firms. *Journal of Business Venturing*, 4(6), 429-442.
- Gicheva, D., & Link, A. N. (2016). On the economic performance of nascent entrepreneurs. *European Economic Review*, 86, 109-117.
- Global Entrepreneurship Monitor (1999). *Executive report*. Kauffman Center for Entrepreneurial Leadership at the Ewing Marion Kauffman Foundation.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate Data Analysis*. 7th ed. NY: Pearson.
- Henrekson, M., & Johansson, D. (2010). Gazelles as job creators: a survey and interpretation of the evidence. *Small Business Economics*, 35(2), 227-244.
- Hirakawa, O., Muendler, M. A., & Rauch, J. E. (2010). *Employee spinoffs and other entrants: Stylized facts from Brazil* (No. w15638). National Bureau of Economic Research.
- IBGE a (2017). *População Brasileira*. Instituto Brasileiro de Geografia e Estatística. Disponível em <https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html>
- IBGE b (2017). *Estatísticas de Empreendedorismo 2015*. Instituto Brasileiro de Geografia e Estatística. Disponível em <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101312.pdf>
- Isenberg, D. J. (2014). What an entrepreneurship ecosystem actually is. *Harvard Business Review*, 5, 1-7.
- Isenberg, D., & Onyemah, V. (2016). Fostering Scale Up Ecosystems for Regional Economic Growth (Innovations Case Narrative: Manizales-Mas and Scale Up Milwaukee).
- Kotler, P. (2013) “O Brasil precisa de marcas globais”, diz guru de marketing. Artigo publicado pelo jornal O Estado de São Paulo, em 13 nov. 2013. Disponível em <http://economia.estadao.com.br/noticias/geral,o-brasil-precisa-de-marcas-globais-diz-guru-do-marketing-imp-,1095478>
- Malhotra, N. K. (2012). *Pesquisa de marketing: uma orientação aplicada*. Bookman Editora.
- Mason, C., & Brown, R. (2014). Entrepreneurial ecosystems and growth oriented entrepreneurship. *Final Report to OECD, Paris*, 30(1), 77-102.
- Mateev, M., & Anastasov, Y. (2011). On the Growth of Micro, Small and Medium-Sized Firms in Central and Eastern Europe: A Dynamic Panel Analysis. *Banking & Finance Review*, 3(2).

- Michelsen, C., Wolf, H., & Schwartz, M. (2013). Regional Entrepreneurial Opportunities in the Biotech Industry: Exploring the Transition from Award-Winning Nascent Entrepreneurs to Real Start-Ups. *European Planning Studies*, 21(11), 1708-1734.
- Morelix, A., & Russell-Fritch, J. (2017). 2017 The Kauffman Index of Growth Entrepreneurship: National Trends. Disponível em <http://www.kauffman.org/kauffman-index/reporting/growth-entrepreneurship>
- Monteiro, G. F. D. A. (2017). Empresas de Alto Crescimento e o Desafio de Scale-up: onde estamos e para onde queremos ir. *Publicado por Endeavor Brasil*. Disponível em <https://www.insper.edu.br/wp-content/uploads/2017/04/White-Paper-Lancamento-Catedra-Endeavor.pdf>
- Motohashi, K. (2013). The role of the science park in innovation performance of start-up firms: an empirical analysis of Tsinghua Science Park in Beijing. *Asia Pacific Business Review*, 19(4), 578-599.
- Nakagawa, M. H. (2008). *Empresa inovadora de base tecnológica: um modelo de desenvolvimento para o contexto brasileiro* (Doctoral dissertation, FEA, Universidade de São Paulo, São Paulo, Brazil).
- OECD. (2005). Oslo Manual. *Guidelines for collecting and interpreting innovation data*.
- OECD. (2007). Eurostat, O. E. C. D. (2007). Eurostat-OECD manual on business demography statistics.
- OECD. (2010). *OECD Studies on SMEs and Entrepreneurship High-Growth Enterprises: What Governments Can Do to Make a Difference*. Organisation for Economic Co-operation and Development.
- Pavani, C. (2003). *O capital de risco no Brasil: conceito evolução perspectivas*. Editora E-papers.
- Politis, D. (2008). Business angels and value added: what do we know and where do we go?. *Venture capital*, 10(2), 127-147.
- Reynolds, P. D., Hay, M., & Camp, S. M. (1999). *Global entrepreneurship monitor* (p. 3). Kansas City, Mo.: Kauffman Center for Entrepreneurial Leadership.
- Robb, A., & Farhat, J. B. (2013). An Overview of the Kauffman Firm Survey: Results from 2011 Business Activities. *Available at SSRN 2277641*.
- Rosa, P. L. P. (2017). *Programas de conexão de empreendedores funcionam? Uma análise de impacto do Programa de Apoio da Endeavor Brasil* (Dissertação de Mestrado). Disponível em <http://bibliotecadigital.fgv.br/dspace/handle/10438/18805>
- Schumpeter, J. A. (1982). The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle (1912/1934). *Transaction Publishers*.— 1982.— January, 1, 244.

