

UNIVERSIDADE DO GRANDE RIO “PROFESSOR JOSE DE SOUZA
HERDY”

ESCOLA DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
CURSO DE MESTRADO EM ADMINISTRAÇÃO

DISSERTAÇÃO DE MESTRADO

ANÁLISE DO *SCORE* DE CRÉDITO NO MERCADO
BRASILEIRO À LUZ DA *BIG DATA* E DA NOVA LGPD

ELISA MOUTTA PINHO

RIO DE JANEIRO

2021

Elisa Moutta Pinho


“Análise do Score de Crédito no Mercado Brasileiro à Luz da Big Data e da Nova LGPD”

Tese apresentada à Universidade do Grande Rio “Prof. José de Souza Herdy”, como parte dos requisitos parciais para obtenção do grau de Mestre em Administração.

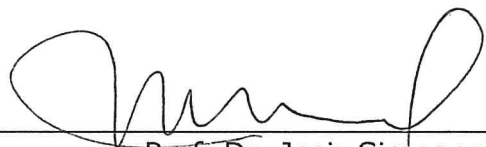
Área de Concentração:
Gestão Organizacional.

Aprovada em 29 de Setembro de 2021.

Banca Examinadora



Prof. Dr. Paulo Vitor Jordão da Gama Silva
Universidade do Grande Rio - UNIGRANRIO



Prof. Dr. Josir Simeone Gomes
Universidade do Grande Rio - UNIGRANRIO



Prof. Dr. Marcelo Cabul Klotzle
Pontifícia Universidade Católica – PUC/RJ

ELISA MOUTTA PINHO

ANÁLISE DO *SCORE* DE CRÉDITO NO MERCADO
BRASILEIRO À LUZ DA *BIG DATA* E DA NOVA LGPD

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Administração, Curso de Mestrado em Administração, Programa de Mestrado e Doutorado em Administração, Universidade do Grande Rio.

Orientador: Prof. Dr. Paulo Vitor Jordão da Gama Silva

RIO DE JANEIRO

2021

CATALOGAÇÃO NA FONTE
UNIGRANRIO – NÚCLEO DE COORDENAÇÃO DE BIBLIOTECAS

P654a

Pinho, Elisa Moutta.

Análise do *score* de crédito no mercado brasileiro à luz da *Big Data* e da nova LGPD / Elisa Moutta Pinho. – Rio de Janeiro, 2021.
205 f. : il. ; 30 cm.

Dissertação (Mestrado em Administração) – Universidade do Grande Rio “Prof. José de Souza Herdy”, Escola de Ciências Sociais Aplicadas, 2021.

“Orientador: Prof. Dr. Paulo Vitor Jordão da Gama Silva”.

Referências: f. 138-156.

1. Administração. 2. Empresas de crédito. 3. *Score* de crédito. 4. *Big Data*. 5. Brasil. [Lei geral de proteção de dados pessoais (2018)]. I. Silva, Paulo Vitor Jordão da Gama. II. Universidade do Grande Rio “Prof. José de Souza Herdy”. III. Título.

CDD – 658

AGRADECIMENTOS

Inicialmente, agradeço a Deus pela vida, pela saúde e por todos os sonhos nos quais venho alcançando e a minha mãe Bianca, verdadeiramente a maior mestra da minha vida e que sempre acreditou em mim, estando ao meu lado nas horas mais difíceis e felizes da minha vida.

A minha filha Cecília, pelo nosso amor, que transborda no peito, sua alegria de criança, trazendo leveza aos meus dias e o meu marido Marcelo, companheiro, amigo e amor, por estar sempre comigo, me incentivando e vibrando com minhas conquistas. Sua ajuda foi de muita valia para mim, me mostrando que o amor existe e mora na paciência e nos detalhes.

A Universidade Unigranrio, seu corpo docente, direção e administração que oportunizaram a janela que hoje vislumbro um horizonte superior, elevado pela acendrada confiança no mérito e ética aqui presentes, bem como à Coordenação de Aperfeiçoamento de Pessoa de Nível Superior (CAPES) pela bolsa, que me permitiu uma formação acadêmica diferenciada e de qualidade.

Ao meu orientador, Professor Paulo Jordão, pelo acolhimento, dedicação, respeito pela construção do conhecimento, pela paciência, por ter lido inúmeras vezes os mesmos capítulos, por seus conselhos, críticas e sugestões bem como a permanente valorização do trabalho desenvolvido e seu entusiasmo determinantes para o final alcançado. Para mim, é uma honra tê-lo como orientador de pesquisa, nesse mestrado. Muito obrigada! Para minha Coordenadora, Professora Rejane Prevot, pelas suas observações criteriosas, que resultaram em um aperfeiçoamento formidável.

Agradeço à Flora Bittencourt, que durante essa jornada do mestrado, tornou-se uma amiga e que deixou a vida mais leve e divertida, bem como aos demais membros da banca, que dedicaram seu tempo em me auxiliar na construção deste trabalho e a todos que de alguma forma contribuíram e me incentivaram pela conclusão deste trabalho.

Esses anos de pesquisa foram marcados por desafios, construções e amadurecimento. Nesse período, aprendi que uma dissertação ou qualquer outro trabalho é a extensão da vida do autor. Dessa forma, para que algo de valor seja produzido, a pessoa deve primeiro criar algo de valor em si mesma.

*“A educação é a arma mais poderosa que você
pode usar para mudar o mundo.”*

Nelson Mandela

Resumo

A sociedade está gerando e consumindo uma gigantesca quantidade de dados, e as empresas que utilizam análises de *Big Data* devem atentar aos desafios impostos por essa realidade. A coleta de dados pessoais permitiu às empresas privadas e às entidades públicas utilizá-los em uma escala sem precedentes. Nesse contexto, o presente estudo propôs realizar uma análise sobre como o sistema de score de crédito pode ser impactado pelo uso da *Big Data* e pela nova LGPD, bem como sua influência em relação aos tomadores de crédito e às empresas que operam com crédito. A metodologia deste estudo é de caráter quantitativa, com análises econométricas por meio de regressões logísticas e testes qui-quadrado de independência, utilizando o software R. Em relação a análise por meio do ponto de vista das empresas, foram utilizados quatro modelos que buscam analisar: as relações no uso do *score* de crédito; o custo/benefício em se usar o *score* de crédito; a prática da LGPD no uso do *score* de crédito; e as relações entre *Score* de Crédito, *Big Data*, LGPD e empréstimos/vendas. Em relação aos tomadores de crédito, foram utilizados dois modelos, que buscam entender as relações, tanto do endividamento quanto da tomada de crédito. Com o intuito de otimizar a avaliação das hipóteses formuladas por meio dos modelos propostos, foi aplicado o princípio da parcimônia, que, em linhas gerais, presume que se deve optar pelo modelo mais simples, caso este consiga explicar o evento. O estudo contou com a participação de 411 respondentes, dos quais 283 assinalaram possuir alguma dívida. Apenas um modelo proposto, dos respondentes que possuem apenas dívidas da empresa, apresentou significância estatística. Aplicando o princípio da simplificação, dois modelos para esse mesmo grupo apresentaram significância estatística. Para aqueles que possuem apenas dívidas pessoais, somente um modelo simplificado foi efetivo em apresentar resultados estatisticamente significativos. Para os respondentes que possuem dívidas da empresa e dívidas pessoais, nem os modelos originais propostos nem os modelos simplificados apresentaram resultados significativos. Foi verificado que a utilização da *Big Data* e da LGPD impacta de maneira significativa o uso do *score* de crédito. Observou-se que a falta de conhecimentos sobre o cálculo do *score* de crédito exerce influência na formação do endividamento.

Palavras-chave:

Big Data, Empresas de crédito, LGPD, *Score* de Crédito, Tomador de crédito.

Abstract

Society is generating and consuming a huge amount of data, and companies that use Big Data analytics must be aware of the challenges imposed by this reality. The collection of personal data has allowed private companies and public entities to use them on an unprecedented scale. In this context, this study proposed to carry out an analysis of how the credit scoring system can be impacted by the use of Big Data and the new LGPD, as well as its influence on borrowers and companies that operate with credit. The methodology of this study is quantitative, with econometric analysis through logistic regressions and chi-square tests of independence, using the R software. Regarding the analysis through the point of view of companies, four models were used that seek to analyze: the relationships in the use of the credit score; the cost/benefit of using the credit score; LGPD's practice in the use of credit scores; and the relationships between Credit Score, Big Data, LGPD and loans/sales. In relation to borrowers, two models were used, which seek to understand the relationships between indebtedness and borrowing. In order to optimize the evaluation of the hypotheses formulated through the proposed models, the principle of parsimony was applied, which, in general terms, assumes that the simplest model should be chosen, if it is able to explain the event. The study had the participation of 411 respondents, of which 283 reported having some debt. Only one proposed model, of respondents who only have debts of the company, showed statistical significance. Applying the simplification principle, two models for this same group showed statistical significance. For those who only have personal debts, only a simplified model was effective in presenting statistically significant results. For respondents who have company debt and personal debt, neither the original models proposed nor the simplified models showed significant results. It was found that the use of Big Data and LGPD significantly impacts the use of the credit score. It was observed that the lack of knowledge about credit score calculation influences the formation of indebtedness.

Keywords:

Big Data, Credit Companies, LGPD, Credit Score, Credit Borrower.

Sumário

1	INTRODUÇÃO.....	20
1.1	OBJETIVO DA PESQUISA	23
1.2	JUSTIFICATIVA E RELEVÂNCIA DO ESTUDO.....	24
1.3	ESTRUTURA DA DISSERTAÇÃO	25
2	QUADRO TEÓRICO DE REFERÊNCIA.....	26
2.1	<i>SCORES</i> DE CRÉDITO	26
2.1.1	Principais modelos de <i>score</i> de crédito	38
2.1.1.1	EUA	38
2.1.1.2	China.....	39
2.1.1.3	Tailândia	39
2.1.1.4	Alemanha.....	40
2.1.1.5	Itália	41
2.1.1.6	Espanha.....	41
2.1.1.7	Portugal.....	43
2.1.2	Modelos de <i>score</i> de crédito no Brasil	43
2.1.3	Análise do perfil do tomador de crédito	48
2.1.3.1	Panoramas do endividamento de risco	48
2.1.3.2	Perfil dos tomadores de crédito por grupo de risco	49
2.1.3.3	Relação entre os indicadores de risco	51
2.2	<i>BIG DATA</i>	53
2.2.1	<i>Big Data</i> como mecanismo de mudança no <i>score</i> de crédito.....	58
2.3	LGPD.....	61
2.3.1	<i>Score</i> de crédito à luz da nova LGPD	63
2.4	ESTUDOS ANTERIORES RELACIONADOS	65
2.4.1	<i>Score</i> de Crédito <i>versus Big Data</i>	65
2.4.2	<i>Score</i> de crédito <i>versus</i> LGPD	66
2.4.3	LGPD <i>versus Big Data</i>	67
2.4.4	<i>Score</i> de crédito <i>versus Big Data versus</i> LGPD.....	68
3	METODOLOGIA E BASE DE DADOS	69
3.1	REGRESSÃO LOGÍSTICA	74
3.2	TESTES DE ROBUSTEZ	76
3.2.1	Teste de Correlação	76
3.2.1.1	Teste de Qui-Quadrado de Independência.....	76
3.2.1.2	Teste Exato de Fisher	77
3.2.2	Teste de Multicolinearidade	77
3.2.3	Análise de variância ANOVA	77
3.2.4	Causalidade de Granger.....	77
4	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....	78

4.1	ANÁLISES SOCIODEMOGRÁFICAS	78
4.1.1	Todos os Respondentes da Amostra.....	78
4.1.2	Respondentes que Não Possuem Dívidas.....	80
4.1.3	Respondentes que Possuem Dívidas	82
4.1.3.1	Respondentes que possuem apenas dívidas pessoais	83
4.1.3.2	Respondentes que possuem apenas dívidas da empresa.....	83
4.1.3.3	Respondentes que possuem dívidas pessoais e dívidas da empresa	85
4.2	ANÁLISES ECONOMÉTRICAS.....	87
4.2.1	Somente Dívidas da Empresa.....	87
4.2.1.1	Modelo 1: Avaliação do uso de score de crédito sob o ponto de vista da empresa	87
4.2.1.2	Modelo 2: Custo-benefício observado pela empresa no uso do score de crédito	91
4.2.1.3	Modelo 3: Avaliação da prática da LGPD sob o ponto de vista da empresa	96
4.2.1.4	Modelo 4: Relação entre uso de score de crédito, Big Data e LGPD, sob o ponto de vista da empresa.....	101
4.2.2	Somente Dívidas Pessoais	104
4.2.2.1	Modelo 5: Avaliação do endividamento sob a ótica do tomador de crédito	104
4.2.2.2	Modelo 6: Influência na formação do endividamento sob a ótica do tomador de crédito	107
4.2.3	Dívidas da Empresa e Dívidas Pessoais	113
4.2.3.1	Modelo 1: Avaliação do uso de score de crédito sob o ponto de vista da empresa	113
4.2.3.2	Modelo 2: Custo/benefício observado pela empresa no uso do score de crédito	115
4.2.3.3	Modelo 3: Avaliação da prática da LGPD sob o ponto de vista da empresa	118
4.2.3.4	Modelo 4: Influência no uso de score de crédito sob o ponto de vista da empresa	120
4.2.3.5	Modelo 5: Avaliação do endividamento sob a ótica do tomador de crédito	123
4.2.3.6	Modelo 6: Influência do endividamento sob a ótica do tomador de crédito	126
4.3	SUMARIZAÇÃO DOS RESULTADOS.....	129
5	CONCLUSÃO.....	132
	REFERÊNCIAS.....	138
	APÊNDICE A – CÓDIGO-FONTE R (MODELO 1).....	157
	APÊNDICE B – CÓDIGO-FONTE R (MODELO 2).....	161
	APÊNDICE C – CÓDIGO-FONTE R (MODELO 3).....	165
	APÊNDICE D – CÓDIGO-FONTE R (MODELO 4).....	169

APÊNDICE E – CÓDIGO-FONTE R (MODELO 5)	173
APÊNDICE F – CÓDIGO-FONTE R (MODELO 6).....	177
APÊNDICE F – CÓDIGO-FONTE R (QUI-QUADRADO)	182
ANEXO A – QUESTIONÁRIO <i>SCORE</i> DE CRÉDITO, <i>BIG DATA</i> E LGPD	197

Lista de tabelas

Tabela 1: Principais contribuições para o Credit Score.....	27
Tabela 2: 5 C's do Crédito.....	37
Tabela 3: Análise das obrigações assumidas pelo tomador de crédito.....	38
Tabela 4: Perfil socioeconômico dos endividados de risco.....	50
Tabela 5: Comprometimento de renda por subpopulações.	52
Tabela 6: Características da Big Data	56
Tabela 7: Principais vantagens da Pesquisa online	70
Tabela 8: Caracterização de todos os respondentes da amostra	78
Tabela 9: Quantitativo de respondentes com dívidas e sem dívidas	79
Tabela 10: Caracterização dos respondentes que não possuem dívidas.....	80
Tabela 11: Caracterização dos respondentes que possuem dívidas.....	81
Tabela 12: Classificação das dívidas dos respondentes da amostra	83
Tabela 13: Informações sobre tomada de crédito e endividamento dos respondentes que possuem apenas dívidas pessoais	83
Tabela 14: Informações sobre tomada de crédito e endividamento dos respondentes que possuem apenas dívidas da empresa	84
Tabela 15: Elementos que podem influenciar positivamente o uso do score de crédito para os respondentes que possuem apenas dívidas da empresa.....	85
Tabela 16: Fatores influenciados positivamente pelo uso das boas práticas da LGPD para os respondentes que possuem apenas dívidas da empresa.....	85
Tabela 17: Informações sobre tomada de crédito e endividamento dos respondentes que possuem dívidas pessoais e dívidas da empresa	86
Tabela 18: Elementos que podem influenciar positivamente o uso do score de crédito para os respondentes que possuem dívidas pessoais e dívidas da empresa	87
Tabela 19: Fatores influenciados positivamente pelo uso das boas práticas da LGPD para os respondentes que possuem dívidas pessoais e dívidas da empresa	87
Tabela 20: Elementos que podem influenciar negativamente o uso do score de crédito para os respondentes que possuem somente dívidas da empresa	88
Tabela 21: Coeficientes do modelo 1 referentes aos respondentes que possuem apenas dívidas da empresa	88
Tabela 22: Tabelas de contingência para os dados utilizados no modelo 1 para os respondentes que possuem somente dívidas da empresa.....	89
Tabela 23: Teste de qui-quadrado aplicado às variáveis do modelo 1 para os respondentes que possuem somente dívidas da empresa.....	89

Tabela 24: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 1 reduzido para os respondentes que possuem somente dívidas da empresa	90
Tabela 25: Razão de chances e intervalo de confiança das regressões logísticas simples das preditoras candidatas ao modelo 1 reduzido para os respondentes que possuem somente dívidas da empresa	90
Tabela 26: Coeficientes da regressão logística múltipla do modelo 1 reduzido para os respondentes que possuem somente dívidas da empresa	91
Tabela 27: Custo-benefício em relação ao uso do score de crédito	92
Tabela 28: Coeficientes do modelo 2 referentes aos respondentes que possuem apenas dívidas da empresa	92
Tabela 29: Razão de chances e intervalo de confiança da regressão logística do modelo 2 para os respondentes que possuem somente dívidas da empresa	93
Tabela 30: Tabelas de contingência para os dados utilizados no modelo 2 para os respondentes que possuem somente dívidas da empresa	93
Tabela 31: Teste de qui-quadrado aplicado às variáveis do modelo 2 para os respondentes que possuem somente dívidas da empresa	94
Tabela 32: Coeficientes da regressão logística do modelo 2 reduzido para os respondentes que possuem somente dívidas da empresa	94
Tabela 33: Razão de chances e intervalo de confiança da regressão logística do modelo 2 reduzido para os respondentes que possuem somente dívidas da empresa	95
Tabela 34: Comparativo entre o modelo 2 e o modelo 2 reduzido para os respondentes que possuem somente dívidas da empresa	95
Tabela 35: Fatores que inibem o uso das boas práticas da LGPD para os respondentes que possuem somente dívidas da empresa	96
Tabela 36: Coeficientes do modelo 3 referentes aos respondentes que possuem apenas dívidas da empresa	97
Tabela 37: Tabelas de contingência para os dados utilizados no modelo 3 para os respondentes que possuem somente dívidas da empresa	97
Tabela 38: Teste de qui-quadrado aplicado às variáveis do modelo 3 para os respondentes que possuem somente dívidas da empresa	98
Tabela 39: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa	98
Tabela 40: Razão de chances e intervalo de confiança da regressão logística simples da preditora candidata, estatisticamente significativa, ao modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa	99
Tabela 41: Coeficientes da regressão logística múltipla do modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa	99

Tabela 42: Razão de chances e intervalo de confiança da regressão logística do modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa	100
Tabela 43: Comparativo entre o modelo 3 e o modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa	100
Tabela 44: Informações sobre Big Data e LGPD e suas influências no uso do score de crédito para os respondentes que possuem somente dívidas da empresa	101
Tabela 45: Coeficientes do modelo 4 referentes aos respondentes que possuem apenas dívidas da empresa	102
Tabela 46: Tabelas de contingência para os dados utilizados no modelo 4 para os respondentes que possuem somente dívidas da empresa	102
Tabela 47: Teste de qui-quadrado aplicado às variáveis do modelo 4 para os respondentes que possuem somente dívidas da empresa	103
Tabela 48: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem somente dívidas da empresa	103
Tabela 49: Razão de chances e intervalo de confiança das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem somente dívidas da empresa	104
Tabela 50: Elementos que podem influenciar no aumento das dívidas para os respondentes que possuem somente dívidas pessoais	105
Tabela 51: Coeficientes do modelo 5 referentes aos respondentes que possuem apenas dívidas pessoais	105
Tabela 52: Tabelas de contingência para os dados utilizados no modelo 5 para os respondentes que possuem somente dívidas pessoais	106
Tabela 53: Teste de qui-quadrado aplicado às variáveis do modelo 5 para os respondentes que possuem somente dívidas pessoais	106
Tabela 54: Coeficientes da regressão logística simples da preditora candidata ao modelo 5 reduzido para os respondentes que possuem somente dívidas pessoais	107
Tabela 55: Fatores que podem influenciar na formação do endividamento para os respondentes que possuem somente dívidas pessoais	108
Tabela 56: Coeficientes do modelo 6 referentes aos respondentes que possuem apenas dívidas pessoais	108
Tabela 57: Tabelas de contingência para os dados utilizados no modelo 6 para os respondentes que possuem somente dívidas pessoais	109
Tabela 58: Teste de qui-quadrado aplicado às variáveis do modelo 6 para os respondentes que possuem somente dívidas pessoais	109
Tabela 59: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais	110

Tabela 60: Razão de chances e intervalo de confiança das regressões logísticas simples das predictoras candidatas ao modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais.....	111
Tabela 61: Coeficientes da regressão logística múltipla do modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais.....	111
Tabela 62: Razão de chances e intervalo de confiança da regressão logística do modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais.....	112
Tabela 63: Comparativo entre o modelo 6 e o modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais.....	112
Tabela 64: Elementos que podem influenciar negativamente o uso do score de crédito para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	113
Tabela 65: Coeficientes do modelo 1 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais.....	113
Tabela 66: Tabelas de contingência para os dados utilizados no modelo 1 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	114
Tabela 67: Teste de qui-quadrado aplicado às variáveis do modelo 1 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	114
Tabela 68: Coeficientes das regressões logísticas simples das predictoras candidatas ao modelo 1 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	115
Tabela 69: Custo-benefício em relação ao uso do score de crédito.....	116
Tabela 70: Coeficientes do modelo 2 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais.....	116
Tabela 71: Tabelas de contingência para os dados utilizados no modelo 2 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	117
Tabela 72: Teste de qui-quadrado aplicado às variáveis do modelo 2 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	117
Tabela 73: Fatores que inibem o uso das boas práticas da LGPD para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	118
Tabela 74: Coeficientes do modelo 3 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais.....	119
Tabela 75: Tabelas de contingência para os dados utilizados no modelo 3 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	119
Tabela 76: Teste de qui-quadrado aplicado às variáveis do modelo 3 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	120
Tabela 77: Informações sobre Big Data e LGPD e suas influências no uso do score de crédito para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	121
Tabela 78: Coeficientes do modelo 4 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais.....	121

Tabela 79: Tabelas de contingência para os dados utilizados no modelo 4 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	121
Tabela 80: Teste de qui-quadrado aplicado às variáveis do modelo 4 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	122
Tabela 81: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	122
Tabela 82: Razão de chances e intervalo de confiança das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	123
Tabela 83: Elementos que podem influenciar no aumento das dívidas para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	124
Tabela 84: Coeficientes do modelo 5 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais	124
Tabela 85: Tabelas de contingência para os dados utilizados no modelo 5 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	125
Tabela 86: Teste de qui-quadrado aplicado às variáveis do modelo 5 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	125
Tabela 87: Coeficientes da regressão logística simples da preditora candidata ao modelo 5 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	126
Tabela 88: Fatores que podem influenciar na formação do endividamento para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	127
Tabela 89: Coeficientes do modelo 6 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais	127
Tabela 90: Tabelas de contingência para os dados utilizados no modelo 6 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	128
Tabela 91: Teste de qui-quadrado aplicado às variáveis do modelo 6 para os respondentes que possuem dívidas da empresa e dívidas pessoais.....	129
Tabela 92: Sumarização dos resultados.....	130

Lista de figuras

Figura 1: Organização dos capítulos da dissertação.....	25
Figura 2: Sistema Credit Scoring	35
Figura 3: Estrutura do sistema de Rating	36
Figura 4: Plataforma de consulta ao score de crédito.....	46
Figura 5: Fórmula da Big Data.....	55
Figura 6: Curva <i>Sigmoid</i> - Regressão Logística.....	75

Lista de gráficos

Gráfico 1: Número de tomadores de crédito (endividados de risco).....	49
Gráfico 2: Endividados por Unidade da Federação.....	51
Gráfico 3: Endividamento das famílias	80

ABREVIATURAS

Sigla	Significado
ACSP	Associação Comercial de São Paulo
AIC	<i>Akaike Information Criterion</i>
Anbima	Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais
Bacen	Banco Central
BIC	<i>Bayesian Information Criterion</i>
CDC	Código de Defesa do Consumidor
CDL	Câmara Dirigente de Lojistas
CMN	Conselho Monetário Nacional
CPF	Cadastro de Pessoa Física
CRC	Central de Responsabilidade de Crédito
DPO	<i>Data Protection Officer</i>
EMS	<i>Emerging Market Scoring</i>
FEBRABAN	Federação Brasileira de Bancos
FMI	Fundo Monetário Internacional
G10	Grupo dos dez países
GDPR	<i>General Data Protection Regulation</i>
IOT	<i>Internet of things</i> – Internet das coisas
LGPD	Lei Geral de Proteção de Dados
PNAD	Pesquisa Nacional por Amostra de Domicílios
OR	<i>Odds Ratio</i> – Razão de Chances
SCHUFA	Schutzgemeinschaft für allgemeine Kreditsicherung (<i>Score</i> na Alemanha)
TIC	Tecnologia da Informação e Comunicação

1 INTRODUÇÃO

De acordo com Simon (2013), não existem dúvidas que a sociedade está gerando e consumindo uma grande quantidade dos mais diferentes tipos de dados. Neste contexto, o grande desafio das organizações não está em receber dados para analisar, e sim em obter informações relevantes que auxiliem a tomada de decisões em meio ao enorme volume de dados disponíveis. Oliveira (2002), comenta que a informação apoia o processo decisório, sendo a base da tomada de decisões.

Cabrera-Sánchez e Villarejo-Ramos (2019) nos dizem que, através dessa enorme transação de dados, abrangendo localização, *internet* das coisas¹, redes sociais, *wearables*² e quaisquer transações *on-line*, as empresas puderam transformar esses dados sobre seus clientes em informações em tempo real e ganharam uma vantagem competitiva substancial mediante exposto por Sivarajah *et al.* (2017).

Analisado por Akerkar (2014), as empresas que utilizam análises de *Big Data* devem estar atentas aos desafios oriundos do ciclo de vida dos dados: incertezas sobre os próprios dados, dificuldades em seu processamento e preocupações sobre a sua gestão. As incertezas circulam em torno de seu volume, veracidade, velocidade, variedade, valor, volatilidade e visualização. O processamento de dados inclui técnicas relacionadas a obtenção, armazenamento, limpeza e transformação dos dados existentes, selecionando o modelo correto de apresentação de resultados. Arruda (2019) comenta que nessas análises, as empresas certamente irão se deparar com dados pessoais, impelindo a aplicação de princípios básicos e legais com relação à privacidade.

Cunha (2018), em seu estudo, identificou novos desafios com relação a proteção de dados pessoais, no atual contexto globalizado. O aumento, exponencial, do compartilhamento e da coleta desses dados permitiram às empresas privadas e às entidades públicas a sua

¹ (IoT) Revolução tecnológica que tem como objetivo conectar os itens usados do dia a dia à rede mundial de computadores.

² São “dispositivos vestíveis”, tecnologias que se apresentam na forma de dispositivos iguais ou similares a peças de roupa ou equipamentos vestíveis, tais como relógios, pulseiras ou até mesmo óculos de realidade virtual.

utilização, em uma escala sem precedentes, no exercício das suas atividades. Os indivíduos cedem cada vez mais as suas informações pessoais de uma forma pública e global e, juntamente com as várias inovações tecnológicas, acabam por transformar a economia e a vida social, contribuindo para promover a livre circulação desses dados pessoais e a sua transferência para organizações nacionais e internacionais (REGULAMENTO GERAL DE PROTEÇÃO DE DADOS, 2016).

Estes desafios demandam um quadro consistente de proteção de dados, assistidos por uma aplicação rigorosa de regras de utilização. Por meio desse conjunto de medidas, os indivíduos poderão controlar a utilização do que é feito com seus dados pessoais, reforçando a segurança jurídica e a confiabilidade em empresas privadas, autoridades públicas e operadores econômicos, aumentando a credibilidade necessária ao desenvolvimento da economia digital no complexo mercado interno.

Recentemente, no Brasil, entrou em vigor a LGPD (LEI GERAL DE PROTEÇÃO DE DADOS – Nº 13.709/2018), fortemente influenciada pelo Regulamento Geral de Proteção de Dados da União Europeia³, sendo uma lei que regula as atividades de tratamentos de dados, alterando também os artigos 7º e 16 do Marco Civil da Internet (LEI Nº 12.965/2014). A legislação tem por fundamentação múltiplos fatores, como o respeito à privacidade; à autodeterminação informativa; à liberdade de expressão, de informação; à inviolabilidade da intimidade, da honra e da imagem entre outros, criando um conjunto de novos conceitos como, por exemplo, dados pessoais e dados sensíveis, que estabelecem as condições em que os dados pessoais deverão ser tratados. Ao mesmo tempo, determina um conjunto de direitos para os titulares desses dados, gerando obrigações singulares para os que, porventura, venham a ser seus controladores, produzindo procedimentos e normas para que haja os devidos cuidados com o tratamento e o compartilhamento com terceiros.