- Sebrae (2016). Diferenças entre Incubadora e Aceleradora. Disponível em <http://www.sebrae.com.br/sites/PortalSebrae/artigos/entenda-a-diferenca-entre-incubadora-e-aceleradora,761913074c0a3410VgnVCM1000003b74010aRCRD>
- Segarra, A., & Teruel, M. (2014). High-growth firms and innovation: an empirical analysis for Spanish firms. *Small Business Economics*, 43(4), 805-821.
- Singer, S., Amorós, J. E., & Arreola, D. M. (2015). Global entrepreneurship monitor 2014 global report. *Global Entrepreneurship Research Association*, 1-116.
- The Economist, Revista. (2017). *Gazelles in the heartland. American entrepreneurship is flourishing, if you know where to look*. Print ed. Business. September 30, 2017. Disponível em <https://www.economist.com/news/business/21729781-high-growth-kind-has-rebounded-sharply-lows-after-great-recession-american>
- The World Bank (2016). GDP and Per-Capita GDP data. Disponível em <https://data.worldbank.org/country/brazil?locale=pt> e <http://pt.knoema.com/sijweyg/world-gdp-per-capita-ranking-2017-data-and-charts-forecast>
- Van Gelderen, M., Thurik, R., & Bosma, N. (2006). Success and risk factors in the pre-startup phase. *Small Business Economics*, 26(4), 319-335.
- Wise, S., & Valliere, D. (2014). The impact on management experience on the performance of start-ups within accelerators. *The Journal of Private Equity*, 9-19.

APÊNDICE 1

Formulário de Pesquisa

Pesquisa Endeavor & FEA / USP sobre Empreendedorismo no Brasil

Agradecemos muito sua participação nesta pesquisa! Nosso objetivo é identificar como diferentes fatores influenciam o crescimento das empresas.

O tempo estimado para o preenchimento da pesquisa é cerca de 10 minutos.

Todas as respostas de sua empresa serão mantidas confidenciais e sob nenhuma hipótese serão divulgadas; somente os resultados consolidados e conclusões da pesquisa serão compartilhados.

*Obrigatório

1. 1. Nome e sobrenome do respondente da pesquisa *

2. 2. Cargo do respondente da pesquisa na empresa *

3. 3. e-mail do respondente da pesquisa (opcional)

4. 4. Nome fantasia da sua empresa *

5. 5. CNPJ da sua empresa *

caso ainda não possua CNPJ ou não queira informar, preencher com "na"

6. 6. Cidade onde fica a sede da sua empresa *

7. 7. Local onde fica a sede atual da sua empresa *

Marcar apenas uma oval.

- ☐ Parque tecnológico privado
☐ Parque tecnológico público
☐ Incubadora privada
☐ Incubadora pública
☐ Espaço de coworking
☐ Escritório próprio *Ir para a pergunta 9.*
☐ Outro: _____ *Ir para a pergunta 9.*

Seção 2

8. 8. Nome da instituição onde fica a sede da sua empresa *

Seção 3

9. 9. Sua empresa mudou alguma vez de endereço? *

Marcar apenas uma oval.

- ☐ Sim
☐ Não *Ir para a pergunta 11.*

Seção 4

10. 10. Local onde ficava anteriormente a sede da sua empresa *

Marcar apenas uma oval.

- ☐ Parque tecnológico privado
☐ Parque tecnológico público
☐ Incubadora privada
☐ Incubadora pública
☐ Espaço de coworking
☐ Escritório próprio
☐ Outro: _____

Seção 5

11. 11. Produtos e serviços da sua empresa *

por favor, informe os produtos e serviços que a sua empresa comercializa(rá) no mercado

12. **12. Modelo de negócios atual da sua empresa ***

por favor, explique o modelo de negócios atual da sua empresa

13. **13. Públicos-alvo da sua empresa ***

por favor, indique os clientes prioritários de sua empresa (pessoas físicas e jurídicas)

14. **14. Data de fundação da sua empresa ***

por favor, informe o mês e ano aproximados de obtenção do CNPJ da sua empresa; caso não tenha o CNPJ, preencher com "na"

15. **15. Data do primeiro faturamento da sua empresa ***

por favor, informe o mês e o ano aproximados do primeiro faturamento da sua empresa; caso não tenha iniciado o faturamento, preencher com "na"

16. **16. Quantidade e gênero dos sócios da sua empresa ***

por favor, indique a quantidade de sócios pessoas físicas da sua empresa, separando em homens e mulheres

Marcar apenas uma oval por linha.

	0	1	2	3	4	5 ou mais
Homens	☐	☐	☐	☐	☐	☐
Mulheres	☐	☐	☐	☐	☐	☐

17. **17. Formação acadêmica dos sócios da sua empresa ***

por favor, indique a quantidade de sócios pessoas físicas da sua empresa, separando pela formação acadêmica em faculdades

Marcar apenas uma oval por linha.