Nesse contexto, a Lei Geral de Proteção de Dados - LGPD, terá enorme repercussão na forma com que empresas tratam e compartilham as informações de seus clientes, pois foi elaborada com propósitos bem definidos: instituir limites, fixar princípios e estabelecer parâmetros de conformidade para que o relacionamento entre as empresas e os clientes seja o mais transparente e seguro possível. Conforme expressado por Rapôso (2019), as empresas

³ *General Data Protection Regulation*; GDPR ou RGPD – conjunto de normas de proteção à privacidade criado pela União Europeia (UE) desde o dia 25 de maio de 2018.

devem adaptar toda a cadeia de produção, adotando princípios éticos e estabelecendo treinamento dos colaboradores, alcançando uma relação de transparência em relação a todo o trabalho.

De Carvalho e Vilhena (2019) descrevem em seu trabalho que as organizações, de um modo geral, precisam regulamentar condições, regimes e procedimentos internos para o tratamento de dados pessoais, procedendo, periodicamente, auditorias internas para avaliar o grau de conformidade com a LGPD, estipulando quais os riscos no tratamento de dados. Essas organizações necessitam gerir com frequência os dados pessoais, relatando quais são esses dados, onde estão e para que estão sendo empregues. Ao devolver a propriedade dos dados aos clientes, a LGPD afetará, diretamente, como os dados serão analisados, promovendo uma revisão nos custos empresariais e impactando na comunicação digital das organizações, culminando numa reformulação das relações entre as empresas e os clientes.

Conforme observado por Tole (2013), a análise de dados no contexto organizacional vem atravessando grandes mudanças. No passado, um programa de relacionamento e fidelidade era necessário para identificar o perfil do consumidor através do seu histórico de compras; atualmente, o perfil deste consumidor pode ser previsto por uma empresa mesmo que nenhuma compra tenha sido realizada por ele, devido ao que este divulga *on-line*, voluntária ou involuntariamente, sobre as suas experiências de compra e de comportamento.

Dada a grande quantidade de informações pessoais que circulam no mundo digital e as contínuas e rápidas mudanças tecnológicas pelas quais as sociedades modernas atravessam, as empresas acabaram por se submeter ao uso dos dados, como moeda de troca. Lançando mão desses dados, conseguem angariar informações do histórico financeiro e, com isso, classificá-lo, imputando rótulos, como “mau ou bom” pagador, conforme aponta Khashman (2011). Os “maus pagadores” com *scores* baixos, teriam dificuldades na hora da concessão de crédito, e aqueles que são ditos como “bons pagadores” alçariam maiores *scores* obtendo maior credibilidade na hora de buscar por recursos financeiros. Todo o histórico fica registrado em um sistema de classificação de crédito, conhecido como Cadastro Positivo⁴.

⁴ Reúne informações sobre como têm sido pagos os compromissos relacionados à contratação de crédito – empréstimos, financiamentos e crediários, por exemplo. Passam a constar no histórico do CPF os totais financiados, quantidades e valores das parcelas, bem como o comportamento e a pontualidade de pagamento demonstrados pelo consumidor.

De acordo com a Associação Comercial de São Paulo (2020), sem esse cadastro, as instituições financeiras acabavam cobrando taxas elevadas mesmo dos bons pagadores. De acordo com essas instituições, apenas com as informações restritivas, que indicam se uma pessoa ou uma empresa está negativada ou não, os bancos, as *fintechs*⁵ e as cooperativas de crédito não tinham dados suficientes para saber quem estaria pagando suas contas em dia.

Mesmo parecendo um cadastro inofensivo, feito para auxiliar a redução do risco para as empresas que disponibilizam crédito, até que ponto o *credit score*, por meio de seus algoritmos⁶ de tratamento de dados, rotulando os tomadores de crédito, estará apto a aplicar o direito de revisão dessas decisões automatizadas? Esse cadastro, cuja base de dados é objeto de tratamento, poderia provocar dificuldades para a obtenção de crédito para aqueles que, porventura, já estiveram classificados como “maus pagadores”? O *credit score* realmente teria transparência frente ao titular dos dados?

1.1 Objetivo da Pesquisa

Visando encontrar respostas acerca das perguntas supracitadas, ao final da seção anterior, **o presente estudo propõe realizar uma análise sobre como o sistema de *score* de crédito pode ser impactado pelo uso de *big data* e pela nova LGPD, bem como sua influência em relação as empresas que operam com crédito e com os tomadores de crédito.**

Dado o objetivo geral, esta pesquisa se desdobrará nos seguintes objetivos específicos:

- a) Analisar os principais modelos de *score* de crédito no Brasil e no mundo;
- b) Verificar a influência da *Big Data* e da LGPD no sistema de *score* de crédito;
- c) Avaliar, em qual medida, os dados objetos de tratamento podem influenciar na obtenção de crédito, tendo como base informações claras a respeito dos critérios e procedimentos adotados;
- d) Verificar se o *score* de crédito atende as condições necessárias de transparência e objetividade na utilização de dados pessoais;
- e) Entender o ponto de vista das empresas e dos tomadores de crédito em relação as modificações envolvendo a *Big Data* e o advindo da LGPD em relação aos modelos de *score* de crédito.

⁵ *Startups* (empresas emergentes), que trabalham para inovar e otimizar serviços do sistema financeiro.

⁶ Sequência finita de ações executáveis que visam obter uma solução para um determinado tipo de problema.

1.2 Justificativa e Relevância do Estudo

O presente estudo se justifica dada à importância que o *score* de crédito alcançou, no mercado nacional e internacional, assim como notado nos estudos de Wassler (2019), sobre o desenvolvimento contínuo da economia de mercado através do aumento da confiabilidade e transparência de devedores e credores, além de sua importância para o processo decisório que envolve análises de riscos acerca da concessão de crédito. Visto que as empresas que trabalham no mercado de crédito possuem um poder de influência que podem impactar a vida financeira de pessoas e de outras empresas, este trabalho contribui no sentido de esclarecer como é a percepção do tomador de crédito e das próprias empresas em relação aos métodos estatísticos empregados pelas organizações responsáveis pelo *score* de crédito brasileiro que se utilizam de análises automatizadas do histórico financeiro, com auxílio de *Big data* e atualmente devem se enquadrar na nova LGPD.

A importância do estudo da *Big Data* deve-se ao número crescente do interesse na temática sobre grandes volumes de dados, nas quais as ferramentas tradicionais ainda não são capazes de lidar na velocidade requerida, marcado por suas características como a velocidade de processamento necessárias e a sua complexidade. Já o estudo da LGPD, torna-se importante para uma melhor compreensão de sua adaptabilidade com a tecnologia, desmitificando as inseguranças de um todo por se tratar de uma nova lei, bem como será sua aplicabilidade e quais atitudes irão ser tomadas visando o *compliance*⁷ da lei.

Dada a justificativa acima, este trabalho torna-se relevante para: o governo e seus órgãos reguladores pois o mercado de crédito é parte essencial da economia; órgãos de proteção ao consumidor, na medida em que poderá fornecer subsídios teóricos e acadêmicos tanto para embasar as soluções de eventuais conflitos que porventura venham a existir quanto para auxiliar em possíveis propostas de políticas públicas direcionadas à concessão de crédito; para os tomadores de crédito, visto que os mesmos, de posse das informações apresentadas neste estudo, podem ter novas percepções acerca das empresas de *score* crédito, podendo assim, desenvolver uma consciência financeira mais saudável; para as respectivas empresas

⁷ Compromisso da empresa em cumprir normas nacionais e internacionais e agir de acordo com regras ou instruções internas.

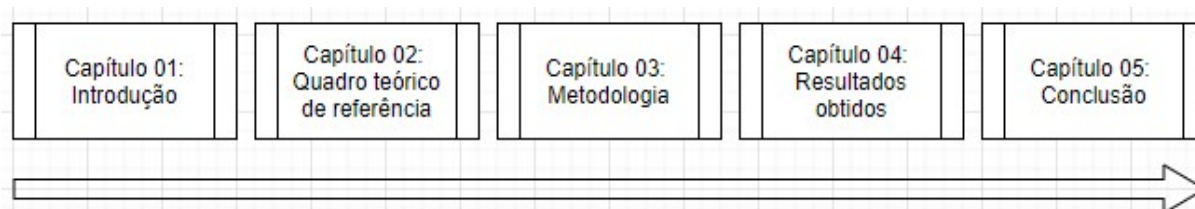
financeiras, nas figuras dos analistas de *big data*⁸ e *data protection officers*⁹, no sentido de aplicar entendimentos sobre questões jurídicas/legislativas e o novo cenário que se apresenta com a vigência da atual LGPD; e para a comunidade científica, delimitando futuros desafios para a área de Administração.

Os questionamentos aqui levantados pretendem despertar uma nova reflexão tanto para os indivíduos tomadores de crédito quanto para os profissionais envolvidos diretamente na manipulação dos dados e para as empresas, levando em conta que tais dados precisam ser resguardados no que concerne ao uso indiscriminado, que pode afetar diretamente a oportunidade financeira como um todo, assim como visto em Silveira (2017).

1.3 Estrutura da Dissertação

O primeiro capítulo procura introduzir conceitos acerca da *Big data*, do LGPD e do *score* de crédito, bem como apresentar os objetivos da pesquisa e a justificativa e relevância do estudo. O segundo capítulo busca apresentar a fundamentação teórica relacionada ao tema proposto para a pesquisa. No terceiro capítulo é apresentada a metodologia utilizada para a condução do presente estudo. O quarto capítulo apresenta os resultados obtidos com a pesquisa e discussões de tais resultados com trabalhos que apresentem similaridade com a temática analisada. No quinto capítulo são apresentadas as considerações finais, os impactos que os sistemas de classificação e *score* de crédito podem gerar na sociedade, bem como sugestões para futuras pesquisas com o intuito de estimular novos trabalhos – a ordenação apontada pode ser refletida por meio da figura 1.

Figura 1: Organização dos capítulos da dissertação.



Fonte: Desenvolvido pela autora

⁸ Especialista que faz a gestão e a análise de dados, a fim de assegurar eficiência e rentabilidade para a empresa, além de propor ideias inovadoras para o negócio.

⁹ Especialista que cuida das questões referentes à proteção de dados (LGPD), com maior foco na segurança das informações.

2 QUADRO TEÓRICO DE REFERÊNCIA

2.1 Scores de Crédito

Crédito é uma palavra originária do latim *'creditum'* e refere-se a 'confiar em' (HOUAISS, VILLAR & FRANCO, 2004). Conforme dito por Nader (2010), transação de crédito é a circulação de bens, sendo composto pelas seguintes partes: credor, aquele que deposita a confiança; devedor, aquele que se beneficia daquela confiança depositada. Uma transação de crédito reúne dois fatores: o tempo e a confiança. A confiança como um ato calculado, contendo todos os seus riscos (RIVES-LANGE, 1995) e o tempo, conforme nota de Covello (1988):

"Toda a operação de troca na qual se realiza uma prestação pecuniária presente contra uma prestação futura de igual natureza é a troca de uma riqueza presente por uma riqueza futura. O que caracteriza o crédito, pois, é disposição efetiva e imediata de um bem econômico em vista de uma contraprestação futura." (COVELLO, 1988)

Sob o ponto de vista de Jiménez e Saurina (2006), o acesso ao financiamento ou crédito criam certos compromissos, isto é, os mutuários pagam em relação aos termos acordados e pela criação de confiança, inclusive pagando pelo prêmio de risco em caso de inadimplência. Todo esse processo de adiantamento de crédito criou o fenômeno de risco de crédito e solvência¹⁰ (BESSIS, 2011). O conceito de emprestar, tomar emprestado e os riscos envolvidos, estão relacionados à história humana há muito tempo (THOMAS, CROOK & EDELMAN, 2017).

Chiavenato (2004), afirma que a maior parte da supervivência social depende de uma certa interação social do ser humano, na qual surgem as compras de bens e imóveis, prestações de serviços, entretenimentos, relações sociais e interações com o poder público. Segundo Mucelin (2018), a relação de consumo é evidente e direta, sendo que um indivíduo fornece algo e a outra pessoa compra e/ou aproveita por motivo de necessidade pessoal. E, em grande parte dessas transações, essa outra pessoa não dispõe de grande quantidade monetária para efetuar a aquisição, carecendo de crédito.

Fazendo um breve resumo dos principais trabalhos, de acordo com a tabela 1, nota-se que no decorrer das décadas, os estudos sobre o *score* de crédito foram aperfeiçoando seus

¹⁰ Refere à possibilidade de uma empresa não receber ou receber em atraso o pagamento de uma venda efetuada a um cliente por causa da insolvência/inadimplência. A inadimplência é um fato difícil de ser evitado, mas pode e deve ser controlado e prevenido pela análise de crédito

métodos de avaliação, de forma a automatizar tais análises, ao mesmo tempo em que se alterava o foco de sua aplicação, inicialmente empregando seu uso no fornecimento de crédito para as empresas, passando para o microcrédito, destinando as pessoas físicas. Alguns estudos recentes já caminham no sentido de observar os impactos que os modelos de *score* de crédito têm sobre a sociedade.

Tabela 1: Principais contribuições para o *Credit Score*

Autor e ano do trabalho	Contribuições
Altman (1968)	Estudos sobre a relação entre índices financeiros e a previsão de falência das empresas
Breiman <i>et al.</i> (1984)	Estudos sobre classificações e regressões de dados utilizando árvores de decisões
Trippi e Turban (1992)	Estudos sobre redes neurais como forma de prever e avaliar risco
Altman <i>et al.</i> (1998)	Estudos sobre como um sistema de <i>score</i> auxilia no investimento em títulos corporativos
Malhotra (2002)	Estudos sobre a diferenciação entre bons e maus empréstimos (gestão de risco) usando sistemas <i>neuro-fuzzy</i>
Corrêa e Machado (2004)	Estudos sobre a construção de um modelo de <i>credit scoring</i> em redes neurais para previsão de inadimplência na concessão de microcrédito
Palmuti e Pichiatti (2004)	Estudos sobre a mensuração do risco de crédito por meio de análise estatística multivariada
Anderson (2007)	Estudos sobre a teoria e prática do gerenciamento de risco de crédito de varejo e automação de decisão como ferramenta
Lima e Kimura (2009)	Estudos sobre a aplicação de redes neurais na análise e na concessão de crédito ao consumidor
Da Silva Filho (2009)	Estudos sobre a aplicação de redes neurais na análise e na concessão de crédito ao consumidor
Pereira <i>et al.</i> (2018)	Estudos e reflexões sobre Data Brokers e os impactos da classificação do <i>score</i> de crédito na sociedade
Abdou e Pointon (2011)	Estudos sobre a pontuação de crédito, técnicas estatísticas e critério de avaliação
O'Neil (2016)	Estudos sobre como a big data pode aumentar a desigualdade
Pasquale (2016)	Estudos sobre os algoritmos secretos que controlam dinheiro e informações
Zuliani (2018)	Estudos sobre a judicialização do <i>score</i> de crédito em empréstimos negados

Fonte: Desenvolvido pela autora

Nesse momento, ao objetivar a realização do seu sonho, o indivíduo vai em busca de uma instituição financeira para obtenção de crédito, e a instituição por si, objetiva obter lucro, elaborando mecanismos eficazes para dirimir os riscos de inadimplência do tomador de crédito solicitante. É apresentado por Zuliani (2018), o método de *score* de crédito que estimulou o STJ - Superior Tribunal de Justiça a preparar a súmula 550, que diz:

“A utilização de *score* de crédito, método estatístico de avaliação de risco que não constitui banco de dados, dispensa o consentimento do consumidor, que terá o direito

de solicitar esclarecimentos sobre as informações pessoais valoradas e as fontes dos dados considerados no respectivo cálculo.” (ZULIANI, 2018)

Segundo Goh e Lee (2019), a pontuação de crédito é usada para avaliar e moldar os dados disponíveis e avaliar sua probabilidade de inadimplência. Usualmente, a pontuação é uma medida da capacidade de crédito dos indivíduos, enquanto a probabilidade de inadimplência analisa a estimativa de um indivíduo de deixar de cumprir sua dívida em um determinado período. O *score*, de acordo com os estudos de Anderson (2007), vem como uma ferramenta para ordenar os tomadores de crédito conforme suas características, visando a padronização e a agilidade das análises em situações, na qual a aprovação ou a recusa do crédito é integralmente automatizada, conforme a perspectiva de Brito (2016).

E, no entanto, o que significa *score* de crédito? De acordo com Yang Yingxu (2007), refere-se à aplicação de meio matemático, em sua vertente probabilística, para avaliar e pontuar o indivíduo consumidor que por assim deseja obter o crédito (empréstimo), levando em conta as suas características pessoais e profissionais, resultando em um sistema de pontuação, que recebe também os nomes de “*credit scoring*”¹¹ ou “*behavior scoring*”.

A expressão *score* de crédito surgiu nos Estados Unidos em meados de 1941, originária do estudo de David Durand, intitulado de “*Risk elements in consumer installment financing*” (DURAND, 1941). O referido autor, foi um dos primeiros pesquisadores do *National Bureau of Economic Research [N.Y – EUA]* a apresentar um modelo que outorgava pesos para cada uma das variáveis operando a análise discriminante¹².

O’Neil (2016) enfatiza que nesta era marcada pela *internet* e *big data* deve-se questionar quais serão os impactos que estes modelos de classificação e pontuação de crédito atuarão nos indivíduos tomadores de crédito e em toda uma sociedade nos próximos anos. Em princípio, os algoritmos utilizados pelas empresas de *score* de crédito apresentam vantagens e utilidades para os indivíduos, no entanto, por serem regidos com critérios pouco conhecidos, podem trazer à tona impactos negativos à sociedade, gerando uma desigualdade e tornando-se um tipo de instrumento de controle social (PASQUALE, 2016).

¹¹ Expressão numérica baseada em uma análise de nível dos arquivos de crédito de uma pessoa, para representar a capacidade creditícia de um indivíduo.

¹² Técnica da estatística multivariada utilizada para discriminar e classificar objetos.

O’Neil (2016) critica que nesta sociedade “algorítmica”, os impactos relacionados e causados por estes modelos matemáticos de *score* não podem ser considerados ferramentas neutras. A autora menciona que estes modelos estão orientados pela visão arraigada de quem os criou, ou seja, muitos preconceitos podem ter sido reforçados por critérios subjetivos. Essas aplicações matemáticas empoderam a economia de dados, já que estes algoritmos são segredos poderosos que visam gerar lucros a algumas empresas e estão salvaguardados pela Lei Nº12.414/2011 – Superior Tribunal Federal, que disciplina a formação e consulta a banco de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para formação de histórico de crédito, cujas fórmulas matemáticas e seus modelos estatísticos não precisam ser divulgados. Além disso, não é necessário exigir o prévio e expresse consentimento do indivíduo tomador de crédito avaliado em questão.

Conforme exposto por Cheregati (2008), o Acordo de Basiléia II¹³, teve suas diretrizes adotadas não somente pelos bancários dos países que formam o G10¹⁴, mas também por outras nações, visando estabelecer requisitos analíticos para as avaliações de riscos, baseados nos dados coletados pelos bancos durante o período do ciclo de vida de um empréstimo bancário. Sua temática principal caracteriza-se por introduzir uma determinada estrutura de capital que esteja mais sensível as incertezas, fomentando o incentivo das boas práticas de gerenciamento de risco, na qual inúmeras instituições bancárias veem estudando e implementando alguns modelos próprios para melhor avaliar as análises de risco.

De acordo com ANBIMA¹⁵ (2010), que faz comentários sobre as diretrizes desse acordo estabelecido, as instituições financeiras utilizam as notas ou *scores* de corte com o objetivo de nortear a concessão de crédito, privilegiando aqueles indivíduos tomadores de crédito que apresentam baixo risco, visando ofertar uma maior segurança ao sistema financeiro. O acordo

¹³ O Acordo fixa-se em três pilares: capital (guardar), supervisão (fiscalizar), transparência e disciplina de mercado (divulgação de dados) e 29 princípios básicos sobre contabilidade e supervisão bancária.

¹⁴ Trata-se de um acordo para a obtenção de empréstimos suplementares, que podem ser obtidos caso os recursos monetários estimados pelo FMI estejam abaixo das necessidades reais do país membro (representantes dos bancos centrais da Alemanha, Bélgica, Canadá, Estados Unidos, França, Itália, Japão, Países Baixos, Reino Unido, Suécia e Suíça).

¹⁵ Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais.

exigiu um novo tipo de estrutura normativa, na qual o Bacen¹⁶ e o CMN¹⁷ emitiram um rol de resoluções que nortearam os métodos mensuradores de riscos de crédito, dentre eles: comunicados nº 12.746 de 2004, nº 16.137 de 2007 e nº 19.028/2009, que estabeleceram procedimentos gerais e a cronologia da implantação do Novo Tratado no Brasil (adaptada as condições, peculiaridades e estágio de desenvolvimento do mercado brasileiro) no intuito de guiar o cumprimento das referidas ações.

Zendersky (2005) destaca que o acordo prevê que as instituições financeiras desenvolvam metodologias próprias para o cálculo regulatório. Cortes (2004) complementa que seus pilares se baseiam na exigência de capital mínimo mantido pelos bancos, fazendo frente a eventuais insolvências, constituindo boas práticas para gerenciamento de risco.

Pela visão de Palmuti e Pichiatti (2004), as instituições financeiras utilizam técnicas para viabilizar uma melhor classificação de seus tomadores de crédito, estabelecendo uma prevenção de riscos que possam ser mensurados antecipadamente, fazendo com que as tomadas de decisão sejam mais assertivas. De acordo com Yanaka e Holland (2010), de posse desse conhecimento, as instituições financeiras obtêm um maior poderio em relação à decisão para a solicitação de crédito, tanto para definir o limite de crédito a ser concedido, privilegiando os tomadores de crédito que possuem menores probabilidades de atraso, quanto para conceder percentuais de limites mais altos, em relação às suas rendas financeiras.

O trabalho de pesquisa elaborado por Abdou e Pointon (2011), apontou que a coleta de informações se tornou um problema crítico na construção do modelo de *score* de crédito, visto que as informações coletadas e utilizadas são sexo, idade, estado civil, salário, entre outras. Essa seleção de variáveis se baseia em uma análise estatística e, de acordo com o estudo, nenhum autor estabeleceu algum tipo de razão teórica para a seleção das variáveis do modelo, que dependem do banco de dados autorizado pelas instituições privadas, segundo Pointon (2011).

¹⁶ Banco central é uma entidade independente ou ligada ao Estado cuja função é gerir a política econômica, ou seja, garantir a estabilidade e o poder de compra da moeda de cada país e do sistema financeiro como um todo.

¹⁷ Conselho Monetário Nacional, criado pela Lei nº 4.595/1964, como poder deliberativo máximo do sistema financeiro do Brasil, sendo responsável por expedir normas e diretrizes gerais para seu bom funcionamento.

Cada autor manteve uma particularidade em relação aos dados dos tomadores de crédito, usando-os cada qual a sua maneira, para definir como parâmetros da construção do *score* de crédito. Autores como Hand, Mcconway e Stanghellini, (1997), Lee e Chen (2005), Zekic-Susac, Sarlija e Bensik (2004), Banasik, Crook e Thomas (2003), Huang, Chen e Wang (2007), Orgler (1971) e Steenackers e Goovarts (1989) acreditam que características como sexo, idade, estado civil, tempo no mesmo endereço, dentre outras, são importantes para a construção do *score*. Já Ong (2002) e Greene (1998) sustentam que emprego atual, casa própria, contas bancárias, hipotecas, possuir bens móveis etc., são boas fontes de dados para construir os modelos de pontuação. Orgler (1971) estende mais um pouco e reforça que a inclusão dos dados do cônjuge pode auxiliar no refinamento dos modelos.

Com relação as técnicas de classificação, para Lee e Chen (2005), Silva, Nakamura e Moraes (2012), Kay e Titterington (1999) e Lenard, Alam e Madey (1995), a seleção das características, em alguns casos, foram baseadas em análises estatísticas, ou seja, regressão logística *stepwise*¹⁸, regressões ou redes neurais. Conforme Hand e Henley (1993), algumas técnicas se dividem em: métodos convencionais e técnicas de estatísticas avançadas. Os primeiros incluem peso da evidência, regressão linear múltipla, análise discriminante, análise *probit* e regressão logística. Já as técnicas de estatísticas avançadas compreendem outras abordagens como algoritmos *fuzzy*¹⁹, algoritmos genéticos²⁰ e sistemas especialistas.

Assim, a seleção de variáveis utilizadas na construção dos modelos de pontuação de crédito, depende dos provedores de dados e da disponibilidade dos dados conforme declarado por cada autor. Todos eles contextualizados à sua maneira para categorizar os indivíduos tomadores de crédito como aceitos ou rejeitados em relação a essas características.

De acordo com Ramos (2016), nos últimos 50 anos, o tema sobre modelos de risco de crédito tem sido algo bastante discutido na literatura acadêmica. Beaver (1966), foi um dos

¹⁸ Modificação da seleção Forward em que cada passo, todas as variáveis do modelo são previamente verificadas pelas suas estatísticas F parciais. Uma variável adicionada no modelo no passo anterior pode ser redundante para o modelo por causa do seu relacionamento com as outras variáveis e se sua estatística F parcial for menor que F_{out} , ela é removida do modelo.

¹⁹ Forma de lógica multivalorada, na qual os valores de verdade das variáveis podem ser qualquer número real entre 0 (correspondente ao valor falso) e 1 (correspondente ao valor verdadeiro).

²⁰ Algoritmos evolutivos que usam técnicas inspiradas pela biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação (ou crossing over).

primeiros a elaborar testes estatísticos, utilizando modelos univariados²¹, prevendo uma provável falência das empresas. De acordo com este autor:

“A insolvência de uma empresa pode ser definida como a probabilidade de que o reservatório se vai esgotar, ponto ao qual a empresa não será capaz de pagar as suas obrigações financeiras à medida que estas vão sendo exigíveis.” (BEAVER, 1966)

O modelo desenvolvido por Beaver (1966), analisava cada rácio financeiro²² de forma isolada, (essa era a principal limitação do seu modelo), pois só poderia ser analisado a média comparativa de rácios financeiros dentre empresas falidas e não-falidas, o que seria determinantemente limitativo quando se faz uma análise financeira, visto que não seria viável agregar diferentes rácios, o que impossibilitaria não só uma visão atual da empresa bem como à sua evolução.

Altman (1968), corrigiu este tipo de falha utilizando análise discriminante multivariada²³ em seu conhecido modelo *ZScore*, sua modelagem é tão influente, que até hoje é difundida tornando-se marcante na literatura acadêmica, onde é considerada como o primeiro modelo de risco de crédito. Altman (2013), revisou seu próprio modelo e ajustou alguns valores de mercados pelos valores contabilísticos dos ativos. Altman *et al.* (1998), basearam o modelo apenas em empresas americanas, o que tornava algo não muito representativo em empresas de outros países. Então, para expandir para os mercados emergentes, surgiu o modelo *Emerging Market Scoring* (EMS).

A utilização de regressão logística foi classificada por Steenackers e Goovarts (1989), como o método mais utilizado para o desenvolvimento do *score* de crédito e os usos de técnicas de inteligência artificial, oriundas da teoria de aprendizagem estatística como as árvores de decisões, conforme Breiman *et al.* (1984), e as redes neurais (MALHOTRA, 2002), cresceram exponencialmente nos sistemas de *score* de crédito, tanto para a sua construção quanto para a previsão de falência (LENSBERG, EILIFSEN & MCKEE, 2006), análise de estresse, e outros, conforme explicado por Gestel (2006).

²¹ Regressão com apenas uma variável explicativa.

²² Valores escolhidos do relatório e contas de uma empresa, com o objetivo de determinar alguma característica financeira dessa empresa, seja a sua solvabilidade, risco, nível de valorização ou qualidade como investimento.

²³ Técnica da estatística multivariada utilizada para discriminar e classificar objetos. Estudando a separação de objetos de uma população em duas ou mais classes.

Trippi e Turban (1992) trouxeram um dos primeiros modelos baseados na aplicação de redes neurais aplicados na área de estatísticas, em meados do início dos anos 90. Em termos gerais, uma rede neural, é uma forma de informação inspirada no cérebro humano (inteligência artificial). Sua base é estabelecida em redes neurais que reagem a estímulos a fim de produzirem determinados resultados. Cada neurônio recebe um estímulo e determinada informação primária, com o objetivo de que um resultado seja criado por elementos desta rede. Conforme Haykin (2007), as redes neurais artificiais (RNA's), assemelham-se a um processador computacional inspirado no comportamento humano, utilizando técnicas computacionais capazes de armazenar um conhecimento experimental prévio e torná-lo disponível para o uso, sendo sua capacidade de aprendizagem um dos pontos mais ressaltados.

Rocha (2020) relatou em seu estudo que podem ser encontrados inúmeros tipos de RNA's, cada qual com suas próprias características que melhor se adaptam na resolução de um determinado problema, dada a sua capacidade de predição de variáveis ou comportamentos para aplicação em finanças e controle. Conforme explicado por Almeida *et al.* (2004), as redes neurais podem ser aplicadas nas análises de crédito, enquanto Lachtermacher e Espenchitt (2001) estudaram suas bases em previsão de falências. Corrêa e Machado (2004) aplicam os conceitos de RNA's construindo modelos de *credit scoring* para o microcrédito baseado no limite do cheque especial, com o objetivo de ampliar a utilização desse limite, identificando os tomadores de crédito como “bons” ou “maus” pagadores.

Entretanto, como em qualquer mecanismo matemático, o modelo de redes neurais, como apresentado no estudo desenvolvido por Pereira *et al.* (2018), pode apresentar algumas falhas e erros no processamento das informações utilizadas para as análises de crédito. Mesmo com “erros”, esses mecanismos contribuem e conferem as instituições financeiras e tomadoras de crédito subsídios e agilidade para as tomadas de decisões.

De acordo com Lima e Kimura (2009), o gestor da instituição credora deve ter em mente dois aspectos: primeiramente, observar o benefício de uma análise com mais profundidade e com menos automatização, com seus devidos custos das perdas esperadas provenientes de uma classificação inadequada; e em segundo lugar, deve-se levar em consideração tanto os casos em que bons pagadores podem ser classificados como maus pagadores, quanto os maus pagadores podem ser classificados como bons pagadores.

Da Silva Filho (2009) ressalta que é importante que a instituição possua um banco de dados que conceda a interferência dessas perdas potenciais. Investimentos em TI, não somente em modelos matemáticos, criam vantagens competitivas para a modelagem de riscos de crédito, permitindo melhores previsões.

Ainda conforme Lima e Kimura (2009), o segundo aspecto a ser observado deve ser as políticas de crédito em relação à estratégia da empresa. Caso, porventura, o objetivo seja elevar sua participação no mercado ou mesmo adotar uma postura mais agressiva para o enfrentamento da concorrência, a instituição financeira pode conceder empréstimo ao indivíduo tomador de crédito, mesmo que ele tenha sido classificado entre os maus pagadores. Em sua contrapartida, se a instituição financeira projetar uma política para minimizar perdas com o fator da inadimplência, uma postura mais conservadora deverá ser adotada, negando o crédito, mesmo para aqueles em que as redes neurais classificaram como bons pagadores.

Neste trabalho adotou-se a classificação de Saunders (2000), embora existam diversas formas de categorização na literatura. Assim sendo, os modelos de *score* de crédito podem ser divididos em três: o modelo de aprovação de crédito (*credit scoring* e *rating*), o modelo comportamental (*behavioural scoring*) e o modelo de concessão de crédito (sistemas especialistas de análise subjetiva).

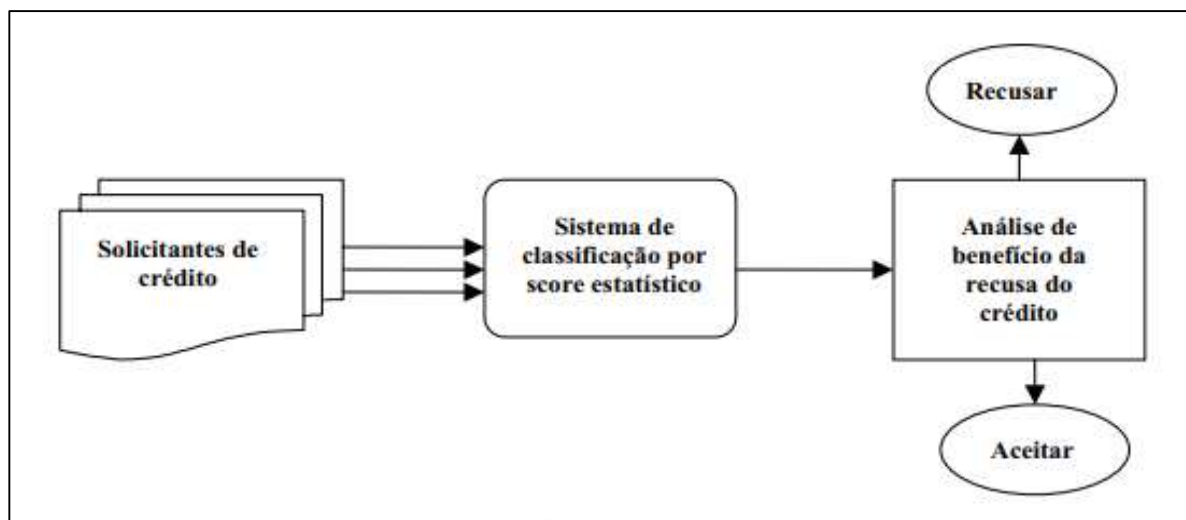
Credit Scoring – De acordo com Saunders (2000), o sistema de *credit scoring* é amplamente difundido e utilizado em praticamente em quase todos os tipos de análise de crédito. Conforme descrito por Silva (2000):

“O sistema de *Credit Scoring* possibilita resposta rápida para decisão de crédito massificado. Você imputa os dados de seu potencial cliente no sistema e imediatamente o computador informa se o crédito foi aprovado. Na realidade o método estatístico também leva em consideração a experiência da instituição com os clientes, [...]. As empresas e instituições podem utilizar metodologias e modelos estatísticos diferentes.” (SILVA, 2000).

Através de uma fórmula originada das variáveis referentes ao proponente de crédito e/ou a operação de crédito, o sistema de *credit scoring* pode ser compreendido como a probabilidade de inadimplência ao se confrontar com a pontuação estabelecida como ponto de corte (ou mínima aceitável) de um determinado crédito. Conforme explica Saunders (2000), o *score* pode ser utilizado para classificar como adimplentes ou inadimplentes, bons ou maus, desejáveis ou não, de acordo com a pontuação obtida por cada crédito. Assim, esta classificação pode orientar a decisão do analista em relação à aprovação ou não do crédito solicitado.

Este sistema parte do pressuposto de que as características dos tomadores de crédito que se tornarão inadimplentes no futuro serão similares às dos tomadores que ficaram inadimplentes no passado. Em sendo assim, ao se utilizar de amostras dos adimplentes e inadimplentes em determinado período passado e aplicando determinadas técnicas estatísticas, o sistema de *credit scoring* (figura 2) permite deduzir a possível inadimplência de determinado tomador de crédito.

Figura 2: Sistema *Credit Scoring*



Fonte: Chaia (2003)

Rating – Já o sistema de *rating* consiste em uma expectativa de aprovação ou reprovação de crédito ao tomador, através de mensuração e ponderação de variáveis determinantes do seu desempenho passado (NASCIMENTO, KOMATI & ANDRADE, 2019).

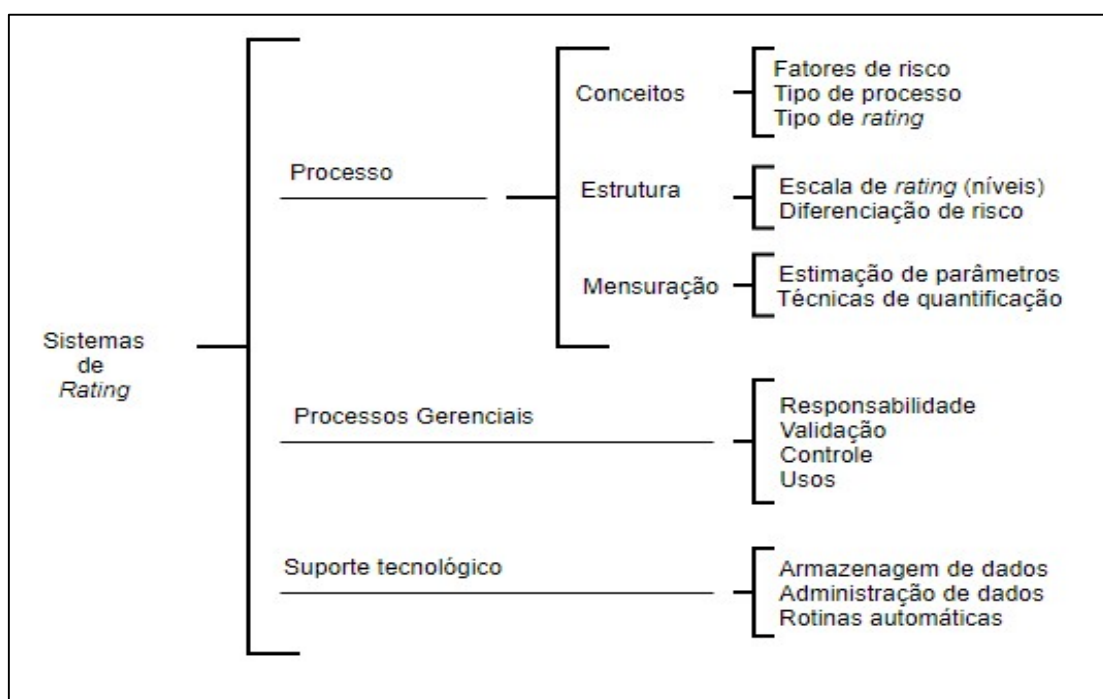
Os sistemas de *rating* classificam os empréstimos em categorias de risco de crédito, de acordo com a possibilidade de perda, conferindo notas conforme a pontuação obtida. A avaliação não é feita caso a caso, mas por meio de faixas de grupos de crédito, obtidos através da distribuição das notas (*ratings*) aos possíveis tomadores de crédito. Conforme Marqués, García e Sánchez (2012), tal *rating* oferece um cenário da capacidade financeira do tomador de crédito e o risco a ele associado.

De acordo com Caouette, Altman e Narayanam (1998), os sistemas de *rating* podem ser desenvolvidos pela própria instituição ou, o que é mais comum, ser obtidos através de birôs de crédito. Estes birôs de crédito são empresas especializadas que fazem a avaliação do risco de crédito para determinada pessoa física ou jurídica. Nas variadas modalidades de concessão de

crédito, é analisa somente a possibilidade econômica do tomador de recursos, tendo em consideração os seus dados.

No final do século XX, foi publicado um estudo intitulado “*Range of Practice in Bank’s Internal Rating Systems*” (BASLER AUSSCHUSS FÜR BANKENAUF SICH, 2000), sendo este o ponto de partida para as análises e classificações de risco nos sistemas de *rating*. De acordo com este estudo, os sistemas de *rating* compreendem três componentes fundamentais: o processo propriamente dito, os procedimentos gerenciais associados a esses procedimentos e as ferramentas tecnológicas utilizadas. A figura 3 exibe esses três componentes de forma esquemática.

Figura 3: Estrutura do sistema de *Rating*



Fonte: Garcia (2002)

Concessão de crédito – Saunders (2000) afirma que os modelos de concessão de crédito são os que têm como respaldo a análise de crédito clássica, baseados no julgamento subjetivo dos analistas especializados na área e na sua decisão em conceder ou não o crédito.

Marqués, García e Sánchez (2013) acrescentam que nos modelos de concessão de crédito, a expertise dos profissionais mais antigos é de suma importância, visto que tal experiência fornece parâmetros mais acurados para a análise do crédito. Neves e Silva (1998) ressaltam que os fatores analisados são extraídos das principais características dos tomadores

de crédito, guiados pelos C's do crédito que, após minuciosa análise e ponderação, culmina em uma conclusão com base na pontuação obtida. Já para Securato (2012), os dados obtidos dos tomadores de crédito são condensados e confrontados com os parâmetros a que se referem, sendo posteriormente pontuados.

Os parâmetros para a concessão de crédito são norteados pelos 5 C's do crédito, conforme explicação de (WESTON & BRIGHAM, 2000) e exibidos na tabela 2.

Tabela 2: 5 C's do Crédito

Caráter	Aferição da reputação da empresa, bem como um prognóstico em relação ao refinanciamento ou liquidação de suas dívidas. Um dos dados levados em consideração é sua boa reputação em saldar suas dívidas.
Capital	É visto como boa medida da probabilidade de falência de uma empresa. Relaciona a contribuição ao capital próprio, realizada pelos proprietários, e a alavancagem (índice de endividamento da empresa). Uma alta alavancagem sugere uma maior probabilidade de falência do que as de baixo índice.
Capacidade	Representa a volatilidade dos ganhos do tomador de empréstimo e sua capacidade de pagamento/refinanciamento. Se os pagamentos/refinanciamentos dos contratos apresentarem fluxo constante ao longo do tempo, entretanto os ganhos forem muito voláteis, a probabilidade de haver períodos cuja capacidade da empresa em quitar as obrigações seja restringida.
Colateral (Garantia Real)	Conforme ocorre a inadimplência, o credor possui direito sobre as garantias reais oferecidas pelo devedor. Quanto maior a garantia, menor o risco de exposição.
Ciclos ou Condições (Econômicas)	Um fator importante na exposição de risco é o estado do ciclo de negócios, em especial para aqueles setores que dependem desses ciclos. Por exemplo, setores de bens duráveis dependem mais do que os não-duráveis. Da mesma forma, os setores que têm exposição a condições competitivas internacionais são mais sensíveis aos ciclos.

Fonte: Blatt (1999).

Comportamental (*behavior scoring*) – no modelo comportamental, o tipo de comportamento econômico é adicionado, sendo analisado todo o histórico progresso de compras e pagamentos do tomador de crédito. Conforme Garcia (2002), essas análises se baseiam na experiência dos analistas ou gerentes de negócios. Securato (2012) aponta que sua composição é complexa, na medida em que envolve grande quantidade de amostras de variáveis particulares, atreladas ao comportamento dos tomadores de recursos. Securato (2012) analisou a metodologia de *behaviour scoring*, compilou e as elencou conforme tabela 3.

Tabela 3: Análise das obrigações assumidas pelo tomador de crédito

Hábitos de consumo	Registro das compras dos tomadores de crédito e por quais meios foram realizadas (exemplo: compra direta em lojas, telefone ou via internet);
Hábitos de lazer	Frequência dos tomadores de crédito em academias, teatros, restaurantes, clubes frequentados;
Viagens	Frequência de viagens (nacionais ou internacionais) para negócios ou lazer;
Tipos de aplicações financeiras	Verificação (carteira de aplicação do tomador de crédito), se ele é avesso, indiferente ou tende a assumir os riscos;
Compatibilidade de renda e patrimônio do tomador de crédito	Análise e comparativo dos hábitos de consumo, lazer, viagens de acordo com sua disponibilidade de renda.

Fonte: Securato (2012).

Nos estudos de Silva (2011), a diferença entre o modelo *behaviour score* para o modelo *credit score* se dá pelo fato de que o segundo avalia a aceitação do tomador de crédito e o primeiro faz a avaliação comportamental do tomador de crédito.

2.1.1 Principais modelos de *score* de crédito no mundo

2.1.1.1 EUA

Como visto em Komorad (2002), os pais fundadores da pontuação de crédito, Bill Fair e Earl Isaac, construíram o primeiro sistema de pontuação de crédito para a *American Investments* em 1958 e, embora a pontuação de crédito seja usada desde então, ela só se tornou amplamente difundida nas últimas duas décadas. Essa expansão teve origem devido ao aumento no número de negócios voltados à concessão de crédito, nas formas de empréstimos ao cliente, empréstimos a pequenas empresas, hipotecas residenciais para *marketing* direto, previsão de falha corporativa, gerenciamento de risco de telecomunicações e assim por diante.

Para Chatterjee *et al.* (2020), as pontuações de crédito são um elemento essencial para o acesso do mutuário ao crédito. Nos Estados Unidos, os birôs de crédito e agências de classificação de crédito cumprem essa função para crédito individual e empresarial, criando e mantendo as pontuações de crédito para tomadores individuais. Existem agências semelhantes em muitos outros países, onde as pontuações de crédito afetam os termos do empréstimo e mudam com o uso e reembolsos do crédito.

2.1.1.2 China

De acordo com Wassler (2019), em meados de 2014, o Partido Comunista Chinês (PCC) traçou seus planos para a construção de um Sistema de Crédito Social (SCS), cuja implementação se deu no ano de 2020, pelo Estadual Conselho da República Popular da China. Seu propósito era o desenvolvimento da economia de mercado através do aumento da confiabilidade e transparência de devedores e credores, reduzindo os riscos com fraudes e o não cumprimento das obrigações de crédito. O sistema de *score* de crédito está sendo monitorado com a gigante chinesa da tecnologia “Alibaba” que utiliza seu robusto banco de dados para compilar as pontuações de crédito. Neles são analisados o histórico de crédito, capacidade de cumprimento, informações pessoais, comportamentos e preferências e relacionamentos interpessoais conforme explicado por Botsman (2017).

2.1.1.3 Tailândia

A Tailândia foi acometida por uma crise econômica em meados dos anos 90, na qual sua dívida externa foi avaliada, de acordo com o FMI²⁴, em US\$ 99 bilhões, ou seja, cerca de 55% do seu PIB²⁵ (MEEAMPOL, 2014). De acordo com Craig (2000), a queda dos mercados de ativos e a falência generalizada dos bancos acabaram por agravar os problemas de fluxo de caixa e inadimplência da maioria das empresas (URAPEEPATANAPONG, SETHSATHIRA & OKANURAK, 1998). Visando reverter tal cenário, Samarakoon e Hasan (2003) mencionaram em seus estudos que o sistema financeiro tailandês passou a utilizar o sistema de *score* (*Emerging Market Score Model*) para inferir uma avaliação de riscos e mitigar as possíveis dificuldades em relação à economia nacional, ajudando, naquele momento, a fornecer sinais de estabilidade. Os investidores teriam as informações necessárias para analisar as empresas que, em um futuro próximo, poderiam apresentar algum tipo de problema financeiro, fazendo com que os gerentes dessas empresas focassem no controle da estabilidade financeira e de sua liquidez, respectivamente.

²⁴ Fundo Monetário Internacional - Organização financeira que possui a função de oferecer ajudas financeiras pontuais e temporárias aos seus países-membros.

²⁵ Produto Interno Bruto - soma de todos os bens e serviços finais produzidos numa determinada região, durante um período determinado. É um dos indicadores mais utilizados na macroeconomia com o objetivo de quantificar a atividade econômica de uma região.

2.1.1.4 Alemanha

Helfrich (2010) publica em seu trabalho que, na Alemanha, os indivíduos tomadores de crédito que deixam de honrar seus pagamentos vão para o *Schufa* (*Schutzgemeinschaft für allgemeine Kreditsicherung*). Esta é uma Associação de Proteção ao Crédito que utiliza um sistema de pontos para avaliar a credibilidade do tomador de crédito, similar aos SPC e Serasa existentes aqui no Brasil. Essa associação utiliza o sistema *Schufa Score*, medido em percentuais, cujo valor decreta a cada vez que o indivíduo deixa de pagar uma conta. Com o decorrer do tempo, se o indivíduo tomador de crédito estiver em um nível muito baixo, torna-se impossível obter empréstimos, abrir contas bancárias, bem como alugar imóveis entre outros.

Assim como no Brasil, o *Schufa Score* utiliza um método matemático-estatístico que calcula um prognóstico de risco, sendo que tal método não pode ser divulgado. Porém, a *homepage* da *Schufa* comenta que, de acordo com a GDPR²⁶, o indivíduo pode receber informações armazenadas sobre eles mesmos e ainda descobrir de onde vieram e para onde os dados foram encaminhados. Quem tiver o interesse em consultar sua pontuação deve fazê-lo na *homepage* da *Schufa* (em consonância com a GDPR), sendo gratuita a consulta somente uma única vez por ano. No Brasil, a plataforma da Serasa pode ser acessada a qualquer tempo para consulta do CPF do titular.

Ainda conforme Helfrich (2010), estima-se que tal método analise as informações pregressas de um indivíduo e as compare com os dados e as taxas de inadimplência de outras pessoas, obtendo-se um resultado cujo valor de probabilidade indicaria os riscos de inadimplência. Como seu cálculo é efetuado a cada três meses e, ainda de acordo com informações contidas na *homepage* da *Schufa*, a fim de orientar as pessoas sobre boas práticas para uma boa avaliação, as mesmas apontam para que o indivíduo cancele contas bancárias e cartões de crédito extras e concentre tudo em um único banco, que evite trocas de banco desnecessárias, mudanças de endereços constantes, pagamento das faturas em dia e evitar fazer constantes pedidos de crédito (toda vez que feito um pedido de crédito (*kreditkonditionenfrage*) o *Schufa Score* gera um registro e são avaliados de forma negativa.

²⁶ *General Data Protection Regulation* (Regulamento Geral sobre a Proteção de Dados) Regulamento do direito europeu sobre privacidade e proteção de dados pessoais, aplicável a todos os indivíduos na União Europeia e Espaço Económico Europeu criado em 2018.

2.1.1.5 Itália

Já na Itália, de acordo com Pennoni e Stanghellini (1999), toda prestação de crédito concedida representa uma operação arriscada, visto que, em primeiro lugar, há um *gap* entre o momento de desembolso do crédito e o do reembolso do capital com seu respectivo pagamento de juros. Em suas concepções, os pesquisadores deduzem que, com base na qualidade do perfil do requerente e nos objetivos da utilização do crédito, é atribuído um grau de risco na qual o volume do empréstimo pode ser reembolsado, o que define a possibilidade ou não para o seu aceite ou recusa.

Filotto (1996) afirma que a área de gestão de risco foi reformulada, convergindo em algumas orientações mais incisivas, incluindo estratégias tais quais: divisão das atividades de análise, especialização por segmentos de mercado e a busca por um maior relacionamento com o cliente. Com isso, houve a necessidade de melhores informações e uma melhor formação dos analistas de risco, não se limitando apenas a dizer se o cliente é “bom” ou “mau” pagador.

2.1.1.6 Espanha

Na Espanha, como visto em Bonilla, Olmeda e Puertas (2003) os principais modelos utilizam redes neurais artificiais, no qual comparam modelos paramétricos e não paramétricos. Segundo Gonzalez e Woods (1992), o maior benefício da modelagem por redes neurais é sua capacidade de determinar planos não-lineares de separação de classes, árvores de decisão²⁷ e técnicas paramétricas²⁸, ou seja, *logit*²⁹ e análise discriminante³⁰. Eles aplicaram esses modelos para avaliação de risco no cartão de crédito e concluíram que a capacidade preditiva dos modelos paramétricos é maior do que nos modelos não paramétricos. Von Zuben (1996) detalha em sua tese que uma rede neural representa um modelo paramétrico sempre que, em sua arquitetura, a estrutura de conexão e números de neurônios forem previamente definidas. Nos modelos não-paramétricos, a arquitetura da rede neural pode ser definida em função do problema de aproximação. Esta flexibilidade se dá pelo fato de que o modelo estará

²⁷ Mapa dos possíveis resultados de uma série de escolhas relacionadas.

²⁸ Presume que os dados são provenientes de um tipo de distribuição de probabilidade e faz inferências sobre os parâmetros da distribuição.

²⁹ Em matemática, a função *logit* é a inversa da função logística.

³⁰ Técnica da estatística multivariada utilizada para discriminar e classificar objetos.

invariavelmente associado à presença de um extenso número de parâmetros e nunca a uma suposta natureza não paramétrica do modelo.

Quando ocorrem mudanças ambientais ou populacionais, os modelos estáticos geralmente falham em se adaptar, precisando ser reconstruídos do zero. Ali e Ghauri (2013) analisaram as características socioeconômicas de clientes tomadores de crédito usando regressão logística e estatísticas descritivas. Bellotti e Crook (2008) introduziram variáveis macroeconômicas como a taxa de desemprego e a taxa de juros em seu modelo de pontuação de crédito, utilizando análise de sobrevivência. Para efeito de estimativa, a análise de sobrevivência é mais poderosa do que a regressão logística para possibilitar o modelo de pontuação de crédito.

De acordo com Lee *et al.* (2002), a pontuação de crédito é considerada uma tarefa importante, dado que o setor de crédito experimentou uma taxa considerável de crescimento. Com ajuda da tecnologia, o uso de redes neurais artificiais tornou-se popular devido à sua característica de memória associativa³¹ e capacidade de generalização. Os autores demonstraram que a inclusão dos resultados de pontuação de crédito da análise discriminante simplificou a estrutura de rede e melhorou os resultados sobre as pontuações de crédito do modelo de rede neural projetado, visto que as tarefas de pontuações de crédito foram realizadas em um conjunto de dados dos cartões de crédito do banco em questão. Seus estudos revelaram que a abordagem híbrida³² converge muito mais rápido do que o modelo de redes neurais convencionais. Além disso, as precisões da pontuação de crédito aumentam nos termos da metodologia proposta, superando a análise de discriminante tradicional e as abordagens de regressão logística.

Conforme Liberati, Camillo e Saporta (2017), com relação as redes neurais artificiais, estas envolvem um imenso esforço de processamento de dados por meio de análise interativa de seleção de variáveis. Isto requer um conhecimento especializado de domínio e profundidade em compreensão de dados. Além disso, esses métodos não são eficazes para problemas com alta dimensão e pequeno tamanho da amostra, por exemplo, novos credores. Nas aplicações do

³¹ Modelos inspirados na habilidade do cérebro humano de recordar por associação.

³² Executa ciclos compostos por uma etapa de otimização de topologia de redes neurais seguida por outra etapa de otimização de pesos. Estes ciclos continuam até serem encontrados uma topologia de rede neural e um ajuste otimizado para os pesos das conexões.

mundo real, algumas suposições devem ser realizadas sobre os dados, de forma linear, para que os dados possam seguir certas distribuições. Mais importante constantemente, com base nesses algoritmos, é difícil automatizar o processo de modelagem e projetar um processo contínuo fluxo de atualização.

2.1.1.7 Portugal

Em 2007, Portugal vivenciou uma crise financeira que culminou na elevação dos preços habitacionais. A necessidade em cobrir bens essenciais e necessidades financeiras das famílias fizeram com que as pessoas tivessem que recorrer a tomada de crédito, sendo a maior parte ligada ao crédito habitacional (BANCO DE PORTUGAL – BOLETIM ECONÔMICO, 2007).

De acordo com Vicente (2015), a análise de crédito é de responsabilidade da CRC (Central de Responsabilidade de Crédito) gerido pelo Banco de Portugal. Essa análise é baseada na modelagem de *scoring*, cujo birô possui responsabilidades sobre as efetivas e potenciais operações de crédito e seus serviços relativos ao processamento e difusão. A CRC apoia as entidades participantes na avaliação do risco de crédito e, através dela, pode-se verificar todos os créditos que o tomador possui, bem como os encargos gerados sobre suas responsabilidades creditícias.

A principal operação gira em torno da habitação própria (casas e terrenos), baseando-se na aquisição de imóveis e construções. O sistema possui uma gama de ferramentas que distinguem as operações morosas das não morosas, colocando como ponto de partida as operações de descumprimento (mutuários com atrasos há mais de trinta dias). Na maioria dos casos, o sistema CRC corresponde a informação positiva, comprovando a capacidade dos tomadores de crédito sobre a sua pontualidade aos pagamentos.

2.1.2 Modelos de *score* de crédito no Brasil

Atualmente, existem quatro birôs autorizados pelo Banco Central que aplicam o *score* de crédito nos indivíduos: o SPC Brasil, o Serasa *Experian*, o Boa Vista e o Quod. Cada um deles possui a própria fórmula para elaborar a nota, além de convênios com diversas redes de lojas, instituições financeiras e prestadoras de serviços.

O SPC (Serviço de Proteção ao Crédito) é de responsabilidade da SPC Brasil, sendo fundado em 1955 e constituindo o primeiro serviço com a finalidade de garantir a proteção de

crédito no país. A estrutura de banco de dados é constituída pelas CDL (Câmaras de Dirigentes Lojistas) de todo o território nacional, armazenando dados e acumulando informações sobre todos os consumidores junto ao comércio.

De acordo com Schiminkoski, Tagliapietra e Bertolin (2015), a Serasa, acrônimo de Serviços de Assessoria S.A., foi fundada em 1968 pela FEBRABAN (Federação Brasileira de Bancos) e, em 2007, associou-se à empresa Experian, revertendo-se em uma empresa global, denominada Serasa *Experian*. Com presença em mais de 40 países prestando serviços de informação, *marketing* e gerenciamento de crédito, atualmente é o maior banco de dados da América Latina, compreendendo consumidores, empresas e grupos econômicos. Com seus mais de 40 anos no mercado brasileiro, a Serasa *Experian* participa da maioria das decisões de crédito e negócios tomadas no Brasil (SPC BRASIL, 2020)

A empresa Boa Vista Serviços administra, desde 2010, o SCPC (Serviço Central de Proteção ao Crédito). Semelhante ao SPC, o SCPC está presente em todo o Brasil por meio de escritórios regionais e distribuidores, além da parceria com mais de duas mil entidades representativas do comércio, da indústria e do setor de serviços. O novo banco de dados foi criado devido a instituições financeiras não estarem satisfeitas com o antigo serviço (SPC). Esta cisão resultou no consórcio da Associação Comercial de São Paulo (ACSP), do fundo brasileiro de investimentos TMG Capital, do Clube Diretores Lojistas do Rio de Janeiro, da Associação Comercial do Paraná e da Câmara de Dirigentes Lojistas de Porto Alegre. Em 2011, a Boa Vista fechou um acordo assumindo a *Equifax*³³, ampliando significativamente sua base de dados sobre transações comerciais de tomadores de crédito e empresas (BOAVISTA, 2020).

A Quod nasceu da união dos cinco maiores bancos em atuação no país, com o intuito de impulsionar o Cadastro Positivo no Brasil. Por ser uma empresa nativa em *Big Data* e inteligência artificial, a plataforma criada é capaz de produzir correlações e análises setenta e cinco vezes mais rápido do que a concorrência, com garantia de segurança das informações. A Quod é certificada pela ISO 27.001 (Gestão da Segurança da Informação) e pela ISO 23.301 (Gestão da Continuidade de Negócios), sendo autorizada pelo Banco Central a operar o Cadastro Positivo (QUOD, 2020).

³³ Bureau de Crédito dos Estados Unidos, considerado um dos três maiores birôs de crédito americano juntamente com a Experian e a Transunion.

Conforme verificado nas páginas oficiais dos birôs de crédito, a classificação do *score* de crédito se dá de maneira semelhante, diferindo apenas em como o *score* é formulado, visto que diferentes birôs apresentam diferentes resultados. Sua classificação é baseada na pontuação do Cadastro Positivo, variando entre 0 e 1000 pontos.