	0	1	2	3	4	5 ou mais
Sócios com curso de pós-graduação	☐	☐	☐	☐	☐	☐
Sócios com curso de graduação	☐	☐	☐	☐	☐	☐
Sócios sem curso de graduação	☐	☐	☐	☐	☐	☐

18. **18. Sua empresa é spin-off de alguma instituição ou de outra empresa? ***

Marcar apenas uma oval.

☐

Sim

☐

Não

Ir para a pergunta 20.

19. **19. Indicar o nome da instituição ou da outra empresa da qual sua empresa é uma spin-off ***

Seção 7

20. **20. Sua empresa mantém ou manteve relacionamento com alguma aceleradora? ***

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não

21. **21. Sua empresa recebeu aporte de capital de: ***

por favor, indique todas as alternativas aplicáveis

Marcar apenas uma oval por linha.

	Sim	Não
Sócios	☐	☐
Familiares ou amigos	☐	☐
Crowdfunding	☐	☐
Investidores Angel	☐	☐
Aceleradoras	☐	☐
Fundos de Capital Semente	☐	☐
Fundos de Venture Capital	☐	☐
Fundos de Private Equity	☐	☐
Empréstimos Bancários	☐	☐

22. **22. Sua empresa recebeu doação de alguma instituição ou empresa? ***

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não *Ir para a pergunta 24.*

Seção 8

23. **23. Indicar os nomes das instituições ou empresas que fizeram doação para sua empresa ***

Seção 9

24. Avalie a importância de cada item a seguir para o crescimento de sua empresa; considere 0 = pouca importância e 10 = muita importância

38. Obter empréstimos bancários *

por favor, indique a importância de sua empresa contar com empréstimos de bancos
Marcar apenas uma oval.

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

39. Obter doações de instituições de fomento ou de outras empresas *

por favor, indique a importância de sua empresa contar com doações de instituições de fomento ou de outras empresas
Marcar apenas uma oval.

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

40. 25. Sua empresa recebeu alguma premiação? *

por favor, indique se a sua startup recebeu algum reconhecimento público ou prêmio
Marcar apenas uma oval.

- ☐ Sim
☐ Não

26. Crescimento do faturamento anual

por favor, indique o percentual (%) de crescimento do faturamento em R\$ de sua empresa ano-a-ano; se em alguns anos sua empresa não teve faturamento, preencher com "na" o respectivo campo

41. Crescimento do faturamento 2014 em comparação com o faturamento 2013 (em %) *

42. Crescimento do faturamento 2015 em comparação com o faturamento 2014 (em %) *

43. Crescimento do faturamento 2016 em comparação com o faturamento 2015 (em %) *

44. Crescimento do faturamento previsto para 2017 em comparação com o faturamento 2016 (em %) *

27. Número de colaboradores

por favor, indique o número de colaboradores, incluindo sócios, funcionários e prestadores de serviços em tempo integral de sua empresa ao final de cada ano; se em alguns anos sua empresa não teve colaboradores, preencher com o número 0 (zero) o respectivo campo

45. Número de colaboradores em 31/12/2013 *

46. Número de colaboradores em 31/12/2014 *

47. Número de colaboradores em 31/12/2015 *

48. Número de colaboradores em 31/12/2016 *

49. Número de colaboradores previsto para 31/12/2017 *

52. 30. Como você avalia sua vida hoje? *

Marcar apenas uma oval.

	0	1	2	3	4	5	6	7	8	9	10	
Pior vida	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Melhor vida

53. 31. Como você acha que sua vida estará daqui a 5 anos? *

Marcar apenas uma oval.

	0	1	2	3	4	5	6	7	8	9	10	
Pior vida	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Melhor vida

54. 32. Você se considera uma pessoa empreendedora? *

Marcar apenas uma oval.

	0	1	2	3	4	5	6	7	8	9	10	
Pouco empreendedora	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Muito empreendedora

55. 33. De modo geral, você diria que começou seu negócio próprio porque: *

Marcar apenas uma oval.

- ☐ Achou que tinha uma oportunidade
- ☐ Teve necessidade
- ☐ Ambas as alternativas

56. **34. Porque você prefere ter um negócio próprio a ser um empregado? (favor marcar todas as alternativas)**

Marque todas que se aplicam.

- ☐ Para evitar incertezas do mercado de trabalho
- ☐ Melhor perspectiva de renda futura
- ☐ Independência pessoal ou auto realização
- ☐ Falta de empregos atrativos no mercado
- ☐ Oportunidade de negócios
- ☐ Para contribuir para a sociedade
- ☐ Independência para escolher local e horário de trabalho
- ☐ Meus amigos e minha família têm seus próprios negócios

57. **35. Sua empresa gostaria de receber os resultados desta pesquisa? ***

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não

58. **36. Espaço destinado para comentários ou sugestões do respondente (opcional)**

Pesquisa completada. Agradecemos muito sua colaboração!
Clique em "Enviar"