Esse *score* de crédito, através do Cadastro Positivo, é elaborado por meio de inúmeras informações cadastrais públicas existentes, tanto positivas quanto negativas, tais como: pagamentos, investimentos, protestos, pendências bancárias, ações judiciais e informações de dados sensíveis como: endereço, idade, sexo. Essa métrica é utilizada pelos órgãos de proteção de crédito para auxiliar as instituições financeiras a traçarem um perfil de risco dos tomadores de recursos e oferecem as oportunidades de crédito mais adequadas. Cada um desses múltiplos dados são analisados de forma peculiar e particular por cada instituição. Os *scorecards* incluem bases comportamentais³⁴, com base nos pagamentos, características de empréstimo, dados financeiros, demografia, a disponibilidade de dados do *bureau*³⁵ de crédito e a natureza do tipo de negócio (DEAN CAIRE, 2004).

Em termos práticos, o *score* é como se fosse uma nota ou “currículo” da sua vida financeira. Quanto maior for o seu *score*, melhores são as chances de os clientes honrarem seus compromissos financeiros. Por conseguinte, quanto menor ele for, maior será a probabilidade de os clientes não cumprirem tais compromissos. O *score* é algo dinâmico e leva em consideração os dados disponíveis na base de dados, sendo sua evolução assíncrona, isto é, acontecendo de forma gradual na medida em que o mercado financeiro volta a confiar no compromisso do indivíduo em honrar seus pagamentos. No mercado brasileiro, a pontuação do *score* é dividida em três faixas:

- *Score* entre 701 e 1.000 pontos: baixo risco de inadimplência;
- *Score* entre 301 e 700 pontos: médio risco de inadimplência; e
- *Score* entre 0 e 300 pontos: alto risco de inadimplência.

³⁴ Modelos comportamentais que preveem a probabilidade de inadimplência.

³⁵ Serviço de informações de crédito, que utiliza informações de adimplência e inadimplência de pessoas físicas ou jurídicas para fins de decisão sobre crédito.

A aprovação de crédito é uma decisão das empresas e não dos birôs de crédito, na qual as empresas avaliam a sua pontuação como parte do processo da liberação de crédito. A figura 4 exemplifica a plataforma de consulta ao *score* da empresa Quod.

Figura 4: Plataforma de consulta ao *score* de crédito

portalcliente.quod.com.br

quod Início Gestão de Usuários Ambiente Financeiro Atendimento Bem vindo, Convidado

quod Consultar Pessoa Física

Consultar outro CPF Gerado em: 17/01/2020 11:13:15 por Luciana Nina.

Score 1

Risco de crédito: MÉDIO
Capacidade de pagamento menor que a média da população brasileira.

Probabilidade de pagamento: 59%
A chance deste cliente não pagar nos próximos meses é de 41%. 2

Entenda o Quod Score

- Pontuação baseada principalmente nos dados do **cadastro positivo** do consumidor
- Nota que vai de 300 a 1000. Quanto mais alto, maior a chance do consumidor pagar
- Banco de dados que vai **além de atrasos em pagamentos**, com mais de 100 atributos
- Análise profunda dos hábitos de pagamentos do cliente e seu histórico de crédito

Quero conhecer mais

Dados do CPF consultado - 000.000.000-00 4

Nome completo: Nome teste

Situação cadastral: **Regular**

CPF: 000.000.000-00

Data de nascimento: 00/00/1999 (21 anos)

Sexo: Masculino

Nome da mãe: Mãe Teste

Endereço atual: Rua teste, 123, apartamento 123
Bairro teste
São Paulo - SP, CEP 0000-000

Telefone atual: (11) 00000-0000

Email: emailteste@quod.com.br

Participações em empresas: 4 participações [ver detalhes](#)

expandir detalhes +

Total de consultas realizadas 5 [ver detalhes](#)

Este CPF foi consultado 5 vezes por outras empresas nos últimos 30 dias

Pendências Financeiras 6

Total de pendências: 7
Valor total: R\$ 1.850,00

- Inadimplências Comunicadas**
R\$ 350,00 | 2 pendências [ver detalhes](#)
- Ações Judiciais**
R\$ 350,00 | 2 pendências [ver detalhes](#)
- Partic. em Recuperações Judiciais e Falências**
R\$ 350,00 | 2 pendências [ver detalhes](#)

O Quod Consulta PF é destinado a apoiar a realização de negócios e análises de crédito. As decisões tomadas com base nas informações acima são de responsabilidade da empresa credora.

Gestora de Inteligência de Crédito S.A.
Alameda Araguaia, nº 2104, 8º andar - CEP 06455-000, Barueri/SP

CNPJ nº 28.042.871/0001-97
Copyright © 2019 Quod. Todos os direitos reservados.

Fonte: homepage Quod (2021)

E com isso surge o seguinte questionamento: esses modelos matemáticos são utilizados, de forma transparente e justa, no Brasil como sistemas para concessão ou negação de crédito? Diante de alguns indivíduos tomadores de recursos que não obtiveram seus créditos aprovados e das inúmeras ações judiciais, o Superior Tribunal de Justiça “deu nascimento” aos temas 710 e 915.

A tese firmada pelo Superior Tribunal de Justiça é a de que, na avaliação de risco do crédito, há de ser respeitado os limites determinados pelo sistema de proteção ao consumidor no sentido da tutela da privacidade e na máxima transparência nas relações de negócios, conforme a previsão do Código de Defesa do Consumidor e da Lei Nº12.414/2011.

Além da preservação no que tange a privacidade das informações, caso o indivíduo consumidor requeira esclarecimentos sobre os critérios utilizados, a instituição não poderá negar-se a informar. Contudo, como ressaltou Cavalcanti (2018), o consumidor não terá direito a saber a metodologia do cálculo, qual a sua fórmula matemática e os dados estatísticos utilizados para pontuação do “*credit scoring*”. Porque a fórmula é fruto de estudos e investimentos, constituindo assim segredo da atividade empresarial. Esse sigilo advém do direito ao sigilo empresarial, previsto na parte final do art. 5º, IV, da lei Nº 12.414/2011.

Cabe salientar que o indivíduo consumidor não precisa autorizar a utilização desse sistema para buscar seu crédito perante as instituições devidamente autorizadas, ainda assim, uma vez que utilizado o critério, desponta para o indivíduo consumidor o acesso as informações que foram utilizadas, bem como dados econômicos, valores pessoais, histórico de compras, dentre outros.

Como consultado no STJ (2014), o Ministro Paulo Sanseverino aponta que a conduta desrespeitosa da instituição que analisa o crédito, desfavorecendo o consumidor, configura abuso de direito (art.187 do Código Civil - Também comete ato ilícito o titular de um direito que, ao exercê-lo, excede manifestamente os limites impostos pelo seu fim econômico ou social, pela boa-fé ou pelos bons costumes), ensejando a sua responsabilidade objetiva e solidária com o consulente, ou seja, pessoa natural ou jurídica que acesse as informações contidas em banco de dados para a finalidade permitida pela Lei do Cadastro Positivo.

2.1.3 Análise do perfil do tomador de crédito

De acordo com o BACEN (2019), a utilização de forma responsável da tomada de crédito impulsiona a economia, favorecendo uma melhoria na vida dos cidadãos. Outrossim, seu uso exacerbado, isto é, sem o devido planejamento, ocasiona sérias consequências para os tomadores de crédito, suas famílias e a sociedade, o que levaria uma condição temerária de endividamento, ocasionando a diminuição do bem-estar.

O Bacen nos revela que, entre 2010 e 2019, o crédito para as pessoas físicas teve um crescimento considerável, passando de 19,9% para 27,8% do PIB, atingindo 58% do total da carteira de instituições financeiras, ante 45% em 2010.

D'Alessio e Iezzi (2013; 2016) apontam que o endividado de risco é o cidadão que possui um volume de dívida superior à sua capacidade de pagamento, cuja persistência e a baixa qualidade do crédito dificultam o gerenciamento dos recursos financeiros e, em último caso, a qualidade de vida. Segundo os autores, para que uma pessoa seja considerada um endividado de risco, é necessário que o tomador possua dois ou mais critérios, abaixo:

- Inadimplemento das parcelas de crédito, ou seja, atrasos superiores a noventa dias no cumprimento das suas obrigações;
- Comprometimento da renda mensal acima de 50%, respectivo ao pagamento do serviço das dívidas;
- Exposição concomitante às seguintes modalidades de crédito: cheque especial, crédito pessoal sem consignação e crédito rotativo (multimodalidades);
- Renda disponível (após o pagamento do serviço das dívidas) mensal abaixo da linha de pobreza.

2.1.3.1 Panoramas do endividamento de risco

Em consonância com os critérios apresentados na seção acima, o Sistema de Informações de Crédito – SCR³⁶ (BACEN, 2020), no ano de 2019, 4,6 milhões de tomadores de crédito foram classificados como endividados de risco, em um universo de 85 milhões de

³⁶ Banco de dados sobre informações remetidas ao Banco Central do Brasil, referente às operações e títulos com características de crédito e respectivas garantias contratados por cliente perante instituições financeiras.

peças. Portanto, significa que, aproximadamente, 2% da população brasileira se encontra nesta situação.

É possível observar, no gráfico 1, a trajetória do endividamento de risco, a partir do ano de 2016, com uma redução de aproximadamente um milhão de tomadores de crédito em 2017. Porém, em 2018, sobe novamente e atinge os mesmos patamares do ano de 2016, com 4,8% milhões de tomadores de crédito, apresentando uma queda no último trimestre da série.

Gráfico 1: Número de tomadores de crédito (endividados de risco³⁷)



Fonte: BACEN (2020)

2.1.3.2 Perfil dos tomadores de crédito por grupo de risco

O perfil socioeconômico dos endividados, conforme tabela 4, foi classificado nas categorias renda, idade, sexo e região segundo dados do BACEN (2020). O percentual de endividados, em termos de faixa etária, apresenta-se crescente com o decorrer da idade, atingindo 7,8% da população acima dos sessenta e cinco anos, sendo quase o dobro dos tomadores de crédito com até trinta e quatro anos. Em relação à renda mensal, os tomadores que possuem rendas entre R\$5 mil e R\$10 mil são os que mais se endividam, chegando aos 6,5%, manifestando comportamento semelhante àqueles que se encontram em faixas intermediárias de renda. Os que estão nas extremidades, isto é, estão na cauda (de menor e

³⁷ A análise temporal, a partir de junho de 2016, tendo em vista o novo limite de identificação no SCR, no qual se consideram as dívidas do cliente, em dia ou em atraso, que no conjunto tenham valor igual ou superior a R\$200,00 (Carta Circular 3.786, de 18 de outubro de 2016).

maior renda), cujos valores são de R\$1 mil ou superior a R\$10 mil reais, apresentam as menores proporções de endividamento de risco. Levando-se em conta o sexo, o percentual de endividados do sexo feminino é levemente superior ao do sexo masculino. Isto sugere que a propensão ao superendividamento pode ser um fenômeno de renda média com leve predominância do público feminino.

Tabela 4: Perfil socioeconômico dos endividados de risco

Dezembro de 2019				
Classificação	Tomadores de crédito		Endividados de risco	
	Milhões		Milhões	%
Total	85,2	4,6	5,4	
Por região				
Centro-Oeste	6,4	0,3	5,3	
Nordeste	19,9	1,1	5,4	
Norte	5,0	0,3	6,1	
Sudeste	40,2	2,1	5,3	
Sul	13,7	0,8	5,5	
Por local				
Fora das capitais	60,9	3,3	5,4	
Em capitais	24,3	1,3	5,4	
Por sexo				
Feminino	44,0	2,4	5,6	
Masculino	41,2	2,2	5,2	
Por faixa de idade				
Até 34 anos	23,9	0,9	3,8	
34 a 54 anos	35,0	1,7	4,9	
55 a 65 anos	14,0	1,0	7,2	
Acima de 65 anos	12,4	1,0	7,8	
Por faixa de renda				
Até R\$1 mil	12,2	0,7	5,7	
De R\$1 mil a R\$2 mil	25,5	1,2	4,7	
De R\$2 mil a R\$5 mil	31,5	1,8	5,6	
De R\$5 mil a R\$10 mil	10,9	0,7	6,5	
Acima de R\$10 mil	5,1	0,2	4,7	

Fonte: BACEN (2020)

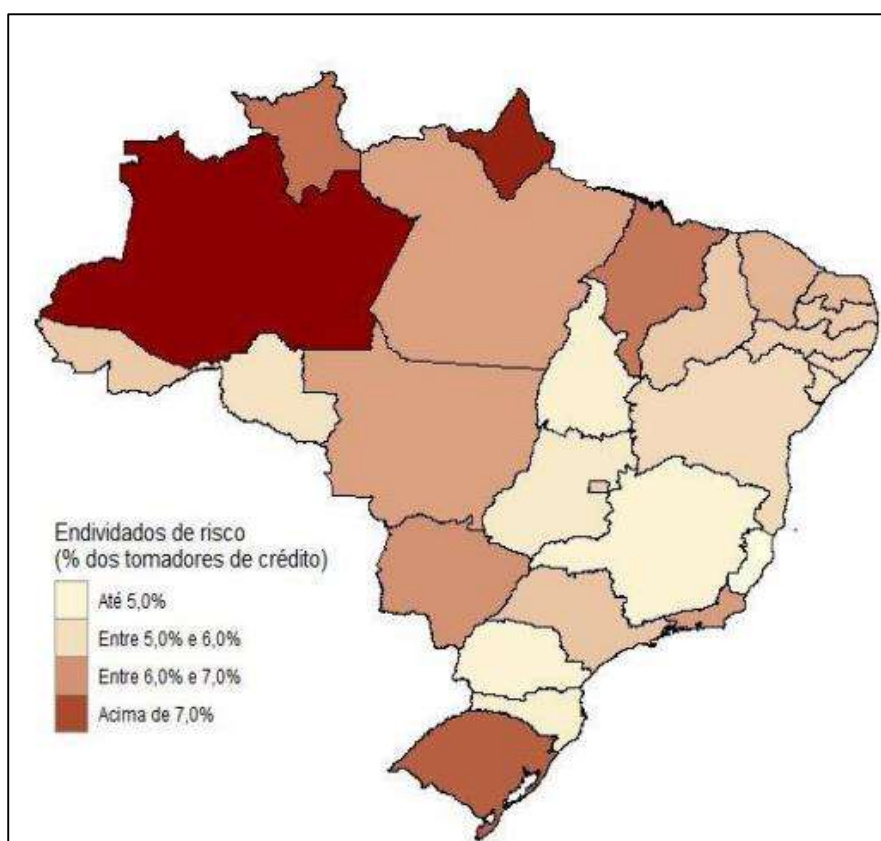
A tabela 4 ainda demonstra que, no geral, a incidência de endividamento é homogênea entre as regiões, sendo a média nacional de 5,4 da população endividada, com exceção da Região Norte, que teve aumento de 6,1%. Nesta mesma região, foram encontrados três dos cinco estados com maiores incidências:

- Amazonas – 7,7%

- Amapá – 7,5%
- Roraima – 6,5%

No Gráfico 2, nota-se o Rio Grande do Sul com 6,7% e Maranhão 6,4% apresentando elevada proporção de endividados. Em relação a localização dentro dos Estados, a proporção de endividados que reside nas capitais é equivalente à proporção de outros municípios.

Gráfico 2: Endividados por Unidade da Federação



Fonte: BACEN (2020)

2.1.3.3 Relação entre os indicadores de risco

Como visto no BACEN (2020), a tabela 5 aponta os critérios de endividamento de risco, a saber: inadimplência, multimodalidades e renda disponível abaixo da linha da pobreza foram comparados com o critério comprometimento de renda, a partir da inexistência ($X = 0$) ou existência ($X = 1$). Foram testadas a média, a mediana e desvio-padrão desse comprometimento e o total dos clientes de cada cruzamento. Como exemplo, na primeira linha, o valor de 3,2% informa a mediana do comprometimento de renda dos tomadores de crédito que não são classificados como renda disponível abaixo da linha da pobreza ($X = 0$).

Os tomadores de crédito na faixa de multimodalidade de crédito, assim como os que estão abaixo da linha da pobreza, isto é, com ($X = 1$), possuem média e mediana de comprometimento da renda acima do público não exposto a esses indicadores ($X = 0$). Todavia, a população inadimplente possui um comprometimento de renda parecido à população adimplente. A distinção ocorre de acordo com a separação por renda, ou seja, renda bruta até R\$2 mil o comprometimento de renda do público inadimplente é inferior ao público adimplente, enquanto a situação é oposta na renda mais alta.

Tabela 5: Comprometimento de renda por subpopulações.

Dezembro de 2019								
Indicador	Mediana (%)		Média (%)		Desvio padrão		Clientes (milhões)	
	X=0	X=1	X=0	X=1	X=0	X=1	X=0	X=1
X = Renda disponível abaixo da linha da pobreza (R\$ 439,03)	3,2	113,1	12,5	141,6	0,2	9,1	77,1	6,1
Por renda bruta:								
Até R\$1 mil	0,2	86,8	9	143	0,1	16,0	7,1	1,9
De R\$1 mil a R\$2 mil	0,6	112,2	10	130	0,2	1,3	22,9	1,7
De R\$2 mil a R\$5 mil	6,5	118,9	15	142	0,2	1,3	28,3	1,5
Acima de R\$5 mil	4,2	127,7	14	157	0,2	1,5	18,7	1,0
X = Multimodalidades	4,0	43,8	20,2	63,0	2,5	1,0	79,7	3,4
Por renda bruta:								
Até R\$1 mil	2,5	84,4	37	106	7,4	4,1	9,0	0,1
De R\$1 mil a R\$2 mil	1,5	36,7	17	61	0,5	0,7	24,1	0,5
De R\$2 mil a R\$5 mil	6,8	42,0	19	58	0,4	0,6	28,3	1,4
Acima de R\$5 mil	3,8	46,8	18	65	0,5	0,7	18,3	1,4
X = Inadimplência	5,1	5,4	22,0	21,7	2,6	1,4	72,8	10,3
Por renda bruta:								
Até R\$1 mil	4,6	0,0	40	20	7,7	3,9	8,0	1,1
De R\$1 mil a R\$2 mil	2,5	0,0	19	14	0,5	0,4	20,9	3,7
De R\$2 mil a R\$5 mil	7,5	10,3	21	24	0,4	0,5	25,9	3,8
Acima de R\$5 mil	4,6	17,9	20	34	0,5	0,5	18,0	1,7

Fonte: BACEN (2020)

O Bacen (2020) estabeleceu os indicadores de endividamento de risco, composto das variáveis quantitativas, extraídas do Sistema de Informações de Crédito (SCR), traçando um recorte da população daqueles indivíduos tomadores de crédito que possuem dificuldades em gerenciar seus recursos financeiros.

As subpopulações com renda acima de R\$2 mil e com idade acima de 55 anos mostraram-se mais suscetíveis às vulnerabilidades financeiras, cuja provável explicação consiste em relacionamentos bancários mais aprofundados e o acesso a uma ampla gama de

produtos financeiros com limites mais altos de crédito. Portanto, a análise de vários cruzamentos entre os indicadores demonstrou que o endividamento de risco é um fenômeno de complexidade que alcança os tomadores de crédito por distintos canais.

2.2 *Big Data*

“Para ganhar dinheiro, você tem que prever duas coisas: o que vai acontecer e o que as pessoas acham que vai acontecer.” (VARIAN, 2006)

Na frase supracitada, é possível refletir nas palavras de Varian (2006), economista-chefe do Google e professor emérito da Universidade da Califórnia em Berkeley, que os dados valem muito, sendo considerados o ouro da atualidade e um dos assuntos mais amplamente difundidos na ciência, na prática e notoriamente presente em mais de 70% dos artigos acadêmicos (POSPIECH & FELDEN, 2013). Esses dados, devido a capacidade de serem preditivos e prescritivos, podem colocar uma determinada organização em vantagem competitiva, através da capacidade de antever situações e com isso dispor de decisões mais assertivas.

De acordo com Cooney (2012), seja na ciência, TI, na administração ou na prática, quando se trata de *Big Data*, grandes expectativas se engendram nas possibilidades em armazenamento, processamento e análise de dados, em especial nas indústrias com uso intensivo de dados como no entretenimento, nas telecomunicações, na saúde, nas engenharias, nos serviços financeiros, entre outros. Os dados que compõem a *Big Data* são oriundos de variadas fontes, não apresentando uma estrutura definida, portanto não seria compatível armazená-la em sistemas de banco de dados padrões, como o SGBDR³⁸.

Logo, os cientistas notaram que esses bancos não suportariam essa imensa massa de dados não estruturados e desenvolveram um modelo de programação (*MapReduce*³⁹, criado pela *Google*) que permitisse processar essa enorme quantidade de dados. Subsequentemente, foi desenvolvido o *Hadoop* (que é uma implementação em código aberto⁴⁰ do *MapReduce*). O *Hadoop*, atualmente, é o tipo de processamento mais utilizado por empresas que trabalham com

³⁸ Sistema Gerenciador de Banco de Dados Relacional – onde os dados são representados por meio de tabelas, com inúmeras linhas e colunas.

³⁹ Modelo de programação desenhado para processar grandes volumes de dados em paralelo, dividindo o trabalho em um conjunto de tarefas independentes.

⁴⁰ Modelo de desenvolvimento, que promove o licenciamento livre para o design ou esquematização de um produto, e a redistribuição universal desses, com a possibilidade de livre consulta, examinação ou modificação do produto, sem a necessidade de pagar uma licença comercial.

a *Big Data*. Múltiplas empresas vêm contribuindo para o seu desenvolvimento, como o *Yahoo*, o *Facebook*, o *Cloudera*, a *IBM* etc.

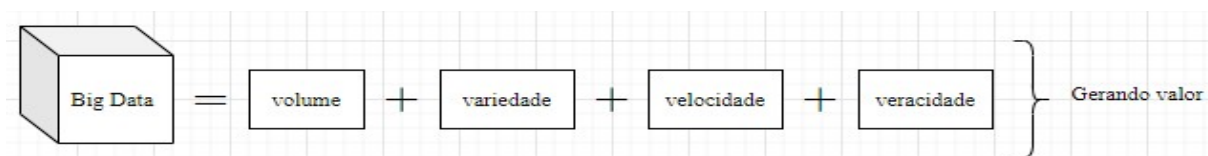
Segundo a *IBM* (2008), são elaborados cerca de 2,5 quintilhões de *bytes* todos os dias, decorrente das adesões de empresas robustas à internet. Pode-se citar como exemplo as redes sociais, os dados de geolocalização, dispositivos móveis e, com o advento da Internet das Coisas, houve um aumento exponencial da quantidade de dados gerados. Dentre os tipos de dados estudados pelos especialistas em *Big Data*, os tipos mais comuns são:

- ***Social Data***: Dados coletados das redes sociais, interação dos usuários, normalmente demográficos ou comportamentais, ditando um certo padrão de um grupo que possuam as mesmas características (usados muito em campanhas de *marketing* – para poder oferecer um serviço mais personalizado);
- ***Enterprise Data***: São os dados empresariais, coletados dos recursos humanos de empresas, setores de vendas, finanças, logística e outras. Utilizados para otimizar processos, identificar falhas, fraudes num determinado setor;
- ***Personal Data***: São os dados pessoais obtidos dos aparelhos de uso pessoal dos indivíduos, demonstrando as preferências pessoais, para desenvolver metodologias personalizadas, no intuito de tornar a relação menos robotizada.

Conforme Silveira, Marcolin e Freitas (2015), o conceito de *Big Data* é um conceito abstrato, que surgiu em meados de 2010 para representar uma tecnologia capaz de gerenciar uma imensa quantidade de dados, das mais variadas origens e formatos (CHEN CHIANG & STOREY, 2012). Davenport e Chow-White (2013) salientaram que o uso de *Big Data* é uma ferramenta para tomada de decisão corporativa e de acordo com Manyika *et al.* (2011) além do volume de dados robusto, outras características fazem parte da literatura *Big Data*: variedade e velocidade.

Para Demchenko *et al.* (2013) a veracidade retrata a necessidade de que os dados coletados possuam genuinidade comprovada e, juntamente com a velocidade adequada, configura o valor imbuído no negócio. De acordo com Taurion (2013) a fórmula da *Big Data* seria:

Figura 5: Fórmula da *Big Data*



Fonte: Autora adaptou do trabalho de Taurion (2013)

Assim, o conceito de *Big Data* pode ser subdividido em 5 categorias (ou “5 V’s”), explicados abaixo:

Volume - Fatores do cotidiano evidenciados pelo conceito *Big Data*: diariamente, troca de e-mails, transações bancárias, interações em redes sociais, registros de chamadas e tráfego em linhas telefônicas. Todos esses exemplos servem de ponto de partida para a compreensão do volume de dados presente no mundo (NETTO, MORO & FERREIRA, 2014).

Variedade – Além do volume, há uma enorme variedade de dados, informações das redes sociais ofertadas a todo segundo e-mails, fotografias, áudios, telefones e cartões de crédito (MCAFEE ET AL. 2012). Empresas que conseguem captar essa variedade, seja de fontes ou de critérios, agregam mais valor ao negócio (SOUZA, MUCHERONI & DE MASSENA, 2018). A *Big Data* pode escalonar vastas informações nas seguintes formas:

- **Dados estruturados:** Armazenados em bancos de dados, sequenciados em tabelas;
- **Dados semiestruturados:** Acompanham padrões heterogêneos, mais difíceis de serem identificados, pois seguem padrões diversificados;
- **Dados não estruturados:** Mistura de dados com fontes variadas como: imagens, áudios e documentos *online*.

Estima-se que até 90% de todos esses dados do mundo (dentre essas três categorias) estão na forma de não estruturados.

Velocidade - Informação é poder e, assim sendo, a velocidade com a qual o indivíduo obtém a informação é uma vantagem competitiva das empresas, podendo limitar a operação de muitos negócios como, por exemplo, quando utilizamos o cartão de crédito – se não obtivermos a aprovação de compra em alguns segundos, normalmente pensamos em utilizar outro método

de pagamento ou, até mesmo, desconsiderar a compra (DUMBILL, 2012). É a operadora perdendo uma oportunidade de negócio pela falha na velocidade de transmissão e análise de dados do indivíduo comprador (TAURION, 2013).

Veracidade – Para obter êxito do processo da *Big Data* é necessário a obtenção de dados verídicos, de acordo com a realidade, isto é, dados que correspondam com a realidade naquele exato momento, pois dados passados podem não ser mais considerados verídicos para o momento atual na qual está sendo analisado. A verificação dos dados ao propósito da análise (adequação e relevância) é ponto chave para angariar dados que agreguem valor ao processo (HURWITZ *ET AL.* 2013).

Valor – Quando se realizam as perguntas certas no início do processo da análise, maior será a riqueza de dados (BAGLEY & BROWN, 2014). Isto faz que seja necessário estar focado na orientação dos negócios, ou seja, no valor que a coleta e a análise de dados trarão para o negócio. Não é viável realizar toda a operação de *Big Data* se não houver questionamentos que auxiliem o negócio de modo realístico. Da mesma forma, é importante atentar-se aos custos envolvidos na operação e o valor agregado a todo esse trabalho desenvolvido. A coleta, o armazenamento e a análise desses dados precisam compensar todo o custo financeiro envolvido (TAURION, 2013).

Conforme a tabela 6, essas são as principais características de uma *Big Data*, por diversos autores.

Tabela 6: Características da Big Data

Características	Autores	Descrição
Volume	Davenport, Barth E Bean, 2012	Centenas de <i>Terabytes</i> ou <i>Petabytes</i>
	McAfee e Brynjolfsson, 2012	Grande quantidade de dados criados por dia
	Demchenko <i>et al.</i> 2013	Relacionado ao tamanho e quantidade de dados
	Agrawal, 2014	Grande quantidade e complexidade de dados
Velocidade	Goldman <i>et al.</i> 2012	Necessidade de respostas em tempo real

	McAfee e Brynjolfsson, 2012	Dados capturados e processados quase em <i>real time</i>
	Demchenko <i>et al.</i> 2013	Dinâmica de crescimento e processamento de dados
	Zikopoulos <i>et al.</i> 2013	Velocidade de captura e análise de dados
Variedade	McAfee e Brynjolfsson, 2012	Grande variedade de fontes e formas de dados com o desafio de encontrar padrões de dados úteis para os negócios
	Demchenko <i>et al.</i> 2013	Diversidade de origens, formas e formatos de dados
Veracidade	Demchenko <i>et al.</i> 2013	Autenticidade, reputação de origem e confiabilidade dos dados
Valor	Demchenko <i>et al.</i> 2013	Dados com significado para os negócios, que contribuam com valor agregado

Fonte: Adaptado de Demchenko *et al.* (2013); Davenport, Barth e Bean, (2012); Agrawal (2014); McAfee e Brynjolfsson (2012); Goldman *et al.* (2012) e Zikopoulos *et al.* (2013)

A célebre frase de Peter Drucker “Você não pode gerenciar o que não pode medir”, faz uma boa alusão ao conceito *Big Data*. Afinal, o que os administradores puderem mensurar e, portanto, deter um melhor conhecimento para os seus negócios, acaba sendo traduzido em decisões mais assertivas.

Galdino (2016) complementa que através dessas mensurações bem executadas, essas correlações de dados podem levar a descoberta de novos padrões, de preferências dos indivíduos, de exponencial aumento nas vendas de um determinado setor e até de prevenção ao contágio de doenças, bem como ficou evidenciado durante a pandemia do coronavírus, no ano de 2020. Através da utilização da *Big Data*, Taiwan pôde integrar o banco de dados do Seguro Nacional com o banco de dados da imigração e alfândega, o que gerou alertas em tempo real com base no histórico de viagens e sintomas clínicos, avançando nas identificações de casos de COVID-19⁴¹ (WANG, NG & BROOK, 2020), dentre outras aplicações à diversas áreas de estudo.

⁴¹ Doença causada pelo coronavírus, denominado SARS-CoV-2, que apresenta um espectro clínico variando de infecções assintomáticas a quadros graves.

“É um erro capital teorizar antes de ter dados. Sem se perceber, começa-se a distorcer os fatos para ajustá-los às teorias, em vez de mudar as teorias para que se ajustem aos fatos.” (DOYLE, 2013)

De acordo com Junior e Albuquerque (2018), é notório que os avanços da tecnologia e comunicação influenciam todo um comportamento da sociedade e, mediante esses avanços, a humanidade produz cada vez mais dados nas mais variadas plataformas digitais, como por exemplo os sensores, computadores, celulares, câmeras etc. Todas essas informações sendo coletadas e armazenadas em enormes quantidades de registros (VAN DER AALST, 2014). Bugnion, Nieh e Tsafirir (2017) complementam que a maioria desses dados produzidos resultam do uso incessante da tecnologia nos últimos anos.

Posto isso, em que os dados são velozmente e abundantemente gerados, funcionando como matéria-prima para tomada de decisão em inúmeras organizações (ECONOMIST, 2017) e, conforme esclarecido por Taurion (2013) em seu livro, a aplicabilidade da *Big Data* auxilia no tratamento desse contingente de volume de dados, advindo de fontes variadas e que necessitam de uma enorme demanda de processamento.

CPBR6 (2013) trouxe em seu estudo o uso de *Big Data* na computação em nuvem⁴² que, devido ao poder de escalabilidade, permite, por exemplo, em épocas como o Natal ou *Black Friday*⁴³, cujas vendas têm um aumento significativo, evitar o aluguel desnecessário de inúmeros servidores para suas operações, dado o pequeno período em questão.

Mesmo a análise desses dados não sendo algo tão simples quanto possa parecer, devido ao grande volume de dados e estes necessitarem de uma gama de ferramentas adequadas para viabilizar sua manipulação, o uso da *Big Data* gera praticidade e o seu suporte auxilia na tomada de decisões.

2.2.1 *Big Data* como mecanismo de mudança no *score* de crédito

Sob o ponto de vista da Analytics10 (2020), a *Big Data* na área financeira é um setor com altas expectativas, posto que as instituições financeiras criam perfis dos clientes tomadores

⁴² *Cloud computing* (computação em nuvem) - tecnologia que usa a conectividade e a grande escala da *internet* para hospedar os mais variados recursos, programas e informações.

⁴³ Temporada de compras com significativas promoções.

de crédito para uma melhor individualização de preços, ofertas de produtos e uma melhor avaliação do alcance da credibilidade.

As soluções tecnológicas que alcançam robustas bases de dados tornam-se o foco de investimento dessas instituições, referenciando o conceito de *Big Data*, na qual o gerenciamento e armazenamento não podem ser feitas por meios computacionais tradicionais, conforme dito por Manyika *et al.* (2011). Sendo assim, a sua utilização para mensurar o *score* de crédito também é aplicada. Logo, os indivíduos que possuem uma boa pontuação no *score* de crédito poderão obter acesso à empregos, moradias, serviços essenciais, obtenção de crédito em bancos, entre outros.

Segundo Hurley e Adebayo (2017), o advento da *Big Data* como forma de tratamento de dados para aferir o *score* de crédito dos indivíduos trouxe uma certa preocupação para os tomadores de crédito, devido às incertezas de como são feitas essas “pontuações” e se teriam a capacidade de contestação caso eles fossem imprecisos, tendenciosos ou injustos ao efetuar uma avaliação de uma determinada pessoa.

“A maioria das pessoas, pelo menos no mundo industrializado, contribui rotineiramente e experimenta os efeitos da *Big Data*. Sob esse conceito, a *Big Data* é a baleia, e todos nós somos Jonas.” (LERMAN, 2013)

De acordo com os estudos de Hurley e Adebayo (2017) as ferramentas de pontuação de crédito, que são constituídas com milhares de dados de milhões tomadores de crédito, são coletadas sem o tomador ter conhecimento de tal fato, criando certos problemas em relação a transparência. Os consumidores entendem que com essa capacidade limitada, fica injusto identificar e contestar quais seriam as etapas que eles poderiam melhorar em seus respectivos créditos.

Para Chen e Cheung (2017), a característica definidora da *Big Data* é a sua capacidade de pesquisa para agregar e cruzar grandes conjuntos de dados para análise e identificar previamente padrões indetectáveis, criando perfis dos indivíduos, calculando riscos, monitorando e até mesmo prevendo certos comportamentos.

Isto levanta questionamentos sobre a precisão desses dados, pairando a dúvida sobre as ferramentas de análises de *Big Data* em criar um sistema de “solvência por associação”, no qual para os tomadores de crédito, as suas afiliações familiares, religiosas e sociais pudessem determinar sua elegibilidade para um empréstimo. Hurley e Adebayo (2017), complementam

que essas ferramentas podem obscurecer uma política de empréstimo discriminatória e subjetiva, por trás de uma pretensa pontuação. Como saber se essa pretensa pontuação não é intencional?

Já Neoway (2019) possui uma opinião controversa e explica em seus estudos que a utilização de *Big Data* garante operações financeiras para os tomadores de crédito de forma eficiente, rápida e segura, conferindo uma concessão de crédito voltada para a prevenção, visando a perda financeira por inadimplência. Por isso, uma análise minuciosa se faz necessário do perfil do cliente, cujas respostas ágeis de um enorme volume de dados fazem com que as instituições fiquem a frente de seus concorrentes, inferindo precisão ao analisar um *score*.

Em seu livro *Weapons of Math Destruction* (Armas da Destruição Matemática) Cathy O’Neil (2016) retrata o risco e o temor em relação ao uso de modelos matemáticos, sintetizados pelos algoritmos, como um modelo de resolução de todas as mazelas da sociedade contemporânea. Seus estudos são baseados na alegação da exatidão matemática como pressuposto para construção de modelos, como o *machine learning*⁴⁴, na qual máquinas seriam capazes de aprender e ampliar preconceitos como raça, etnia, gênero, religião, classe social, posicionamentos políticos entre outros, ampliando as possibilidades de desigualdades (NOGUEIRA, 2020).

Lin (2013) complementa que amizades no âmbito das redes sociais com “amigos inadimplentes”, diminui sua pontuação de crédito. Tais descobertas também foram confirmadas por Freedman e Jin (2014), com uma advertência de que laços online por si só não revelam informações verdadeiras sobre a credibilidade do tomador de crédito e podem ser manipuladas.

De Cnudde *et al.* (2019) desenvolveram um modelo de pontuação para micro finanças usando informações extraídas das redes sociais (*Facebook*). Seus resultados sugeriram que amigos que interagem podem ser mais preditivos do que aqueles que não interagem, pelo fato de se atribuir um valor ao comportamento dos semelhantes. Na indústria, as redes sociais já são exploradas como método para avaliar a credibilidade, utilizando as conexões para analisar o risco de inadimplência das pessoas.

⁴⁴ Método de análise de dados que automatiza a construção de modelos analíticos. É um ramo da inteligência artificial baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana.

Silveira, Marcolin e Freitas (2015) apontam que o uso da *Big Data* vem para amparar as empresas que a utiliza no que tange ao armazenamento e análise dos dados comportamentais dos tomadores de crédito, de forma cada vez mais particular, traçando perfis mais elaborados com grande valor para entidades comerciais, campanhas políticas, governos ou quaisquer outras entidades que desejem monitorar, controlar, monetizar e prever o comportamento humano.

A aplicabilidade da *Big Data* vem amparando diversos setores e, com o seu uso associado a Lei Geral de Proteção de Dados, ainda percorre um caminho recentemente explorado, acerca dos mecanismos de preservação da privacidade. Infelizmente, somente após violações de dados em larga escala é que os provedores de serviços, juntamente com os usuários, começam a compreender e avaliar a importância da incorporação da proteção à privacidade como elemento central para um bom uso da *Big Data*. Essa compreensão demanda uma certa quantidade de tempo, na qual a *Big Data*, de forma crescente, se torna “pessoal, de confiança”, ao mesmo tempo em que os usuários passam a fazer questionamentos e entender seus direitos (D’ACQUISTO, 2015).

2.3 LGPD

Os avanços tecnológicos trouxeram uma intensa geração de fluxo de negócios, inúmeras trocas de informações e, em consequência, mudaram o cenário dos negócios digitais, que atravessaram transformações ao que se refere à proteção de dados. O que antes era comum, inúmeras empresas vendendo banco de dados (perfis de clientes), atualmente com a então promulgada LGPD, conhecida como a Lei Geral de Proteção de Dados Nº13.709/2018, trazendo em seu bojo, aspectos visando o *compliance* da lei no que tange o assunto da preservação dos direitos dos cidadãos sobre seus dados, sob a pena de incorrer em responsabilidade, muita coisa mudou (FILHO, 2013):

“a partir das décadas de 60 e 70, com o advento das tecnologias da informação. O grande poder de processamento de dados pelos computadores foi o fator responsável pela germinação da moderna legislação nessa área. O aumento do poder de controle e processamento de dados prontamente desencadeou a demanda por uma legislação específica para regular a coleta e manuseio de informações pessoais.” (FILHO, 2013)

Considerando este cenário, Bulos (2014) abarcou uma definição muito precisa a respeito dos direitos, consubstanciando ao fato de perfazer um conjunto de normas que promovesse uma convivência harmoniosa entre os indivíduos da sociedade, dando importância suas condições pessoais individuais, ou seja, o respeito e o consentimento com o trato dos seus dados. Cabe

ressaltar que, nos moldes de Mendes e Branco (2017), os direitos fundamentais possuem grande relevância em um contexto social, pois é através deles que há o reconhecimento que o indivíduo tem direitos em relação ao Estado, bem como seus respectivos deveres.

Inúmeras empresas lucram com os dados dos indivíduos, utilizando indevidamente e vazando-os, como, por exemplo, em 2015 o escândalo envolvendo a *Cambridge Analytica* e o *Facebook*, na qual a empresa obteve acesso aos dados coletados de um aplicativo (*thisisyourdigitalife*⁴⁵) da referida rede social, e de antemão sabia seus nomes, profissões, endereços, preferências, seus interesses políticos e sua rede de contatos, ou seja, foram expostos sem ao menos ter conhecimento disso. De posse dessa gama de dados, a empresa traçou perfis, na qual sugeria que tipo de propaganda seria mais eficaz para persuadir uma pessoa em particular e influenciar na eleição da presidência de Donald Trump.

Gomes (2019), nesse contexto, identificou em seus estudos que, visando a privacidade, adveio a Lei Geral de Proteção de Dados, na qual regras foram implementadas, tanto para pessoas físicas quanto as jurídicas, para legislar sobre a segurança e a privacidade de dados, regulamentando e fiscalizando o modo como as empresas coletam, armazenam e utilizam os dados que foram coletados. Para sua criação, a maior inspiração veio da GDPR (*General Data Protection Regulation*) que por sua vez usou como referência a ISO 27000⁴⁶, Cobit⁴⁷ e Itil⁴⁸ como relata Marinho (2019).

Sua criação aproximou os países ricos do Brasil e, em contrapartida, trouxe credibilidade e notoriedade internacional, atraindo novos investimentos. Implantá-la requer uma equipe multidisciplinar dos setores de administração, tecnologia da informação e jurídico, cada qual contribuindo para a adequação da empresa, conforme explica Santiago e Tamba (2018):

⁴⁵ Tradução (essa é sua vida digital), aplicativo disponibilizado na rede social do *Facebook* para recolhimento de dados pessoais.

⁴⁶ Normas do Sistema de Segurança da Informação, relacionadas à segurança de dados digitais ou sistemas de armazenamento eletrônico.

⁴⁷ Sigla para (*Control Objectives for Information and related Technology*) e que, na prática, significa uma estrutura capaz de fornecer governança de TI.

⁴⁸ Sigla para (*Information Technology Infrastructure Library*), conjunto de boas práticas detalhadas para o gerenciamento de serviços de TI que se concentra no alinhamento de serviços de TI com as necessidades dos negócios.

“A promulgação da lei é um marco para o Brasil já que anteriormente não existia legislação específica sobre proteção de dados. O que existiam eram algumas normativas esparsas, como o Código de Defesa do Consumidor, a lei do Cadastro Positivo e a lei 12.965/14 que ficou conhecida popularmente como o Marco Civil da Internet e que foi parcialmente alterada pela LGPD”. (SANTIAGO & TAMBA, 2018)

Pensar, criar e implementar uma lei de proteção de dados que esteja em conformidade com o cenário mundial contribui para o avanço da economia, trazendo o favorecimento do diálogo com outros países. É certo de que a LGPD irá atingir todos os setores que atuam no tratamento de dados, cuja necessidade de compreender e adequar-se possui caráter imediato, embora a lei dê um período para a empresa captar, entender e amoldar-se, por meio de instituição de políticas públicas corporativas adequadas, designando treinamento de pessoal no sentido de resguardar os direitos relativos aos dados pessoais dos cidadãos (MEYER, 2018).

2.3.1 *Score* de crédito à luz da nova LGPD

Oliva e Viégas (2019) retratam que o Superior Tribunal de Justiça, embora reconheça que o *credit score* não constitua especificamente um banco de dados, fomenta que seja duvidoso que os dados estatísticos que constituam a fórmula dependam dos dados pessoais para se tornarem operantes e, em consequência disso, não se possa evitar a análise da origem e qualidade dos dados alimentados por esse sistema à luz da LGPD.

Cortazio (2019) complementa que, mesmo diante dessas técnicas e modelos no que diz respeito ao tratamento com os dados pessoais, a Lei Geral de Proteção de Dados veio para fortalecer na proteção do indivíduo tomador de crédito, imputando deveres e assegurando o *compliance* da lei, permitindo o acesso livre sobre a forma de tratamento de dados pelo sistema *credit scoring*, conforme o art. 6º, IV da LGPD.

O art. 9º da LGPD comenta que o titular dos dados possui direito de acessar seus dados, que deverão ser disponibilizados de forma clara e objetiva, abarcando, entre outros aspectos, a finalidade específica do tratamento, a identificação do controlador e as informações de contato do controlador. (OLIVA & VIÉGAS, 2019)

“Não há crédito sem conhecimento. Não há conhecimento sem informação.” (BESSA, 2011)

Bessa (2011) menciona através dessa frase, que deter o conhecimento dos dados do tomador de crédito, possua uma finalidade de segurança para financeira de acordo com o grau de confiança no indivíduo envolvido. Esse sentido de confiança envolvido está diretamente atrelado as informações que se detém sobre ele, assim sendo, quanto menor o conhecimento

envolvido, maior seria o risco de crédito e quanto mais informações reunidas sobre determinada pessoa, melhor as instituições financeiras poderiam dispor de taxas mais vantajosas.

Bessa (2011) aponta que a maior dificuldade, atualmente, é a exponencial diminuição dessa interação entre as pessoas e comenta:

“Em que pese a identificação dos tempos atuais como a era da informação, os atores do mercado são anônimos, raramente se conhecem. As relações de compra e venda de produtos e serviços são fugazes e automáticas. É justamente nesse contexto de anonimato dos atores do mercado de bens e serviços que se destacam as atividades exercidas pelos bancos de dados de proteção ao crédito, vale dizer, das entidades que têm por principal objeto a coleta, o armazenamento e a transferência a terceiros (credores potenciais) de informações pessoais dos pretendentes à obtenção de crédito.” (BESSA, 2011)

De acordo com Freitas (2020), a Lei Geral de Proteção de Dados entende que o titular dos dados pode solicitar revisão das decisões por meios automatizados e que a guarda dos cadastros realizadas pelos gestores dos bancos de dados estejam sob rígido controle, evitando a utilização dos dados dos tomadores de créditos para outras finalidades, gerando uma certa discriminação na oferta de produtos e serviços. Diante do exposto da lei, entende-se, que o controlador deve fornecer informações claras a respeito dos critérios adotados para definição do perfil de crédito do tomador.

A Lei do Cadastro Positivo – Lei nº 12.414/2011 – estipula que todo o histórico do tomador de crédito fique preservado em banco de dados por até quinze anos, conforme seu artigo 14, porém, em consonância com a LGPD, o titular pode confirmar a existência de tratamento de dados até a eliminação dos seus dados pessoais, salvo em condições de cumprimento da lei. De fato, esse prazo de quinze anos pode ensejar algum tipo de obstrução de créditos por aqueles tomadores de crédito que estejam com o *score* baixo, inviabilizando novas tomada de empréstimos. Freitas complementa que em quinze anos a vida de uma pessoa pode mudar drasticamente - de estagiário a funcionário público ou empreendedor e, no entanto, mesmo com suas dívidas sanadas, suas dívidas ficariam registradas e constando na base de dados, “julgando” seu *score*.

A legislação obviamente mantém em mente que é imputado benefícios àqueles que são referidos como “bons pagadores”, mas, em contrapartida, há um entendimento que pode ser gerado novos obstáculos para aqueles tomadores de crédito que estão em dificuldades para executar o adimplemento das suas obrigações e, conseqüentemente, entrando em uma esfera sem saída para a tomada de novos créditos (FREITAS, 2020).

2.4 Estudos anteriores relacionados

Sendo a LGPD uma lei recente, estudos anteriores que relacionem, na mesma pesquisa, *Big Data* e *Score* de Crédito com a primeira são pouco recorrentes. Contudo, estudos relacionando *Big Data* e *Score* ou LGPD e *Score* já se mostram relevantes. Abaixo, um compêndio dos principais estudos acerca dos três principais pilares, *Big Data*, *Score* de Crédito e LGPD, que embasam esta pesquisa.

2.4.1 *Score* de Crédito versus *Big Data*

Goulart (2016), em seu trabalho, correlaciona *Score* de Crédito e *Big Data*, perfazendo uma reflexão em direção a uma concepção renovada dos arquivos de consumo⁴⁹. O autor inicia o estudo com o relato das insuficiências da teoria tradicional dos arquivos de consumo frente às novas tecnologias de processamento dados.

Ferreira e Costa (2017) analisaram as crescentes disponibilidades de grandes volumes de dados, discutindo os aspectos conceituais e epistemológicos da utilização de dados intensivos e seus reflexos para a área de Finanças.

Teles (2017) abarcou em seus estudos um panorama da utilização de *Big Data* pelo sistema financeiro, demonstrando a sua eficácia, a melhoria na competitividade e os lucros em relação aos seus competidores, especificamente na pontuação de crédito (*credit scoring*).

Freire (2019) aborda em seus estudos uma evolução temporal sobre pontuações de crédito utilizando *Big Data*, nos quais analisou lacunas e limitações. Com isso, identificou que a pontuação de crédito com o uso de *Big Data* é uma temática nova, apresentando interesse acadêmico crescente em comparação a outros temas de finanças. Por fim, as linhas de pesquisas estão bastante voltadas para a melhoria de decisões de empresas e instituições financeiras, inclusão social e a melhoria no desempenho dos modelos de pontuações de crédito.

Dória (2019) buscou criar variáveis a partir de dados de varejos, categorizando comportamentos inadimplentes e complementando sistemas de avaliação já existentes no mercado. Seus dados visaram a quantificação de parâmetros comportamentais variados, através da criação de *features*, o que possibilitou sua aplicação em modelos de classificações, na qual utilizou, como variável, dados de uma plataforma online de modelagem preditiva, o que

⁴⁹ Têm como única função orientar os fornecedores de produtos e serviços no que se refere à concessão de crédito no mercado de consumo, de forma que os respectivos associados mensurem os riscos envolvendo tal atividade. Como por exemplo: SPC, SERASA, CCF, CADIN etc.

permitiu comparar resultados através de métricas estatísticas, apontando quais apresentariam melhores rendimentos.

2.4.2 *Score de crédito versus LGPD*

Monteiro (2018) trouxe em seus estudos uma análise dos aspectos jurídicos do direito à explicação, no contexto de decisões automatizadas, que com sua força no cotidiano dos indivíduos é acompanhada de pouca transparência com relação ao seu funcionamento. Isso acaba por acarretar uma maior complexidade na identificação de práticas abusivas, discriminatórias ou, ainda, monopolísticas. Para mitigar tais efeitos, as legislações buscam assegurar à transparência a essas decisões automatizadas, concluindo que a Legislação Nacional, mas especificamente a Lei de Proteção de Dados Pessoais, foi além de outras regulações internacionais.

Villela (2019) buscou, em seu trabalho, apresentar os principais aspectos da Lei Geral de Proteção de Dados sob a ótica da atividade bancária e financeira, contextualizando a problemática e perfazendo um breve histórico da legislação, atinente ao tema de privacidade de dados pessoais e informações. Sua metodologia foi baseada no estudo e pesquisa de doutrinas, textos acerca da LGPD, Lei do Sigilo Bancário, Marco Civil da Internet, artigos e matérias jornalísticas.

Em seu trabalho, Sebben (2020) contextualizou a evolução dinâmica das metodologias de análise de crédito e o forte arcabouço legal, no que tange os riscos de créditos. As instituições financeiras, estando sujeitas a esses riscos, devem, na sequência, analisar o conceito de privacidade e as implicações de coleta, tratamento e armazenamento de dados pessoais. As empresas carecem em destacar uma perspectiva interpretativa que possa vir a atender uma harmonização das regras de proteção de dados.

De Freitas e Maffini (2020) trazem, em sua pesquisa, um trabalho com caráter exploratório, utilizando a pesquisa bibliográfica e documental, por meio de análise da lei, no que concerne à proteção do titular dos dados pessoais na concessão de crédito bancário, frente ao Cadastro Positivo. Os autores concluem que os artigos 7º e 20 da LGPD, juntamente com a possibilidade de solicitar revisões das decisões tomadas, podem contribuir para uma maior transparência na utilização e no tratamento dos dados pessoais dos titulares.

Polettini (2020) buscou, em seu artigo, impulsionar a importância da LGPD frente a uma mudança na vida dos cidadãos, uma vez que a lei é condição para o Brasil seguir dentro da economia global, cuja tendência aponta para os países desenvolvidos desenvolverem as suas leis de proteção de dados. Os autores fomentaram que a observância ao CDC e à LGPD são pressupostos essenciais para que as empresas evitem sanções que possam, inclusive, inviabilizar os negócios. Afirmam que, nesse sentido, o Direito do Consumidor ganhou um novo aliado. Por fim, avaliaram em que ponto a relação entre o mercado e o consumidor irá ser modificada com essa nova legislação e como irão garantir o consentimento para o uso dos dados pessoais e a privacidade.

2.4.3 LGPD versus *Big Data*

Cervantes e Rodrigues (2019) apresentaram conceitos, em seu estudo, por meio de análise da LGPD e do Regulamento Europeu 2016/679 do Parlamento Europeu e do Conselho de 27 de Abril de 2016 (GDPR), em conjunto com a doutrina nacional e estrangeira, já existente acerca do tema, analisando requisitos e alternativas que conduzam à exploração adequada da *Big Data* em um cenário de valorização e uma busca intensa pela proteção de dados pessoais na sociedade atual.

Guerra Martins, Jorgetto e Sutti (2019) analisam o quadro normativo da proteção ao direito constitucional à privacidade e sua possível violação em relação ao uso de *Big Data*, cujo tema foi confrontado no direito brasileiro e no direito europeu (GDPR), com incursões pelo Marco Civil da Internet e pela recente LGPD, visando estabelecer as obrigações das empresas que colhem dados pessoais dos indivíduos, buscando estabelecer direito aos titulares, bem como um sistema de responsabilidade administrativa e judicial.

Szinvelski, Arceno e Francisco (2019) abordaram, em seus estudos, o fenômeno da *Big Data* como efeito inevitável da quarta revolução industrial, levando em consideração o incremento da quantidade de dados pessoais compartilhados diariamente na internet. O autor discute a compatibilização entre a *Big Data* e a proteção de dados pessoais, trazendo uma reflexão de que não cabe ao Direito determinar o que a tecnologia deve ser, cujo desdobramento pode chegar ao ponto de impedir o desenvolvimento tecnológico, mas sim como uma ferramenta subsidiária, isto é, uma barreira frente à violação do uso indiscriminado dos dados pessoais sem o consentimento dos titulares.

Azambuja, Granville e Sarmiento (2020) também contextualizaram sobre a *Big Data*, com suas definições e características, abordando questões como a ética, privacidade, variedade e veracidade no ecossistema do Big Data. O autor discorre sobre os riscos e preocupações com a privacidade, a organização e segurança das informações, trazendo referências para entender melhor os processos e melhorias, para mitigar os impactos sociais decorrentes de Segurança Cibernética e o alinhamento com a LGPD.

2.4.4 *Score de crédito versus Big Data versus LGPD*

Nogueira (2020) aborda sobre a temática *Big Data* e suas características, bem como os indivíduos são monitorados por seus dispositivos do dia a dia como: *smartphones*, sites, cartões de crédito entre outros. Adicionando o debate sobre a LGPD, busca saber se a coleta, o armazenamento e a análise de milhares de cidadãos irão, de fato, servir para aumentar o acesso ao crédito e diminuir as taxas de juros. Por fim, trata o *Score* de Crédito e todo tipo de informação disponibilizada pelos cidadãos, com foco em estabelecer um *ranking* social e econômico.

Nogueira (2020) apresenta, em seu trabalho, o foco nas mudanças trazidas pela associação entre metodologias de informação e tecnologia, que possibilitaram alterações significativas para as sociedades industrializadas. O autor nos mostra como o controle social e a segmentação são potencializados pela combinação de dados e algoritmos, abrindo debate acerca da responsabilização dessas metodologias e os impactos da utilização de decisões automatizadas, nas quais afetam diretamente a sociedade contemporânea, levantando hipóteses de que levam a fomentar o crescimento das diferenças sociais e econômicas.

Este trabalho de dissertação, semelhante ao estudo de Nogueira (2020), também retrata aspectos acerca de *Big Data*, LGPD e *Credit Score*, confrontando tratamento de dados com o poder que o indivíduo tomador de crédito possui sobre seus próprios dados e o direito que lhe é assegurado em revisar as decisões automatizadas. Relaciona, também, as dificuldades para a obtenção de crédito daqueles que estão classificados como “maus pagadores”, bem como as dificuldades de transição entre as categorias.

Diferentemente da pesquisa desenvolvida por Nogueira (2020), cuja essência dogmática, através de uma corrente teórica, traz o debate para o crescimento das diferenças sociais e econômicas que surgem devido a categorização do indivíduo, este trabalho tenta

equilibrar a parte conceitual com a parte empírica, de modo a comparar as concepções teóricas com as observações do cotidiano dos atores⁵⁰ envolvidos, de forma a responder à pergunta de pesquisa proposta.

3 METODOLOGIA E BASE DE DADOS

Como apontado anteriormente, o presente estudo se propôs a realizar uma análise sobre como o sistema de *score* de crédito pode ser impactado pelo uso de *Big Data* e pela nova LGPD, bem como suas influências em relação aos tomadores de crédito e as empresas que operam com crédito. Inicialmente, para realização deste projeto, foi feito um levantamento de artigos acerca do tema abordado, cujas bases utilizadas para a pesquisa foram o *Google Scholar Metrics*, a *Scientific Electronic Library Online* (SciELO) e alguns livros científicos. Observou-se uma carência relativa ao tema, posto que as referências identificadas são posteriores ao ano de 2015.

A partir do levantamento bibliográfico realizado, foi conduzida uma *survey online*⁵¹, exploratória-descritiva com abordagem quantitativa, desenvolvida por meio da aplicação de questionário via *Google Forms*®, conforme Anexo – A, para armazenar as informações dos respondentes, cujo *link* para a enquete foi disponibilizado em redes sociais e aplicativos de trocas de mensagens, tais quais o *Facebook* e o *WhatsApp*. Esse tipo de estudo exploratório-descritivo permite ao pesquisador ter um aumento na experiência ao possibilitar explorar, conhecer e descrever possíveis situações por intermédio de dados subjetivos e situações desconhecidas (LEOPARDI, 2002).

As pesquisas online possuem suas devidas vantagens devido ao fácil acesso a um grupo populacional, otimizando o tempo do pesquisador, além da facilidade do entrevistado responder através de computadores, celulares e outros dispositivos, de forma oportuna (REGMI *et al.*, 2016). Flick (2013) abarca quatro principais vantagens da pesquisa online, conforme tabela 7:

⁵⁰ No contexto da *Big Data*, LGPD e Score de Crédito, os atores são: empresários, lojistas e consumidores.

⁵¹ O objetivo da *survey* (enquete) é a obtenção de informações quantitativas sobre um determinado grupo de pessoas. É mais indicada quando se deseja responder questões que expressem opiniões, costumes ou características de um determinado público-alvo.

Tabela 7: Principais vantagens da Pesquisa online

Baixo Custo	Questionários respondidos de forma online evitam gastos desnecessários com impressões e/ou envio/deslocamento para a coleta de dados.
Tempo	As respostas retornam mais rapidamente do que se fosse enviada pelos correios.
Facilidade	Pelo contexto em que as pessoas possuem cada vez mais acesso à internet e estão familiarizadas com os recursos tecnológicos.
Ausência de restrições espaciais	Acesso fácil para pessoas que estão geograficamente em outros locais.

Fonte: Autora adaptou do trabalho de Flick (2013).

Entretanto, de acordo com Mendes (2009), as principais desvantagens da pesquisa online são a incerteza sobre a validade dos dados que o respondente emite sobre as questões, além do baixo retorno e/ou devolução dos questionários.

Antes da coleta de dados propriamente, foi realizado um teste do questionário, cujo propósito foi verificar se a formulação e a apresentação das questões estavam em acordo com o propósito deste trabalho e se elas ofereceriam alguma dificuldade aos respondentes da pesquisa. Visto que o questionário de fato atendera corretamente as especificações para realização da pesquisa, esta foi aberta ao público durante o período de 13 de fevereiro de 2021 a 09 de maio de 2021, identificando o perfil sociodemográfico dos respondentes e como estão economicamente inseridos no contexto do presente estudo. Ao total, foram colhidas as respostas de quatrocentos e onze respondentes.

Durante a coleta de dados foram tomadas as devidas providências para garantir o anonimato e a privacidade dos respondentes. Nenhum dado sensível como nome, CPF ou qualquer outro dado que pudesse identificar o respondente foi utilizado.

Para a avaliação do perfil sociodemográfico, as variáveis utilizadas foram: sexo, etnia, faixa etária, escolaridade atual, estado civil e renda mensal bruta. Em relação à variável sexo, apenas as opções masculino e feminino foram consideradas para estudar se existiriam diferenças significativas entre essas opções, sabendo-se que a questão de gênero não é limitada a essas duas. A etnia foi classificada em cinco modalidades (branco, preto, pardo, amarelo e

indígena), conforme classificação adotada pelo PNAD⁵² – 2019. A faixa etária foi dividida em dez grupos (entre 18 anos e 24 anos, entre 25 anos e 29 anos, entre 30 anos e 34 anos, entre 35 anos e 39 anos, entre 40 anos e 44 anos, entre 45 anos e 49 anos, entre 50 anos e 54 anos, entre 55 anos e 59 anos, entre 60 anos e 64 anos e igual ou maior que 65 anos). A escolaridade foi dividida em oito níveis (fundamental, médio, superior e pós-graduação incompletos ou completos). O estado civil foi classificado em cinco modalidades (solteiro, casado, divorciado, união estável e viúvo). Por fim, a renda mensal bruta sofreu sete divisões (igual ou menor que dois salários-mínimos, entre dois e quatro salários-mínimos, entre quatro e seis salários-mínimos, entre seis e oito salários-mínimos, entre oito e dez salários-mínimos, entre dez e vinte salários-mínimos e maior que vinte salários-mínimos).

Para classificação dos dados coletados, sob o ponto de vista sociodemográfico, utilizou-se os procedimentos da estatística descritiva, na qual as variáveis foram avaliadas por meio das frequências absoluta e relativa (percentual). As respostas foram analisadas e organizadas em uma planilha no *Microsoft Excel*®, agrupadas por sua similaridade, divididas nos seguintes grupos:

- Todos os respondentes da amostra;
- Os respondentes que não possuem dívidas; e
- Os respondentes que possuem dívidas:
 - Somente dívidas pessoais;
 - Somente dívidas da empresa; e
 - Dívidas pessoais e dívidas da empresa.

Para a verificação dos dados coletados dos respondentes que assinalaram ser possuidores de dívidas, sob a ótica econométrica, a análise estatística inferencial foi efetuada com auxílio da *biblioteca R e do software RStudio for Windows*, cujas respostas foram analisadas e organizadas em uma planilha no *Microsoft Excel*®, agrupados, por sua similaridade, nos três grupos da categoria:

- Somente dívidas da empresa;
- Somente dívidas pessoais; e
- Dívidas da empresa e dívidas pessoais.

⁵² PNAD – Pesquisa Nacional por Amostra de Domicílios realizada pelo IBGE no ano de 2019.

Para a verificação dos lojistas/empresas, foram desenvolvidos quatro modelos que buscaram analisar: as relações no uso do *score* de crédito; o custo/benefício em sem usar o *score* de crédito; a prática da LGPD no uso do *score* de crédito; e as relações entre *Score* de Crédito, *Big Data*, LGPD e empréstimos/vendas. Após, na verificação dos tomadores de crédito, foram utilizados dois modelos, que são as relações, tanto no endividamento quanto na tomada de crédito - todos os modelos descritos foram elaborados pela autora.

Inicialmente, optou-se por avaliar o uso de *score* de crédito sob o ponto de vista da empresa, sendo para isso utilizadas algumas variáveis independentes, como a agilidade a ser percebida pela empresa na utilização do *score*, a assertividade em relação ao uso do *score* de crédito em relação ao empréstimo ou parcelamento que, porventura, foi liberado, a segurança que a empresa deduz na utilização do *score* de crédito, a credibilidade que a empresa percebe em relação ao *score* de crédito, a facilidade de utilização do crédito e o preço pago pelo serviço de *score* de crédito, como pode ser visto na equação 1.

$$Y_{Uso\ Score\ Cred} = \alpha + \beta_1 Agilidade + \beta_2 Assertividade + \beta_3 Seguran\c{c}a + \beta_4 Credibilidade + \beta_5 Facilidade + \beta_6 Pre\c{c}o\ do\ Servi\c{c}o + \varepsilon$$

Eq. (1)

O segundo modelo avaliou o custo/benefício observado pela empresa, verificando qual variável gera mais impacto. As variáveis independentes são o treinamento dos funcionários que a empresa deve fornecer, a redução da inadimplência que possa ter ocorrido na empresa e o aumento da venda parcelada, como observado na equação 2.

$$Y_{Custo\ x\ Benf} = \alpha + \beta_1 Treinamento + \beta_2 Red\ Inadm + \beta_3 Aumt\ Vend\ Parcel + \varepsilon$$

Eq. (2)

No terceiro modelo, foi analisada a prática da LGPD com relação a segurança dos dados cadastrados, o treinamento de funcionários, o investimento na segurança dos dados e sobre vazamento de dados, conforme representado na equação 3.

$$Y_{Pr\c{a}tica\ LGPD} = \alpha + \beta_1 Seg\ Dados\ Cad + \beta_2 Treinamento\ Func + \beta_3 Invest\ Seg\ Dados + \beta_4 Vazamt\ Dados + \varepsilon$$

Eq. (3)

Por fim, no quarto modelo, foi observado o que poderia influenciar o *score* de crédito, comparando-o com o uso da *Big Data* e o uso da LGPD, como visto na equação 4.

$$Y_{Score\ Cred} = \alpha + \beta_1 Big\ Data + \beta_2 LGPD + \varepsilon$$

Eq. (4)

Após a análise para com as empresas, os modelos abaixo fazem referência aos tomadores de crédito, analisando sobre a ótica do endividamento e da tomada de crédito.

Com relação ao modelo na equação 5, foi avaliado o que poderia potencializar o endividamento do tomador de crédito, isto é, o que estaria aumentando o seu endividamento. As variáveis independentes são: facilidade de crédito, parcelamento de contas, aumento salarial, quantidade de cartões de crédito, quantidade de contas bancárias, quantidade de dívidas ativas e se o nome não está negativado junto aos birôs de crédito.

$$Y_{Endividamento} = \alpha + \beta_1 Fac\ de\ Cred + \beta_2 Parcel\ de\ Cont + \beta_3 Aum\ Sal + \beta_4 Qnt\ Cart\ de\ Cred + \beta_5 Qnt\ Conta\ Banc + \beta_6 Qnt\ Dívida\ Ativa + \beta_7 Nome\ Limpo + \varepsilon$$

Eq. (5)

Finalizando a análise sobre os tomadores de crédito, o último modelo visto na equação 6 examinou o que influenciaria na tomada de crédito, sendo as variáveis independentes descritas como: o conhecimento que o tomador de crédito possui sobre o que é *score* de crédito, o entendimento acerca do cálculo do *score* de crédito, não se confundindo com o entendimento da metodologia específica, haja vista que o algoritmo está salvaguardado pelo segredo comercial; se há confiança nas empresas, ou seja, se ela é idônea; se há confiança na segurança dos dados pelos das empresas; a facilidade de obtenção de crédito; os valores parcelados e a consulta no Serasa/SPC.

$$Y_{Tomd\ de\ Cred} = \alpha + \beta_1 Conhe\ Score\ Cred + \beta_2 Entend\ Calc\ Score\ Cred + \beta_3 Conf\ Empresa + \beta_4 Conf\ Seg\ Dados + \beta_5 Fac\ de\ Cred + \beta_6 Valores\ Parcel + \beta_7 Consulta\ Serasa/SPC + \varepsilon$$

Eq. (6)

Conforme os estudos de Figueiredo (2017) e Santos (2020), para a elaboração de modelos simplificados, ou seja, com menos variáveis independentes, cujo objetivo é comparar

com os modelos originais (complexos) visando testar o princípio da parcimônia⁵³, as análises foram realizadas através do teste qui-quadrado (utilizando o teste exato de *Fisher* quando necessário) para verificar possíveis associações entre variáveis categóricas dicotômicas. Dessas verificações, selecionou-se como covariáveis de interesse aquelas cujo p-valor foi menor ou igual a 0,20. Em seguida, foram realizadas análises por meio de regressões logísticas simples para cada uma das preditoras que atenderam os requisitos dos testes de associação. Essas regressões logísticas simples, cujas variáveis independentes tiveram uma significância menor ou igual a 0,10, compuseram os modelos (reduzidos) de regressão logística múltipla para a comparação, quando possível, com os modelos originais propostos.

Tanto nos modelos originais quanto nos modelos reduzidos de regressão logística múltipla, o método de seleção das variáveis independentes foi o método *enter*, na qual todas as variáveis independentes são inseridas ao mesmo tempo. Essa escolha se justifica pelo fato de que, nos modelos complexos, o objetivo era testar se todas as preditoras estimariam os parâmetros dos modelos. Já para os modelos simplificados, como as covariáveis foram submetidas a um processo de seleção, o objetivo foi testar se todas essas variáveis independentes selecionadas também estimariam os parâmetros.

3.1 Regressão Logística

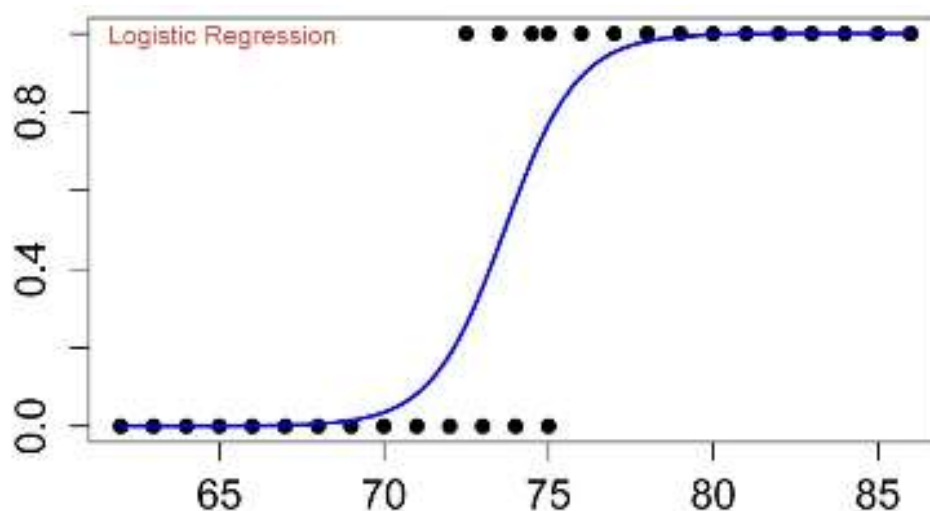
Conforme apontado no trabalho de Rosa (2000) a análise de regressão logística é a técnica mais utilizada para o desenvolvimento de modelos de *credit scoring*. Enquanto na regressão logística simples apenas uma covariável é considerada no modelo, na regressão múltipla a variável dependente é tratada em função de várias variáveis independentes. Esta técnica é indicada para casos nos quais a variável dependente é de natureza dicotômica ou binária. Quanto as preditoras, tanto podem ser categóricas ou não. Sendo assim, podemos entender a regressão logística como o análogo da regressão linear, para problemas que envolvam classificação.

A regressão logística avalia a probabilidade de um evento ocorrer, podendo estar entre 0 e 1. Em vez de acharmos a reta que melhor se ajusta aos dados, como na regressão linear, a

⁵³ Princípio da parcimônia – possui características desejáveis, tais quais a baixa dependência dos dados observados e um menor erro padrão. Essas características geram uma estabilidade numérica suficiente para tornar o modelo altamente generalizável

relação entre as variáveis independentes e a variável dependente se assemelha a uma curva em forma de S, que melhor se adequa aos dados. Esta curva é chamada de curva *sigmoid*, conforme mostra a Figura 6.

Figura 6: Curva *Sigmoid* - Regressão Logística



Fonte: Página de Mukesh Chaudhary no medium.com⁵⁴

Os coeficientes $\beta_0, \beta_1, \dots, \beta_n$ da regressão logística são estimados a partir do conjunto de dados, pelo método da máxima verossimilhança, no qual encontra uma combinação de coeficientes que maximiza a probabilidade da amostra ter sido observada. Do ponto de vista da interpretação, um coeficiente positivo aumenta a probabilidade e um coeficiente negativo diminui a probabilidade.

Contudo, para uma interpretação mais precisa da relação entre as covariáveis e a variável dependente é necessário que seja efetuada a exponenciação dos coeficientes, haja vista que estes estão na forma logarítmica. O valor dessa exponenciação é chamado de razão de chances (OR – *odds ratio* em inglês) e significa a chance de sucesso de uma categoria em relação à categoria de referência. Se a razão de chances for maior que 1, isto equivale a dizer que o aumento de uma unidade da variável explicativa aumenta em $[(OR - 1) * 100]\%$ a variável dependente em pertencer a categoria. Se a razão for menor que 1, isto significa que o aumento de uma unidade da variável explicativa diminui em $[(OR - 1) * 100]\%$ a variável dependente em pertencer a categoria.

⁵⁴ Disponível em: <<https://medium.com/@cmukesh8688/logistic-regression-sigmoid-function-and-threshold-b37b82a4cd79>>. Acesso em: 23 jun. 2021

3.2 Testes de Robustez

Tão importante quanto determinar se um dado sistema comporta-se de acordo com a especificação – objetivo do teste de conformidade – é definir se o sistema reage de maneira admissível fora do contexto especificado, ou seja, na presença de entradas inválidas ou sob condições estressantes. Sendo assim, a realização de testes de robustez visa verificar todo o comportamento do sistema fora da conjuntura descrita na especificação, tornando-se um passo comum para a validação de modelos (LEMOS, 2012).

3.2.1 Teste de Correlação

Segundo Gujarati (2011) a análise de correlação tem por objetivo medir a força ou o grau de associação entre duas variáveis através do coeficiente de correlação. Para Hoffmann (2016), esta análise procura determinar o grau de relacionamento entre duas variáveis.

3.2.1.1 Teste de Qui-Quadrado de Independência

O teste qui-quadrado é um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis categóricas nominais e avaliar a associação existente. A estatística qui-quadrado testa a significância da associação observada em uma tabulação cruzada. Calculam-se as frequências esperadas como se não houvesse associação nenhuma entre as variáveis (MALHOTRA, 2019).

O teste avalia se duas variáveis dispostas em uma tabela de contingência são dependentes ou independentes. A partir dos totais marginais observados na tabela, calculam-se os valores esperados em cada casela, sendo apresentada a estatística de teste calculada, o número de graus de liberdade e o nível de significância do teste. Contudo, alguns pressupostos devem ser observados:

- No caso de tabelas de contingência 2 x 2, usa-se a correção de *Yates*. Além disso, sempre que n for menor ou igual a 20 ou as frequências esperadas forem inferiores a 5, deve-se recorrer ao teste exato de Fisher, no qual fornece valores exatos para os p-valores do teste.
- Se n for maior que 20 e o número de categorias for maior que 2, não deverá existir mais do que 20% das células com frequências esperadas inferiores a 5 e nem deverá existir nenhuma célula com frequência esperada inferior a 1.

3.2.1.2 Teste Exato de Fisher

O Teste Exato de *Fisher* é utilizado em tabelas de contingência 2 x 2 e tem como objetivo testar se a variável da linha e a variável da coluna são independentes. Além disso, o p-valor fornecido pelo teste é exato, não exigindo técnica de aproximação. Este valor é preciso para todos os tamanhos amostrais, enquanto os resultados provenientes do teste qui-quadrado, que examina as mesmas hipóteses, podem ser imprecisos quando o número de células é pequeno. Ademais, o teste exato de *Fisher* é baseado na distribuição hipergeométrica. (CONTADOR; SENNE, 2016)

3.2.2 Teste de Multicolinearidade

A multicolinearidade é a correlação das variáveis independentes que possuem relações lineares exatas ou aproximadamente exatas, tendo seu indício de existência mais claro quando R^2 é bastante elevado (SARTORIS, 2003). Entretanto, nenhum dos coeficientes de regressão são estatisticamente significativos.

3.2.3 Análise de variância ANOVA

Para verificar as diferenças entre as estimativas médias em diferentes níveis de instrução (se são significativas ou não) serão aplicados o teste de hipóteses, denominado análise de variância ANOVA. Seu método consiste em testar a igualdade de duas ou mais médias, baseado na análise de variâncias amostrais (TRIOLA, 2005).

Sendo assim, se o valor obtido pela equação da estatística de teste for maior do que o valor crítico tabelado, referente à estatística teórica, rejeita-se a hipótese nula. Isto indicará que existe pelo menos uma diferença significativa entre as médias que estão sendo comparadas (TIMOSSI *ET AL.* 2010).

3.2.4 Causalidade de Granger

Para identificar as relações: 1) do uso do *score* de crédito e seu custo/benefício; 2) a prática da LGPD no uso do *score* de crédito e, 3) as relações entre *Score* de Crédito, *Big Data*, LGPD, empréstimos/vendas, tomadores de crédito e as associações, tanto no endividamento quanto na tomada de crédito, utilizou-se o teste de relação causal. Como exposto pela literatura, a correlação não implica, necessariamente, em causa (MATOS, 2000).

Além do mais, para Figueiredo Filho e Silva Junior (2009), não se aplica a divisão de causalidade simples ou recursiva por meio da correlação, ou seja, duvidosamente pode ser afirmado quem varia em função de quem. O teste de causalidade de Granger para Pereira *et al.* (2018) é uma técnica econométrica cuja função é estimar se uma variável X causa Y ou se Y causa X. Outrossim, pode-se concluir por meio do teste que X não causa Y ou que Y não causa X.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

4.1 Análises Sociodemográficas

4.1.1 Todos os Respondentes da Amostra

Na tabela 8 podemos verificar a caracterização sociodemográfica de toda a amostra, na qual teve 411 respondentes. Com relação aos quesitos elencados que tratam das condições sociodemográficas, a maioria respondente foi do sexo feminino, com 261 respondentes (63,50%); etnia branca, com 257 respondentes (62,53%); faixa etária entre 30 e 34 anos, com 67 respondentes (16,30%); estado civil casado, com 182 respondentes (44,28%); e com nível superior completo, com 106 respondentes (25,79%). A renda predominante, com 109 respondentes, foi igual ou menor que R\$ 2.200,00 (26,52%).

Tabela 8: Caracterização de todos os respondentes da amostra

Variáveis	n	%
Sexo		
Feminino	261	63,50
Masculino	150	36,50
Etnia		
Branco(a)	257	62,53
Negro(a)	53	12,90
Pardo(a)	93	22,63
Amarelo(a)	5	1,21
Indígena	3	0,73
Idade		
Entre 18 anos e 24 anos	51	12,41
Entre 25 anos e 29 anos	38	9,25
Entre 30 anos e 34 anos	67	16,30
Entre 35 anos e 39 anos	62	15,09
Entre 40 anos e 44 anos	47	11,44
Entre 45 anos e 49 anos	44	10,71
Entre 50 anos e 54 anos	26	6,33
Entre 55 anos e 59 anos	46	11,19
Entre 60 anos e 64 anos	17	4,14
Igual ou maior que 65 anos	13	3,16

Escolaridade		
Ensino fundamental incompleto	1	0,24
Ensino fundamental completo	5	1,22
Ensino médio incompleto	3	0,73
Ensino médio completo	39	9,49
Ensino superior incompleto	95	23,11
Ensino superior completo	106	25,79
Pós-graduação (MBA, Mestrado, Doutorado) incompleto	98	23,85
Pós-graduação (MBA, Mestrado, Doutorado) completo	64	15,57
Estado civil		
Solteiro(a)	159	38,69
Casado(a)	182	44,28
Divorciado(a)	28	6,81
União Estável	27	6,57
Viúvo(a)	15	3,65
Renda		
Igual ou menor que R\$ 2.200,00 (Igual ou menor que 2 SM ⁵⁵)	109	26,52
Entre R\$ 2.200,01 e R\$ 4.400,00 (Entre 2 e 4 SM)	87	21,17
Entre R\$ 4.400,01 e R\$ 6.600,00 (Entre 4 e 6 SM)	62	15,08
Entre R\$ 6.600,01 e R\$ 8.800,00 (Entre 6 e 8 SM)	46	11,19
Entre R\$ 8.800,01 e R\$ 11.000,00 (Entre 8 e 10 SM)	47	11,44
Entre R\$ 11.000,01 e R\$ 22.000,00 (Entre 10 e 20 SM)	47	11,44
Maior que R\$ 22.000,00 (Maior que 20 SM)	13	3,16

Fonte: Elaborado pela autora, 2021

Na tabela 9, como podemos notar, foi realizada uma divisão na amostra, sendo segmentada em duas categorias: os que possuem dívidas, com 283 respondentes (68,86%) e os que não possuem dívidas, com 128 respondentes (31,14%).

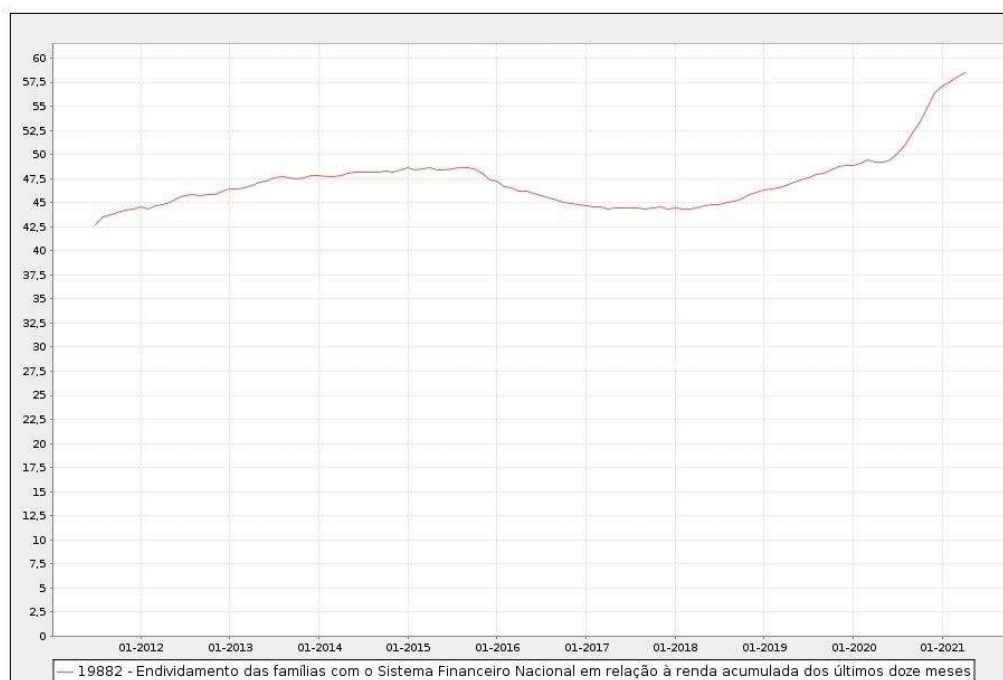
Tabela 9: Quantitativo de respondentes com dívidas e sem dívidas

Dívidas	n	%
Não	128	31,14
Sim	283	68,86

Fonte: Elaborado pela autora, 2021

Cabe ressaltar que os valores apresentados na tabela 9, acima, vão ao encontro dos resultados apresentados no gráfico 3, que representa a série histórica que estuda o endividamento das famílias com o Sistema Financeiro Nacional em relação à renda acumulada dos últimos doze meses. (BACEN, 2021)

⁵⁵ Salário-mínimo (SM) vigente em 2021 no Brasil = R\$ 1.100,00

Gráfico 3: Endividamento das famílias

Fonte: BACEN (2021)

4.1.2 Respondentes que Não Possuem Dívidas

A tabela 10 apresenta a caracterização dos respondentes que não possuem dívidas (128 respondentes de acordo com a tabela 9), divididos da seguinte forma (apresentando aqueles com os maiores percentuais): sexo feminino (63,41%), etnia branca (69,53%), idade entre a faixa dos 18 até os 24 anos (20,31%), escolaridade com pós-graduação completa (29,69%), estado civil solteiro (42,97%) e renda igual ou menor que R\$2.200,00 (28,91%).

Tabela 10: Caracterização dos respondentes que não possuem dívidas

Variáveis	<i>n</i>	%
Sexo		
Feminino	85	63,41
Masculino	43	36,59
Etnia		
Branco(a)	89	69,53
Negro(a)	9	7,03
Pardo(a)	29	22,66
Amarelo(a)	1	0,78
Indígena	0	0,00
Idade		
Entre 18 anos e 24 anos	26	20,31
Entre 25 anos e 29 anos	11	8,59
Entre 30 anos e 34 anos	21	16,41
Entre 35 anos e 39 anos	17	13,28
Entre 40 anos e 44 anos	14	10,94

Entre 45 anos e 49 anos	6	4,69
Entre 50 anos e 54 anos	8	6,25
Entre 55 anos e 59 anos	15	11,72
Entre 60 anos e 64 anos	6	4,69
Igual ou maior que 65 anos	4	3,13
Escolaridade		
Ensino fundamental incompleto	1	0,78
Ensino fundamental completo	1	0,78
Ensino médio incompleto	1	0,78
Ensino médio completo	14	10,94
Ensino superior incompleto	28	21,88
Ensino superior completo	25	19,53
Pós-graduação (MBA, Mestrado, Doutorado) incompleto	20	15,63
Pós-graduação (MBA, Mestrado, Doutorado) completo	38	29,69
Estado civil		
Solteiro(a)	55	42,97
Casado(a)	52	40,63
Divorciado(a)	8	6,25
União Estável	9	7,03
Viúvo(a)	4	3,13
Renda		
Igual ou menor que R\$ 2.200,00 (Igual ou menor que 2 SM ⁵⁶)	37	28,91
Entre R\$ 2.200,01 e R\$ 4.400,00 (Entre 2 e 4 SM)	31	24,22
Entre R\$ 4.400,01 e R\$ 6.600,00 (Entre 4 e 6 SM)	22	17,19
Entre R\$ 6.600,01 e R\$ 8.800,00 (Entre 6 e 8 SM)	14	10,94
Entre R\$ 8.800,01 e R\$ 11.000,00 (Entre 8 e 10 SM)	9	7,03
Entre R\$ 11.000,01 e R\$ 22.000,00 (Entre 10 e 20 SM)	15	11,72
Maior que R\$ 22.000,00 (Maior que 20 SM)	0	0,00

Fonte: Elaborado pela autora, 2021

Na tabela 11 podemos observar a outra face da divisão, ou seja, os respondentes que possuem dívidas (283 respondentes conforme tabela 9), sumarizados da seguinte forma (sendo apresentados aqueles com os maiores percentuais): sexo feminino, com 176 respondentes (62,19%), etnia branca, com 168 respondentes (59,36%), faixa etária entre 30 anos e 34 anos, com 46 respondentes (16,25%), escolaridade apresentando nível superior completo, com 81 respondentes (28,62%), estado civil casado, com 130 respondentes (45,94%) e renda igual ou menor que R\$2.200,00, com 72 respondentes (25,44%).

Tabela 11: Caracterização dos respondentes que possuem dívidas

Variáveis	n	%
Sexo		
Feminino	176	62,19
Masculino	107	37,81
Etnia		
Branco(a)	168	59,36
Negro(a)	44	15,55
Pardo(a)	64	22,61
Amarelo(a)	4	1,41

⁵⁶ Salário-mínimo (SM) vigente em 2021 no Brasil = R\$ 1.100,00

Indígena	3	1,06
Idade		
Entre 18 anos e 24 anos	25	8,83
Entre 25 anos e 29 anos	27	9,54
Entre 30 anos e 34 anos	46	16,25
Entre 35 anos e 39 anos	45	15,90
Entre 40 anos e 44 anos	33	11,66
Entre 45 anos e 49 anos	38	13,43
Entre 50 anos e 54 anos	18	6,36
Entre 55 anos e 59 anos	31	10,95
Entre 60 anos e 64 anos	11	3,89
Igual ou maior que 65 anos	9	3,18
Escolaridade		
Ensino fundamental incompleto	0	0,00
Ensino fundamental completo	4	1,41
Ensino médio incompleto	2	0,71
Ensino médio completo	25	8,83
Ensino superior incompleto	67	23,67
Ensino superior completo	81	28,62
Pós-graduação (MBA, Mestrado, Doutorado) incompleto	44	15,55
Pós-graduação (MBA, Mestrado, Doutorado) completo	60	21,20
Estado civil		
Solteiro(a)	104	36,75
Casado(a)	130	45,94
Divorciado(a)	20	7,07
União Estável	18	6,36
Viúvo(a)	11	3,89
Renda		
Igual ou menor que R\$ 2.200,00 (Igual ou menor que 2 SM ⁵⁷)	72	25,44
Entre R\$ 2.200,01 e R\$ 4.400,00 (Entre 2 e 4 SM)	56	19,79
Entre R\$ 4.400,01 e R\$ 6.600,00 (Entre 4 e 6 SM)	40	14,13
Entre R\$ 6.600,01 e R\$ 8.800,00 (Entre 6 e 8 SM)	32	11,31
Entre R\$ 8.800,01 e R\$ 11.000,00 (Entre 8 e 10 SM)	38	13,43
Entre R\$ 11.000,01 e R\$ 22.000,00 (Entre 10 e 20 SM)	32	11,31
Maior que R\$ 22.000,00 (Maior que 20 SM)	13	4,59

Fonte: Elaborado pela autora, 2021

4.1.3 Respondentes que Possuem Dívidas

Partindo para análises em um contexto socioeconômico, na tabela 12, o conjunto dos respondentes possuidores de dívidas (283 respondentes conforme tabela 9) foi seccionado em três grupos distintos: os respondentes que possuem apenas dívidas pessoais, com 178 respondentes (62,90%), os respondentes que possuem apenas dívidas da empresa, com 60 respondentes (21,20%) e, por último, os respondentes que estão inseridos nos dois grupos, isto é, os que possuem dívidas pessoais e dívidas da empresa, com 45 respondentes (15,90%).

⁵⁷ Salário-mínimo (SM) vigente em 2021 no Brasil = R\$ 1.100,00

Tabela 12: Classificação das dívidas dos respondentes da amostra

Dívidas	<i>n</i>	%
Apenas dívidas pessoais	178	62,90
Apenas dívidas da empresa	60	21,20
Dívidas pessoais e dívidas da empresa	45	15,90

Fonte: Elaborado pela autora, 2021

4.1.3.1 Respondentes que possuem apenas dívidas pessoais

Em relação aos respondentes que possuem apenas dívidas pessoais (178 respondentes conforme tabela 12), podemos verificar, na tabela 13, que a quantidade de contas bancárias e de cartões de créditos, na média, foi de 2,25 e 1,78, respectivamente. Os desvios padrões foram de 1,38 para a quantidade de contas bancárias e de 1,45 para a quantidade de cartões de crédito, demonstrando uma concentração em torno das médias. 97,75% dos respondentes acreditam que ter nome limpo é de suma importância; 69,66% já tinham ouvido falar em *score* de crédito; 69,96% já haviam buscado o entendimento de como é realizado o cálculo do *score*; e 70,74% não acham justo ocultar as informações utilizadas para o cálculo de *score* de crédito.

Tabela 13: Informações sobre tomada de crédito e endividamento dos respondentes que possuem apenas dívidas pessoais

Variáveis	<i>n</i>	%	\bar{X}	σ	min - máx.
Contas Bancárias					
Quantidade	-	-	2,25	1,38	0 – 10
Cartões de Crédito					
Quantidade	-	-	1,78	1,45	0 – 10
Ter nome limpo é importante					
Sim	174	97,75	-	-	-
Não	4	2,25	-	-	-
Já ouviu falar em <i>Score de Crédito</i>					
Sim	124	69,66	-	-	-
Não	54	30,34	-	-	-
Já buscou verificar como se calcula o <i>Score de Crédito</i>					
Sim	52	29,21	-	-	-
Não	126	70,79	-	-	-
Acha justo ocultar informações utilizadas no <i>Score de Crédito</i>					
Sim	52	29,21	-	-	-
Não	126	70,79	-	-	-

Fonte: Elaborado pela autora, 2021

4.1.3.2 Respondentes que possuem apenas dívidas da empresa

Na tabela 14 podemos analisar as informações sobre os respondentes que possuem dívidas apenas da empresa (60 respondentes conforme tabela 12). Cada respondente possui, na média, 2,87 contas bancárias e 2,42 cartões de crédito. Os desvios padrões foram de 1,37 para

a quantidade de contas bancárias e de 1,24 para a quantidade de cartões de crédito, demonstrando, também, uma concentração em torno das médias. Para 91,67% dos respondentes, possuir nome limpo se faz importante. 93,33% dos respondentes já ouviram falar em *score* de crédito e 65% já realizaram alguma consulta ao *score*. Enquanto 81,67% dos respondentes nunca buscaram verificar como se calcula o *score* de crédito, 98,33% não acham justo ocultar as informações utilizadas no *score*. Perguntados se já começaram a adequar suas empresas à LGPD, 61,67% responderam que sim. Com relação ao vazamento de dados na empresa, 46,67% já tiveram alguma ocorrência nesse sentido. Por fim, 61,67% dos respondentes afirmaram utilizar *Big Data* para solucionar as questões que envolvam *score* de crédito.

Tabela 14: Informações sobre tomada de crédito e endividamento dos respondentes que possuem apenas dívidas da empresa

Variáveis	<i>n</i>	%	\bar{X}	σ	min - máx.
Contas Bancárias					
Quantidade	-	-	2,87	1,37	0 – 10
Cartões de Crédito					
Quantidade	-	-	2,42	1,24	0 – 10
Ter nome limpo é importante					
Sim	55	91,67	-	-	-
Não	5	8,33	-	-	-
Já ouviu falar em <i>Score de Crédito</i>					
Sim	56	93,33	-	-	-
Não	4	6,67	-	-	-
Faz alguma consulta ao <i>Score de Crédito</i>					
Sim	39	65,00	-	-	-
Não	21	35,00	-	-	-
Já buscou verificar como se calcula o <i>Score de Crédito</i>					
Sim	11	18,33	-	-	-
Não	49	81,67	-	-	-
Acha justo ocultar informações utilizadas no <i>Score de Crédito</i>					
Sim	1	1,67	-	-	-
Não	59	98,33	-	-	-
Já começou a adequar a empresa à LGPD					
Sim	37	61,67	-	-	-
Não	23	38,33	-	-	-
Já houve algum vazamento de dados na empresa					
Sim	28	46,67	-	-	-
Não	32	53,33	-	-	-
A empresa utiliza <i>Big Data</i> para questões que envolvem <i>Score de Crédito</i>					
Sim	37	61,67	-	-	-
Não	23	38,33	-	-	-

Fonte: Elaborado pela autora, 2021

Para os respondentes que assinalaram já ter realizado alguma consulta ao *score* de crédito (39 respondentes conforme exibido na tabela 14) foi realizado um questionamento

acerca do que mais lhe parecia interessante no uso do *score*, tendo como opções de resposta os seguintes elementos: a agilidade, a assertividade, a segurança proporcionada, a credibilidade proporcionada, a facilidade e o preço do serviço. Os resultados podem ser vistos na tabela 15, sendo a credibilidade proporcionada o elemento mais assinalado, com 74,36%.

Tabela 15: Elementos que podem influenciar positivamente o uso do *score* de crédito para os respondentes que possuem apenas dívidas da empresa

Variáveis	n	%
Agilidade	12	30,77
Assertividade	23	58,97
Segurança proporcionada	24	61,54
Credibilidade proporcionada	29	74,36
Facilidade	9	23,08
Preço do serviço	2	5,13

Fonte: Elaborado pela autora, 2021.

Sobre os respondentes que assinalaram já terem começado a adequar a empresa à LGPD (37 respondentes conforme exibido na tabela 14), foi realizado um questionamento sobre quais fatores seriam influenciados pelo uso das boas práticas da LGPD. Os resultados podem ser vistos na tabela 16, na qual os respondentes afirmaram que o investimento em segurança de dados é o fator mais influenciado, com 81,08%, seguidos pelo uso de segurança para o cadastramento e armazenamento de dados, com 54,05%, do treinamento de pessoas sobre o uso das boas práticas da LGPD, com 48,65% e de problemas com vazamentos de dados, com 40,54%.

Tabela 16: Fatores influenciados positivamente pelo uso das boas práticas da LGPD para os respondentes que possuem apenas dívidas da empresa

Variáveis	n	%
Uso de segurança para o cadastramento e armazenamento de dados	20	54,05
Treinamento de pessoas sobre o uso das boas práticas da LGPD	18	48,65
Investimentos em segurança de dados	30	81,08
Problemas com vazamentos de dados	15	40,54

Fonte: Elaborado pela autora, 2021

4.1.3.3 Respondentes que possuem dívidas pessoais e dívidas da empresa

Na tabela 17 observamos as características dos respondentes que possuem dívidas pessoais e dívidas da empresa (45 respondentes conforme tabela 12). A quantidade de contas bancárias e de cartões de créditos, na média, foi de 3,18 e 2,96, respectivamente. Os desvios padrões foram de 1,45 para a quantidade de contas bancárias e de 1,66 para a quantidade de cartões de crédito, demonstrando uma concentração em torno das médias. Para 88,89% dos

respondentes, possuir nome limpo é importante. 93,33% dos respondentes já ouviram falar em *score* de crédito e 68,89% já realizaram alguma consulta ao *score*. Enquanto 60% dos respondentes nunca buscaram verificar como se calcula o *score* de crédito, 91,11% não acham justo ocultar as informações utilizadas no *score*. 53,33% dos respondentes já começaram a adequar suas empresas à LGPD. 40% afirmaram que já tiveram algum vazamento de dados na empresa. Finalizando, 53,33% dos respondentes afirmaram utilizar *Big Data* para solucionar as questões que envolvam *score* de crédito.

Tabela 17: Informações sobre tomada de crédito e endividamento dos respondentes que possuem dívidas pessoais e dívidas da empresa

Variáveis	<i>n</i>	%	\bar{X}	σ	min - máx.
Contas Bancárias					
Quantidade	-	-	3,18	1,45	0 – 10
Cartões de Crédito					
Quantidade	-	-	2,96	1,66	0 – 10
Ter nome limpo é importante					
Sim	40	88,89	-	-	-
Não	5	11,11	-	-	-
Já ouviu falar em <i>Score de Crédito</i>					
Sim	42	93,33	-	-	-
Não	3	6,67	-	-	-
Faz alguma consulta ao <i>Score de Crédito</i>					
Sim	31	68,89	-	-	-
Não	14	31,11	-	-	-
Já buscou verificar como se calcula o <i>Score de Crédito</i>					
Sim	18	40,00	-	-	-
Não	27	60,00	-	-	-
Acha justo ocultar informações utilizadas no <i>Score de Crédito</i>					
Sim	4	8,89	-	-	-
Não	41	91,11	-	-	-
Já começou a adequar a empresa à LGPD					
Sim	24	53,33	-	-	-
Não	21	46,67	-	-	-
Já houve algum vazamento de dados na empresa					
Sim	18	40,00	-	-	-
Não	27	60,00	-	-	-
A empresa utiliza <i>Big Data</i> para questões que envolvem <i>Score de Crédito</i>					
Sim	24	53,33	-	-	-
Não	21	46,67	-	-	-

Fonte: Elaborado pela autora, 2021.

Para os respondentes que assinalaram já ter realizado alguma consulta ao *score* de crédito (31 respondentes de acordo com a tabela 17), foi realizado um questionamento acerca do que mais lhe parecia interessante no uso, tendo como opções de resposta a agilidade, a assertividade, a segurança proporcionada, a credibilidade proporcionada, a facilidade e o preço

do serviço. Os resultados podem ser vistos na tabela 18, sendo a credibilidade proporcionada o elemento mais assinalado, com 74,19%.

Tabela 18: Elementos que podem influenciar positivamente o uso do *score* de crédito para os respondentes que possuem dívidas pessoais e dívidas da empresa

Variáveis	<i>n</i>	%
Agilidade	8	25,81
Assertividade	21	67,74
Segurança proporcionada	12	38,71
Credibilidade proporcionada	23	74,19
Facilidade	4	12,90
Preço do serviço	2	6,45

Fonte: Elaborado pela autora, 2021.

Sobre os respondentes que assinalaram já terem começado a adequar a empresa à LGPD (24 respondentes conforme exibido na tabela 17), foi realizado um questionamento sobre quais fatores seriam influenciados pelo uso das boas práticas da LGPD. Os resultados podem ser vistos na tabela 19, na qual os respondentes afirmaram que o investimento em segurança de dados é o fator mais influenciado com 62,50%.

Tabela 19: Fatores influenciados positivamente pelo uso das boas práticas da LGPD para os respondentes que possuem dívidas pessoais e dívidas da empresa

Variáveis	<i>n</i>	%
Uso de segurança para o cadastramento e armazenamento de dados	12	50,00
Treinamento de pessoas sobre o uso das boas práticas da LGPD	11	45,83
Investimentos em segurança de dados	15	62,50
Problemas com vazamentos de dados	14	58,33

Fonte: Elaborado pela autora, 2021

4.2 Análises Econométricas

4.2.1 Somente Dívidas da Empresa

4.2.1.1 Modelo 1: Avaliação do uso de *score* de crédito sob o ponto de vista da empresa

Com relação aos respondentes que assinalaram possuir apenas dívidas da empresa (60 respondentes de acordo com a tabela 12), foi realizado um questionamento acerca do que mais parecia ser ruim no uso do *score* de crédito, tendo como opções de escolha: a agilidade, a assertividade, a segurança proporcionada, a credibilidade proporcionada, a facilidade e o preço do serviço. Os resultados podem ser vistos na tabela 20, na qual os respondentes afirmaram que o preço do serviço, com 86,67%, é o elemento que mais parece ser ruim no uso do *score*.

Tabela 20: Elementos que podem influenciar negativamente o uso do *score* de crédito para os respondentes que possuem somente dívidas da empresa

Variáveis	n	%
Agilidade	5	8,33
Assertividade	8	13,33
Segurança	2	3,33
Credibilidade	7	11,67
Facilidade	4	6,67
Preço do serviço	52	86,67

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 1 para lojistas/empresas que possuem dívidas contraídas apenas por suas pessoas jurídicas, de acordo com a tabela 21, não foi possível obter observações conclusivas sobre as probabilidades relativas as covariáveis, uma vez que os seus p-valores não foram estatisticamente significativos.

Tabela 21: Coeficientes do modelo 1 referentes aos respondentes que possuem apenas dívidas da empresa

Coeficientes do Modelo 1				
$Y_{Uso\ Score\ Cred} = \alpha + \beta_1 Agilidade + \beta_2 Assertividade + \beta_3 Seguran\c{c}a + \beta_4 Credibilidade + \beta_5 Facilidade + \beta_6 Pre\c{c}o\ do\ Servi\c{c}o + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-17,4189	3675,4403	-0,005	0,996
Agilidade	36,5865	6090,2526	0,006	0,995
Assertividade	-0,8023	1,4531	-0,552	0,581
Segurança	19,1809	8352,6710	0,002	0,998
Credibilidade	-19,2394	3965,8962	-0,005	0,996
Facilidade	35,8525	5496,8995	0,007	0,995
Preço do Serviço	18,2212	3675,4403	0,005	0,996
Método: Enter	***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, que utiliza uma quantidade menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado, para os dados que versam sobre o que mais parece ser ruim no uso do *score* de crédito, são apresentadas, de forma compacta, na tabela 22. Enquanto a primeira linha apresenta a quantidade real de observações, a segunda linha apresenta a quantidade esperada de observações.

Tabela 22: Tabelas de contingência para os dados utilizados no modelo 1 para os respondentes que possuem somente dívidas da empresa

Usa_score	<u>Agilidade</u>		<u>Assertividade</u>		<u>Segurança</u>		<u>Credibilidade</u>		<u>Facilidade</u>		<u>Preço do Serviço</u>	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim	5	34	2	37	2	37	1	38	4	35	33	6
Não	0	21	6	15	0	21	6	15	0	21	19	2
(Real)												
Sim	3,3	35,8	5,2	33,8	1,3	37,7	4,6	34,5	2,6	36,4	33,8	5,2
Não	1,8	19,3	2,8	18,2	0,7	20,3	2,5	18,6	1,4	19,6	18,2	2,8
(Esp.)												

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 23, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Para as tabelas de contingência que apresentaram alguma casela com frequência esperada menor que 5, foram calculados, também, os p-valores pelo teste exato de *Fisher*, que neste caso, foram os valores considerados para a seleção das variáveis candidatas e para a verificação da significância estatística.

Apenas as variáveis *Agilidade*, *Assertividade* e *Credibilidade* apresentaram um p-valor menor que 0,20, sendo consideradas candidatas ao uso no modelo reduzido. Paralelo a isso, as variáveis *Assertividade* e *Credibilidade* foram significativas com um p-valor (*Fisher*) menor que 0,05, indicando que a variável *Uso Score Cred* possuiu uma associação de dependência com ambas, quando realizadas as análises bivariadas.

Tabela 23: Teste de qui-quadrado aplicado às variáveis do modelo 1 para os respondentes que possuem somente dívidas da empresa

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Agilidade	1,4985	1	0,2209	0,1519
Assertividade	4,6217	1	0,0315	* 0,0175 *
Segurança	0,0909	1	0,7630	0,5373
Credibilidade	6,6130	1	0,0101	* 0,0057 **
Facilidade	0,9536	1	0,3288	0,2873
Preço do Serviço	0,0570	1	0,8112	0,7007

**** p < 0.001; *** p < 0.01; ** p < 0.05; * p < 0.1

Fonte: Elaborado pela autora, 2021.

Os coeficientes das regressões logísticas simples das variáveis candidatas são mostrados na tabela 24. As predictoras *Assertividade* e *Credibilidade*, quando estudadas de maneira individual, tiveram uma significância menor que 0,10, sendo utilizadas, portanto, na construção do modelo reduzido. Além disso, ambas foram estatisticamente significativas. Podemos

observar, também, que os sinais dos seus coeficientes são negativos, indicando uma redução da probabilidade em usar o *score* de crédito.

Tabela 24: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 1 reduzido para os respondentes que possuem somente dívidas da empresa

Coeficientes da Regressão Logística Simples (Agilidade)					
$Y_{Uso\ Score\ Cred} = \alpha + \beta_{Agilidade} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	0,4818	0,2775	1,736	0,083	.
Agilidade	17,0842	1769,2577	0,010	0,992	
Coeficientes da Regressão Logística Simples (Assertividade)					
$Y_{Uso\ Score\ Cred} = \alpha + \beta_{Assertividade} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	0,9029	0,3061	2,950	0,003	**
Assertividade	-2,0015	0,8720	-2,295	0,022	*
Coeficientes da Regressão Logística Simples (Credibilidade)					
$Y_{Uso\ Score\ Cred} = \alpha + \beta_{Credibilidade} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	0,9295	0,3049	3,048	0,002	**
Credibilidade	-2,7213	1,1223	-2,425	0,015	*
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1				

Fonte: Elaborado pela autora, 2021.

Na tabela 25 observamos as razões de chances (OR – *odds ratio*) e os intervalos de confiança (I.C.), com nível de confiança de 95%, das regressões logísticas simples cujas preditoras candidatas apresentaram significância estatística. Observamos que, ao escolher a variável independente *Assertividade*, a probabilidade em usar o *score* de crédito diminui em 86%. De maneira similar, a escolha da variável independente *Credibilidade* diminui em 93% as chances de utilizar o *score*.

Tabela 25: Razão de chances e intervalo de confiança das regressões logísticas simples das preditoras candidatas ao modelo 1 reduzido para os respondentes que possuem somente dívidas da empresa

Razão de Chances da Regressão Logística Simples (Assertividade)			
$Y_{Uso\ Score\ Cred} = \alpha + \beta_{Assertividade} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	2,47	1,3538	4,4943
Assertividade	0,14	0,0247	0,7465
Razão de Chances da Regressão Logística Simples (Credibilidade)			
$Y_{Uso\ Score\ Cred} = \alpha + \beta_{Credibilidade} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%

(Intercepto)	2,53	1,3936	4,6052
Credibilidade	0,07	0,0073	0,5936

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística múltipla do modelo reduzido são mostrados na tabela 26. Tanto a variável *Assertividade* quanto a variável *Credibilidade* apresentaram um p-valor maior que 0,05, fazendo com que o modelo reduzido, de maneira semelhante ao modelo proposto, não possua significância estatística.

Tabela 26: Coeficientes da regressão logística múltipla do modelo 1 reduzido para os respondentes que possuem somente dívidas da empresa

Coeficientes do Modelo 1 (Reduzido)				
$Y_{Uso\ Score\ Cred} = \alpha + \beta_1\ Assertividade + \beta_2\ Credibilidade + \varepsilon$				
Variáveis	B	S.E.	z-valor	p-valor
(Intercepto)	0,9963	0,3161	3,152	0,002 **
Assertividade	-1,0182	1,0532	-0,967	0,334
Credibilidade	-2,1393	1,2502	-1,711	0,087 .
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Embora os modelos proposto e reduzido não possuam evidências suficientes para rejeitar a hipótese nula, podemos observar, no modelo reduzido, que o coeficiente da variável *Credibilidade* possui sinal negativo, podendo contribuir de alguma forma, no caso de existir uma significância econômica ou prática⁵⁸, para a diminuição da probabilidade em se usar o *score* de crédito.

4.2.1.2 Modelo 2: Custo-benefício observado pela empresa no uso do *score* de crédito

Com relação aos respondentes que assinalaram já ter realizado alguma consulta ao *score* de crédito (39 respondentes conforme tabela 14), tendo em vista o custo-benefício no que diz respeito a utilização do *score*, foi indagado se era justo o valor pago pelo uso do serviço. Na sequência, foram realizados os seguintes questionamentos:

- Teve grandes gastos com treinamento de pessoal para usar o *score* de crédito?
- Desde que começou a utilizar o *score* de crédito, notou alguma redução na sua inadimplência?

⁵⁸ Um resultado estatisticamente significante pode ter um efeito prático fraco ou vice-versa.

- Desde que começou a utilizar o *score* de crédito, notou algum aumento nas vendas parceladas?

Os resultados podem ser vistos na tabela 27. Não acham justo o valor pago pelo serviço, 24 respondentes (61,54%). Assinalaram que não tiveram grandes gastos com treinamento de pessoal, 26 respondentes (66,67%). Notaram alguma redução da sua inadimplência, 38 respondentes (97,44%). Notaram algum aumento nas vendas parceladas, 36 respondentes (92,31%).

Tabela 27: Custo-benefício em relação ao uso do *score* de crédito

Variáveis	n	%
Acha justo o valor pago pelo uso do serviço de <i>score</i> de crédito?		
Sim	15	38,46
Não	24	61,54
Teve grandes gastos com treinamento de pessoal para usar o <i>score</i> de crédito?		
Sim	13	33,33
Não	26	66,67
Desde que começou a utilizar o <i>score</i> de crédito, notou alguma redução na sua inadimplência?		
Sim	38	97,44
Não	1	2,56
Desde que começou a utilizar o <i>score</i> de crédito, notou algum aumento nas vendas parceladas?		
Sim	36	92,31
Não	3	7,69

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 2 para lojistas/empresas que possuem dívidas contraídas apenas por sua empresa, de acordo com a tabela 28, somente a covariável *Treinamento* foi estatisticamente significativa. Podemos ver que o seu coeficiente possui sinal negativo, contribuindo para diminuir a relação custo-benefício em usar o *score* de crédito.

Tabela 28: Coeficientes do modelo 2 referentes aos respondentes que possuem apenas dívidas da empresa

Coeficientes do Modelo 2					
$Y_{Custo \times Benf} = \alpha + \beta_1 Treinamento + \beta_2 Red \ Inadm + \beta_3 Aumt \ Vend \ Parcel + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	-0,1823	4568,2036	0,000	1,000	
Treinamento	-1,8871	0,8799	-2,145	0,032	*
Red Inadm	-17,3837	3956,1803	-0,004	0,996	
Aumt Vend Parcel	17,7484	2284,1018	0,008	0,994	
Método: Enter	*** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1				

Fonte: Elaborado pela autora, 2021.

Na tabela 29 vemos as razões de chances (OR – *odds ratio*) e os intervalos de confiança (I.C.) referentes ao modelo proposto. Observamos que o gasto com treinamento diminuiu em 85% o custo-benefício no uso do *score* de crédito.

Tabela 29: Razão de chances e intervalo de confiança da regressão logística do modelo 2 para os respondentes que possuem somente dívidas da empresa

Razão de Chances da Regressão Logística do Modelo 2			
$Y_{Custo \times Benf} = \alpha + \beta_1 Treinamento + \beta_2 Red \ Inadm + \beta_3 Aumt \ Vend \ Parcel + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,83	0,0000	Inf
Treinamento	0,15	0,0270	0,8500
Red Inadm	0,00	0,0000	Inf
Aumt Vend Parcel	51053774	0,0000	Inf

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, com menos variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado de independência são apresentadas, de forma compacta, na tabela 30, cuja primeira linha apresenta a quantidade real de observações e a segunda linha apresenta a quantidade esperada de observações.

Tabela 30: Tabelas de contingência para os dados utilizados no modelo 2 para os respondentes que possuem somente dívidas da empresa

Custo x Benf	<u>Treinamento</u>		<u>Red Inadm</u>		<u>Aumt Vend Parcel</u>	
	Sim	Não	Sim	Não	Sim	Não
Sim (Real)	2	13	14	1	15	0
Não (Real)	11	13	24	0	21	3
Sim (Esp.)	5	10	14,6	0,4	13,9	1,2
Não (Esp.)	8	16	23,4	0,6	22,2	1,9

Fonte: Elaborado pela autora, 2021.

Os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência são apresentados na tabela 31. Também podemos ver os p-valores calculados pelo teste exato de *Fisher*, para as tabelas de contingência que apresentaram alguma casela com frequência esperada menor que 5. Neste caso, os p-valores do teste exato de *Fisher* foram os valores considerados para a seleção das variáveis candidatas e para verificar a significância estatística.

Apenas a variável *Treinamento* apresentou um p-valor menor que 0,20, sendo a única utilizada no modelo reduzido. Nenhuma das variáveis obteve um p-valor menor que 0,05, indicando que a variável *Custo x Benf* não possui correlação com as demais variáveis, quando realizadas as análises bivariadas.

Tabela 31: Teste de qui-quadrado aplicado às variáveis do modelo 2 para os respondentes que possuem somente dívidas da empresa

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Treinamento	3,0469	1	0,0809	-
Red Inadm	0,0577	1	0,8101	0,3846
Aumt Vend Parcel	0,6523	1	0,4193	0,2713
‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1				

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística simples da única variável candidata e, por consequência, do modelo reduzido são mostrados na tabela 32. A variável independente *Treinamento*, além de ser estatisticamente significativa, apresentou coeficiente com sinal negativo, contribuindo para reduzir o custo-benefício ao usar o *score* de crédito.

Tabela 32: Coeficientes da regressão logística do modelo 2 reduzido para os respondentes que possuem somente dívidas da empresa

Coeficientes do Modelo 2 (Reduzido)				
$Y_{Uso\ Score\ Cred} = \alpha + \beta_1\ Treinamento + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	0,0000	0,3922	0,000	1.000
Treinamento	-1,7047	0,8630	-1,975	0,048 *
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Na tabela 33 observamos as razões de chances (OR – *odds ratio*), juntamente com os intervalos de confiança (I.C.), do modelo reduzido. Observamos que o gasto com treinamento diminui em 82% o efeito do custo-benefício no uso do *score* de crédito.

Tabela 33: Razão de chances e intervalo de confiança da regressão logística do modelo 2 reduzido para os respondentes que possuem somente dívidas da empresa

Razão de Chances da Regressão Logística do Modelo 2 (Reduzido)			
$Y_{UsoscoreCred} = \alpha + \beta_1 Treinamento + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	1,00	0,4636	2,1571
Treinamento	0,18	0,0335	0,9867

Fonte: Elaborado pela autora, 2021.

A tabela 34 apresenta um comparativo entre o modelo proposto e o modelo reduzido. Embora o pseudo- R^2 (*Nagelkerke*) tenha sido maior no modelo proposto em comparação com o modelo reduzido, os valores do AIC⁵⁹, do BIC⁶⁰ e da acurácia geral não apresentaram variações expressivas. Somado a isso, o p-valor do teste qui-quadrado, que foi maior que 0,05, nos mostra que não existem diferenças estatisticamente significativas entre os dois modelos.

Tabela 34: Comparativo entre o modelo 2 e o modelo 2 reduzido para os respondentes que possuem somente dívidas da empresa

Comparativo entre os Modelos						
Modelo	Pseudo- R^2 (<i>Nagelkerke</i>)	df	AIC	BIC	χ^2 p-valor	Acurácia Geral
Modelo 2	0,32	4	49,4788	56,1331	0,0571	0,69
Modelo 2 reduzido	0,16	2	51,2061	54,5332		0,62
*** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1						

Fonte: Elaborado pela autora, 2021.

Dado esses resultados, é possível aplicar o princípio da parcimônia, ou seja, podemos dizer que o modelo reduzido explica o evento de maneira mais simples, quando comparado ao modelo proposto. Podemos dizer que, no caso específico do modelo 2, o modelo com apenas o gasto com o treinamento de pessoal para utilizar o *score* de crédito consegue prever, de maneira similar ao modelo proposto (mais complexo), o custo-benefício observado pela empresa.

⁵⁹ Critério de Informação de Akaike (AIC).

⁶⁰ Critério de Informação de Bayesiano (BIC).

4.2.1.3 Modelo 3: Avaliação da prática da LGPD sob o ponto de vista da empresa

Com relação aos respondentes que assinalaram possuir apenas dívidas da empresa (60 respondentes de acordo com a tabela 12), foi realizado um questionamento sobre a adequação da empresa em relação à LGPD. Em seguida, foi realizado um questionamento acerca do que pode inibir o uso das boas práticas da LGPD, tendo como opções de escolha os seguintes fatores:

- Uso de segurança para o cadastramento e armazenamento de dados;
- Treinamento de pessoal sobre o uso das boas práticas da LGPD;
- Investimentos em segurança de dados (programas seguros, banco de dados, nuvem etc.); e
- Problemas com vazamentos de dados.

Os resultados podem ser vistos na tabela 35, na qual os respondentes afirmaram que o investimento em segurança de dados é o fator que mais inibe o uso, com 76,67%. Na sequência, temos: problemas com vazamentos de dados, com 30,00%; treinamento de pessoas sobre o uso das boas práticas da LGPD, com 26,67%; e o uso de segurança para o cadastramento e armazenamento de dados, com 13,33%.

Tabela 35: Fatores que inibem o uso das boas práticas da LGPD para os respondentes que possuem somente dívidas da empresa

Variáveis	<i>n</i>	%
Uso de segurança para o cadastramento e armazenamento de dados	8	13,33
Treinamento de pessoas sobre o uso das boas práticas da LGPD	16	26,67
Investimentos em segurança de dados	46	76,67
Problemas com vazamentos de dados	18	30,00

Fonte: Elaborado pela autora, 2021

Avaliando o modelo 3 para lojistas/empresas que possuem dívidas contraídas apenas por suas pessoas jurídicas, conforme tabela 36, não foi possível obter observações conclusivas sobre as probabilidades relativas as preditoras, uma vez que os seus p-valores não foram estatisticamente significativos. Contudo, os coeficientes das variáveis *Treinamento Func* e *Invest Seg Dados* possuem sinais negativos, podendo contribuir de alguma forma, no caso de

existir uma significância econômica ou prática⁶¹, para a diminuição da probabilidade em se usar a LGPD.

Tabela 36: Coeficientes do modelo 3 referentes aos respondentes que possuem apenas dívidas da empresa

Coeficientes do Modelo 3				
$Y_{Prática\ LGPD} = \alpha + \beta_1 Seg\ Dados\ Cad + \beta_2 Treinamento\ Func + \beta_3 Invest\ Seg\ Dados + \beta_4 Vazamt\ Dados + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	2,1678	1,0340	2,097	0.036 *
Seg Dados Cad	-0,6269	0,8688	-0,722	0,471
Treinamento Func	-1,0810	0,6524	-1,657	0,098 .
Invest Seg Dados	-1,6990	0,9725	-1,747	0,081 .
Vazamt Dados	0,3739	0,7729	0,484	0,629
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, que utiliza uma quantidade menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado são apresentadas, de forma compacta, na tabela 37, na qual a primeira linha apresenta a quantidade real de observações e a segunda linha apresenta a quantidade esperada de observações.

Tabela 37: Tabelas de contingência para os dados utilizados no modelo 3 para os respondentes que possuem somente dívidas da empresa

Usa_LGPD	Seg Dados Cad		Treinamento Func		Invest Seg Dados		Vazamt Dados	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim (Real)	4	33	7	30	25	12	14	23
Não (Real)	4	19	9	14	21	2	4	19
Sim (Esp.)	4,9	32,1	9,9	27,1	28,4	8,6	11,1	25,9
Não (Esp.)	3,1	19,9	6,1	16,9	17,6	5,4	6,9	16,1

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 38, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. O p-valor do teste exato de Fisher foi calculado apenas entre a variável *Prática LGPD* e a variável *Seg Dados Cad*, visto que essa tabela de contingência foi a única a apresentar caselas com a frequência esperada menor que 5 (conforme visto na tabela 37).

⁶¹ Um resultado estatisticamente significante pode ter um efeito prático fraco ou vice-versa.

As variáveis *Treinamento Func*, *Invest Seg Dados* e *Vazamt Dados* apresentaram um p-valor menor que 0,20, sendo consideradas candidatas ao uso no modelo reduzido. Contudo, nenhuma delas foi estatisticamente significativa, com um p-valor menor que 0,05, indicando que a variável *Prática LGPD* não possuiu correlação com as demais variáveis, quando realizadas as análises bivariadas.

Tabela 38: Teste de qui-quadrado aplicado às variáveis do modelo 3 para os respondentes que possuem somente dívidas da empresa

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Seg Dados Cad	0,1146	1	0,7350	0,4684
Treinamento Func	2,0194	1	0,1553	-
Invest Seg Dados	3,2389	1	0,0719	-
Vazamt Dados	1,9339	1	0,1643	-
****' p < 0.001; ***' p < 0.01; **' p < 0.05; '.' p < 0.1				

Fonte: Elaborado pela autora, 2021.

Os coeficientes das regressões logísticas simples das variáveis candidatas *Treinamento Func*, *Invest Seg Dados* e *Vazamt Dados* são mostrados na tabela 39. As predictoras *Treinamento Func* e *Invest Seg Dados*, quando estudadas individualmente, tiveram uma significância menor que 0,10, sendo utilizadas, portanto, na construção do modelo reduzido. Além disso, a variável independente *Invest Seg Dados* mostrou-se significativa estatisticamente, com um p-valor de 0,048. Podemos observar que os sinais dos coeficientes das variáveis *Treinamento Func* e *Invest Seg Dados* são negativos, reduzindo a probabilidade em usar a LGPD. Em contrapartida, embora não significativo estatisticamente, o coeficiente da variável *Vazamt Dados* é positivo, contribuindo para aumentar a probabilidade em usar a LGPD.

Tabela 39: Coeficientes das regressões logísticas simples das predictoras candidatas ao modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa

Coeficientes da Regressão Logística Simples (Treinamento)					
$Y_{Prática\ LGPD} = \alpha + \beta_{Treinamento\ Func} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	0,7621	0,3237	2,355	0,019	*
Treinamento Func	-1,0135	0,5989	-1,692	0,091	.
Coeficientes da Regressão Logística Simples (Investimentos)					
$Y_{Prática\ LGPD} = \alpha + \beta_{Invest\ Seg\ Dados} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	1,7918	0,7638	2,346	0,019	*
Invest Seg Dados	-1,6174	0,8191	-1,975	0,048	*
Coeficientes da Regressão Logística Simples (Vazamentos)					

$Y_{Prática\ LGPD} = \alpha + \beta_{Vazamt\ Dados} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	0,1911	0,3100	0,616	0,538
Vazamt Dados	1,0617	0,6462	1,643	0,100
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Na tabela 40 observamos as razões de chances (OR – *odds ratio*) e os intervalos de confiança (I.C.) das regressões logísticas simples cuja preditora candidata apresentou significância estatística. Observamos que o aumento de uma unidade da variável independente *Invest Seg Dados* diminui em 80% a probabilidade do uso das boas práticas da LGPD.

Tabela 40: Razão de chances e intervalo de confiança da regressão logística simples da preditora candidata, estatisticamente significante, ao modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa

Razão de Chances da Regressão Logística Simples (Investimentos)			
$Y_{Prática\ LGPD} = \alpha + \beta_{Invest\ Seg\ Dados} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	6,00	1,3429	26,8082
Invest Seg Dados	0,20	0,0399	0,9881

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística múltipla do modelo 3 reduzido são mostrados na tabela 41. Enquanto a variável *Treinamento Func* não se mostrou estatisticamente significativa, a variável *Invest Seg Dados* apresentou um p-valor de 0,039, com sinal negativo em seu coeficiente, indicando uma diminuição na probabilidade do uso das boas práticas da LGPD.

Tabela 41: Coeficientes da regressão logística múltipla do modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa

Coeficientes do Modelo 3 (Reduzido)					
$Y_{Prática\ LGPD} = \alpha + \beta_1_{Treinamento\ Func} + \beta_2_{Invest\ Seg\ Dados} + \varepsilon$					
Variáveis	B	S.E.	z-valor	p-valor	
(Intercepto)	2,2255	0,8356	2,663	0,008	**
Treinamento Func	-1,1602	0,6434	-1,803	0,071	.
Invest Seg Dados	-1,7491	0,8493	-2,059	0,039	*
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1				

Fonte: Elaborado pela autora, 2021.

Na tabela 42 observamos as razões de chances (OR – *odds ratio*), juntamente com os intervalos de confiança (I.C.) do modelo reduzido. Observamos que o fato de ter que investir na segurança dos dados diminui em 83% a probabilidade no uso da LGPD.

Tabela 42: Razão de chances e intervalo de confiança da regressão logística do modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa

Razão de Chances da Regressão Logística do Modelo 3 (Reduzido)			
$Y_{Prática\ LGPD} = \alpha + \beta_1 Treinamento\ Func + \beta_2 Invest\ Seg\ Dados + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	9,26	1,8001	47,6158
Treinamento Func	0,31	0,0888	1,1062
Invest Seg Dados	0,17	0,0330	0,9190

Fonte: Elaborado pela autora, 2021.

A tabela 43 apresenta um comparativo entre o modelo proposto e o modelo reduzido. Embora o pseudo- R^2 (*Nagelkerke*) tenha sido maior no modelo proposto em comparação com o modelo reduzido, os valores do AIC⁶², do BIC⁶³ e da acurácia geral não apresentaram variações expressivas. Paralelamente, o p-valor do teste qui-quadrado, que foi maior que 0,05, nos mostra que não existem diferenças estatisticamente significativas entre os dois modelos.

Tabela 43: Comparativo entre o modelo 3 e o modelo 3 reduzido para os respondentes que possuem somente dívidas da empresa

Comparativo entre os Modelos						
Modelo	Pseudo- R^2 (<i>Nagelkerke</i>)	df	AIC	BIC	χ^2 p-valor	Acurácia Geral
Modelo 3	0,19	5	80,6567	91,1283	0,6519	0,70
Modelo 3 reduzido	0,18	3	77,5124	83,7955		0,68
**** p < 0.001; *** p < 0.01; ** p < 0.05; . p < 0.1						

Fonte: Elaborado pela autora, 2021.

De posse desses resultados, é possível aplicar o princípio da parcimônia, ou seja, podemos dizer que o modelo reduzido explica o evento de maneira mais simples, quando comparado ao modelo proposto. Podemos dizer que, no caso específico do modelo 3, o modelo com apenas as covariáveis que avaliam o treinamento de pessoal sobre o uso das boas práticas da LGPD e os investimentos em segurança de dados (programas seguros, banco de dados,

⁶² Critério de Informação de Akaike (AIC).

⁶³ Critério de Informação de Bayesiano (BIC).

nuvem etc.) consegue prever, de maneira similar ao modelo proposto, que esses dois fatores inibem o uso das boas práticas da LGPD, embora a variável que versa sobre o treinamento não possua significância estatística.

4.2.1.4 Modelo 4: Relação entre uso de score de crédito, Big Data e LGPD, sob o ponto de vista da empresa

Com relação aos respondentes que assinalaram possuir apenas dívidas da empresa (60 respondentes de acordo com a tabela 12), foram realizados três questionamentos:

- Na sua empresa, você já fez alguma consulta ao *score* de crédito?
- Sua empresa utiliza *Big Data* para questões envolvendo o *score* de crédito?
- Você já começou a adequar a sua empresa em relação à LGPD?

Os resultados podem ser vistos na tabela 44, na qual 39 respondentes (65%) afirmaram que já fizeram alguma consulta ao *score* de crédito. 37 respondentes (61,67%) afirmaram que utilizam *Big Data* em questões envolvendo o *score*. 37 respondentes (61,67%) afirmaram que já começaram a adequar a empresa em relação à LGPD.

Tabela 44: Informações sobre *Big Data* e LGPD e suas influências no uso do *score* de crédito para os respondentes que possuem somente dívidas da empresa

Variáveis	<i>n</i>	%
Na sua empresa, você faz alguma consulta ao <i>score</i> de crédito?		
Sim	39	65,00
Não	21	35,00
Sua empresa utiliza <i>Big Data</i> para questões envolvendo o <i>score</i> de crédito?		
Sim	37	61,67
Não	23	38,33
Você já começou a adequar a sua empresa em relação à LGPD?		
Sim	37	61,67
Não	23	38,33

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 4 para lojistas/empresas que possuem somente dívidas da empresa, não foi possível obter observações conclusivas sobre as probabilidades relativas as preditoras, uma vez que os seus p-valores não foram estatisticamente significativos, conforme exibido na tabela 45.

Tabela 45: Coeficientes do modelo 4 referentes aos respondentes que possuem apenas dívidas da empresa

Coeficientes do Modelo 4				
$Y_{Score\ Cred} = \alpha + \beta_1 Big\ Data + \beta_2 LGPD + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,5390	0,4756	-1,133	0.257
Big Data	-17,5283	2223,8198	-0,008	0,994
LGPD	19,7901	2223,8198	0,009	0,993
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, que utiliza um número menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado são apresentadas, de forma compacta, na tabela 46. A primeira linha apresenta a quantidade real de observações e a segunda linha apresenta a quantidade esperada de observações.

Tabela 46: Tabelas de contingência para os dados utilizados no modelo 4 para os respondentes que possuem somente dívidas da empresa

Score Cred	<u>Big Data</u>		<u>LGPD</u>	
	Sim	Não	Sim	Não
Sim (Real)	28	11	32	7
Não (Real)	9	12	5	16
Sim (Esp.)	24,1	15	24,1	15
Não (Esp.)	13	8,1	13	8,1

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 47, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Os p-valores do teste exato de *Fisher* não foram calculados, visto que as tabelas de contingência não apresentaram células com a frequência esperada menor que 5.

Ambas as variáveis apresentaram um p-valor menor que 0,20, sendo consideradas candidatas ao uso no modelo reduzido. Paralelo a isso, a variável *LGPD* foi estatisticamente significativa, com um p-valor de 0,00003, indicando que a variável *Score Cred* possuiu uma associação de dependência, quando realizadas as análises bivariadas.

Tabela 47: Teste de qui-quadrado aplicado às variáveis do modelo 4 para os respondentes que possuem somente dívidas da empresa

Variáveis	qui-quadrado (χ^2)	df	p-valor		p-valor (Fisher)
Big Data	3,6887	1	0,0548	.	-
LGPD	17,2010	1	0,00003	***	-
‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1					

Fonte: Elaborado pela autora, 2021.

Os coeficientes das regressões logísticas simples das variáveis candidatas são mostrados na tabela 48. As preditoras *Big Data* e *LGPD*, quando estudadas de maneira individual, tiveram uma significância menor que 0,10, sendo utilizadas, portanto, na construção do modelo reduzido. Além disso, ambas foram significativas, apresentando sinais positivos nos coeficientes, indicando um aumento na probabilidade no uso do *score* de crédito.

Tabela 48: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem somente dívidas da empresa

Coeficientes da Regressão Logística Simples (Big Data)				
$Y_{Score\ Cred} = \alpha + \beta_{Big\ Data} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,0870	0,4174	-0,208	0.835
Big Data	1,2220	0,5666	2,157	0,031 *
Coeficientes da Regressão Logística Simples (LGPD)				
$Y_{Score\ Cred} = \alpha + \beta_{LGPD} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,8267	0,4532	-1,824	0.068 .
LGPD	2,6830	0,6608	4,060	0,00005 ***
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Na tabela 49 observamos as razões de chances (OR – *odds ratio*), juntamente com os intervalos de confiança (I.C.), das regressões logísticas que avaliaram, individualmente, as preditoras candidatas. Observamos que utilizar a *Big Data* aumenta em 239% a probabilidade do uso do *score* de crédito. De maneira similar, o uso da *LGPD* aumenta em 1363% as chances de utilizar o *score*.

Tabela 49: Razão de chances e intervalo de confiança das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem somente dívidas da empresa

Razão de Chances da Regressão Logística Simples (Big Data)			
$Y_{Score\ Cred} = \alpha + \beta_{Big\ Data} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,92	0,4045	2,0774
Big Data	3,39	1,1179	10,3043
Razão de Chances da Regressão Logística Simples (LGPD)			
$Y_{Score\ Cred} = \alpha + \beta_{LGPD} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,44	0,1800	1,0634
LGPD	14,63	4,0065	53,4126

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística múltipla do modelo reduzido são os mesmos do modelo proposto, haja vista que as covariáveis são as mesmas. Sendo assim, como os modelos são iguais, não é possível obter observações conclusivas. Contudo, as únicas informações sobre a influência da *Big Data* e da LGPD no uso do *score* de crédito são àquelas fornecidas através das análises das regressões logísticas simples e dos testes de associação bivariados.

4.2.2 Somente Dívidas Pessoais

4.2.2.1 Modelo 5: Avaliação do endividamento sob a ótica do tomador de crédito

Com relação aos respondentes que assinalaram possuir somente dívidas pessoais (178 respondentes de acordo com a tabela 12), foi questionado sobre quais elementos influenciariam no aumento das dívidas, tendo como opções de escolha: a facilidade de crédito, a possibilidade de poder parcelar contas, o aumento salarial, a quantidade de cartões de crédito, a quantidade de contas no banco, já possuir dívidas e ter o nome limpo.

Os resultados podem ser vistos na tabela 50, na qual os respondentes afirmaram que a quantidade de cartões de crédito, com 63,48%, é o elemento que mais interfere no aumento das dívidas. Na sequência, foram assinalados: a quantidade de contas no banco, com 58,99%; a facilidade de crédito, com 57,87%; a possibilidade de poder parcelar contas, com 53,37%; já possuir dívidas, com 37,08%; o aumento salarial, com 15,17%; e ter o nome limpo, com 10,67%.

Tabela 50: Elementos que podem influenciar no aumento das dívidas para os respondentes que possuem somente dívidas pessoais

Variáveis	n	%
Facilidade de crédito	103	57,87
Possibilidade de poder parcelar contas	95	53,37
Aumento salarial	27	15,17
Quantidade de cartões de crédito	113	63,48
Quantidade de contas no banco	105	58,99
Já possuir dívidas	66	37,08
Ter nome limpo	19	10,67

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 5 para os tomadores de crédito que possuem apenas dívidas pessoais, de acordo com a tabela 51, não foi possível obter observações conclusivas sobre as probabilidades relativas as covariáveis, uma vez que os p-valores não foram estatisticamente significativos.

Tabela 51: Coeficientes do modelo 5 referentes aos respondentes que possuem apenas dívidas pessoais

Coeficientes do Modelo 5				
$Y_{Endividamento} = \alpha + \beta_1 Fac\ de\ Cred + \beta_2 Parcel\ de\ Cont + \beta_3 Aum\ Sal + \beta_4 Qnt\ Cart\ de\ Cred + \beta_5 Qnt\ Conta\ Banc + \beta_6 Qnt\ Dívida\ Ativa + \beta_7 Nome\ Limpo + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	0,2081	0,3869	0,538	0,591
Fac de Cred	0,7006	0,3597	1,948	0,051
Parcel de Cont	-0,6092	0,8383	-0,727	0,467
Aum Sal	-0,1671	0,4742	-0,352	0,725
Qnt Cart de Cred	0,2545	0,3578	0,711	0,477
Qnt Conta Banc	0,4263	0,8599	0,496	0,620
Qnt Dívida Ativa	0,5059	0,3646	1,387	0,165
Nome Limpo	0,0111	0,5715	0,019	0,985
Método: Enter	*** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo 5 reduzido, que utiliza menos variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado são apresentadas, de forma compacta, na tabela 52, cuja linha que apresenta a expressão (*Real*) exibe a quantidade real de observações e a linha que apresenta a expressão (*Esp.*) exibe a quantidade esperada de observações.

Tabela 52: Tabelas de contingência para os dados utilizados no modelo 5 para os respondentes que possuem somente dívidas pessoais

Endividamento	<u>Fac de Cred</u>		<u>Parcel de Cont</u>		<u>Aum Sal</u>		<u>Qnt Cart Cred</u>		<u>Qnt Conta Banc</u>		<u>Qnt Dívida Ativa</u>	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim	77	47	66	58	18	106	81	43	74	50	49	75
Não	26	28	29	25	9	45	32	22	31	23	17	37
Sim	71,8	52,2	66,2	57,8	18,8	105,2	78,7	45,3	73,1	50,9	46	78
Não	31,2	22,8	28,8	25,2	8,2	45,8	34,3	19,7	31,9	22,1	20	34
	<u>Nome Limpo</u>											
	Sim	Não										
Sim	13	111										
Não	6	48										
Sim	13,2	110,8										
Não	5,8	48,2										

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 53, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Os p-valores do teste exato de Fisher não foram calculados, visto que as tabelas de contingência não apresentaram caselas com a frequência esperada menor que 5 (conforme apresentado na tabela 52).

Apenas a variável *Fac de Cred* apresentou um p-valor menor que 0,20, sendo considerada candidata ao uso no modelo reduzido. Nenhuma das variáveis obteve um p-valor menor que 0,05, indicando que a variável *Endividamento* não possuiu correlação com as demais variáveis, quando das análises bivariadas.

Tabela 53: Teste de qui-quadrado aplicado às variáveis do modelo 5 para os respondentes que possuem somente dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Fac de Cred	2,4571	1	0,1170	-
Parcel de Cont	0,0000	1	1,0000	-
Aum Sal	0,0197	1	0,8883	-
Qnt Cart Cred	0,3637	1	0,5465	-
Qnt Conta Banc	0,0138	1	0,9066	-
Qnt Dívida Ativa	0,7250	1	0,3945	-
Nome Limpo	0,0000	1	1,0000	-

*** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística simples da única variável candidata e, por consequência, do modelo reduzido são mostrados na tabela 54. A preditora *Fac de Cred*, apesar de possuir uma significância estatística maior que 0,05, apresentou coeficiente com sinal positivo, indicando um aumento nas chances de influenciar o aumento das dívidas.

Tabela 54: Coeficientes da regressão logística simples da preditora candidata ao modelo 5 reduzido para os respondentes que possuem somente dívidas pessoais

Coeficientes da Regressão Logística Simples (Facilidade de Crédito)				
$Y_{\text{Endividamento}} = \alpha + \beta_{\text{Fac de Cred}} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	0,5179	0,2387	2,170	0.030 *
Fac de Cred	0,5678	0,3293	1,724	0,085 .
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Embora os modelos proposto e reduzido não possuam evidências suficientes para rejeitar a hipótese nula, podemos observar, no modelo reduzido, que o coeficiente da variável *Fac de Cred* possui sinal positivo, indicando uma ampliação na probabilidade de influenciar o aumento das dívidas. Fazendo um comparativo, no modelo proposto, os coeficientes das variáveis *Parcel de Cont* e *Aum Sal* possuem sinais negativos, indicando uma possível atenuação no aumento das dívidas. Tais análises são apenas percepções, visto as considerações e os limites acerca da significância estatística e da significância econômica ou prática⁶⁴.

4.2.2.2 Modelo 6: Influência na formação do endividamento sob a ótica do tomador de crédito

Com relação aos respondentes que assinalaram possuir somente dívidas pessoais (178 respondentes de acordo com a tabela 12), foi inquirido quais fatores influenciariam na formação do endividamento, tendo como opções de escolha: a falta de conhecimentos sobre o *score* de crédito, a falta de conhecimentos sobre o cálculo do *score* de crédito, a confiança na empresa que vai parcelar a dívida/obter crédito, a confiança na segurança dos dados pessoais, a facilidade na obtenção de crédito, o valor das parcelas (possibilidade de pagamento do valor acordado) e ter o nome limpo.

Os resultados podem ser vistos na tabela 55, na qual os respondentes afirmaram que a facilidade de obtenção crédito, com 49,44%, é o fator que possui mais influência na formação

⁶⁴ Um resultado estatisticamente significante pode ter um efeito prático fraco ou vice-versa.

do endividamento. Na sequência, temos: o valor das parcelas (possibilidade de pagamento do valor acordado), com 46,07%; a falta de conhecimentos sobre o cálculo do *score* de crédito, com 35,39%; a falta de conhecimentos sobre o cálculo do *score* de crédito, com 23,03%; a confiança na empresa que vai parcelar a dívida/obter crédito, com 21,91%; a confiança na segurança dos dados pessoais, com 10,11%; e ter o nome limpo, com 10,11%.

Tabela 55: Fatores que podem influenciar na formação do endividamento para os respondentes que possuem somente dívidas pessoais

Variáveis	n	%
Falta de conhecimentos sobre o <i>score</i>	63	35,39
Falta de conhecimentos sobre o cálculo do <i>score</i>	41	23,03
Confiança na empresa que vai parcelar/obter	39	21,91
Confiança na segurança dos dados pessoais	18	10,11
Facilidade na obtenção de crédito	88	49,44
Valor das parcelas	82	46,07
Ter nome limpo	18	10,11

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 6 para os tomadores de crédito que possuem apenas dívidas pessoais, não foi possível obter observações conclusivas sobre as probabilidades relativas as variáveis independentes, conforme tabela 56, uma vez que os p-valores associados às covariáveis não foram estatisticamente significativos. Contudo, o coeficiente da variável *Entend Calc Score Cred* possui sinal positivo, podendo contribuir de alguma forma, para o aumento na influência na formação do endividamento.

Tabela 56: Coeficientes do modelo 6 referentes aos respondentes que possuem apenas dívidas pessoais

Coeficientes do Modelo 6					
$Y_{Tomada\ de\ Cred} = \alpha + \beta_1\ Conhec\ Score\ Cred + \beta_2\ Entend\ Calc\ Score\ Cred + \beta_3\ Conf\ Empresa + \beta_4\ Conf\ Seg\ Dados + \beta_5\ Fac\ de\ Cred + \beta_6\ Valores\ Parcel + \beta_7\ Consulta\ Serasa/SPC + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	-1,0080	0,3430	-2,939	0,003	**
Conhec Score Cred	0,2289	0,3823	0,599	0,549	
Entend Calc Score Cred	0,7956	0,4299	1,851	0,064	.
Conf Empresa	0,6036	0,4244	1,422	0,155	
Conf Seg Dados	-0,2982	0,5833	-0,511	0,609	
Fac de Cred	-0,2396	0,3542	-0,676	0,499	
Valores Parcel	-0,2644	0,3662	-0,722	0,470	
Consulta Serasa/SPC	-0,0010	0,5680	-0,002	0,999	
Método: Enter	***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1				

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo 6 reduzido, que utiliza uma quantidade menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado de independência são apresentadas, de forma compacta, na tabela 57, cuja linha que apresenta a expressão (*Real*) exibe a quantidade real de observações e a linha que apresenta a expressão (*Esp.*) exibe a quantidade esperada de observações.

Tabela 57: Tabelas de contingência para os dados utilizados no modelo 6 para os respondentes que possuem somente dívidas pessoais

Tomada de Crédito		<u>Conhec Score Cred</u>		<u>Entend Calc Score Cred</u>		<u>Conf Empresa</u>		<u>Conf Seg Dados</u>		<u>Fac de Cred</u>		<u>Valores Parcel</u>	
		Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim	(Real)	38	86	21	103	23	101	12	112	65	59	61	63
Não		25	29	20	34	16	38	6	48	23	31	21	33
Sim	(Esp.)	43,9	80,1	28,6	95,4	27,2	96,8	12,5	111,5	61,3	62,7	57,1	66,9
Não		19,1	34,9	12,4	41,6	11,8	42,2	5,5	48,5	26,7	27,3	24,9	29,1
		<u>Consulta Serasa/SPC</u>											
		Sim	Não										
Sim	(Real)	12	112										
Não		6	48										
Sim	(Esp.)	12,5	111,5										
Não		5,5	48,5										

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 58, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Os p-valores do teste exato de Fisher não foram calculados, visto que as tabelas de contingência não apresentaram células com a frequência esperada menor que 5.

Apenas a variáveis *Conhec Score Cred*, *Entend Calc Score Cred* e *Conf Empresa* apresentaram p-valores menores que 0,20, sendo consideradas candidatas ao uso no modelo reduzido. Ao mesmo tempo, a variável *Entend Calc Score Cred* foi significativa, com um p-valor de 0,0062, indicando que a variável *Tomd de Cred* possuiu uma associação de dependência com aquela, quando realizadas as análises bivariadas.

Tabela 58: Teste de qui-quadrado aplicado às variáveis do modelo 6 para os respondentes que possuem somente dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Conhec Score Cred	3,3745	1	0,0662	.
Entend Calc Score Cred	7,4777	1	0,0062	**

Conf Empresa	2,0910	1	0,1482	-
Conf Seg Dados	0,0005	1	0,9830	-
Fac de Cred	1,0867	1	0,2972	-
Valores Parcel	1,2197	1	0,2694	-
Consulta Serasa/SPC	0,0005	1	0,9830	-
‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1				

Fonte: Elaborado pela autora, 2021.

Os coeficientes das regressões logísticas simples das variáveis candidatas *Conhec Score Cred*, *Entend Calc Score Cred* e *Conf Empresa* são mostrados na tabela 59. As predictoras *Conhec Score Cred* e *Entend Calc Score Cred*, quando estudadas de forma individual, tiveram uma significância menor que 0,10, sendo utilizadas, portanto, na construção do modelo reduzido. Além disso, essas variáveis mostraram-se significativas estatisticamente, com p-valores de 0,046 e 0,004, respectivamente. Podemos observar que os sinais de seus coeficientes são positivos, contribuindo para a aumentar a influência na formação do endividamento.

Tabela 59: Coeficientes das regressões logísticas simples das predictoras candidatas ao modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais

Coeficientes da Regressão Logística Simples (Conhecimento do score)					
$Y_{Tomd\ de\ Cred} = \alpha + \beta_{Conhec\ Score\ Cred} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	-1,0871	0,2147	5,062	0,000 ⁶⁵	***
Conhec Score Cred	0,6683	0,3353	-1,993	0,046	*
Coeficientes da Regressão Logística Simples (Cálculo do score)					
$Y_{Tomd\ de\ Cred} = \alpha + \beta_{Entend\ Calc\ Score\ Cred} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	-1,1084	0,1978	5,604	0,000 ⁶⁶	***
Entend Calc Score Cred	1,0596	0,3698	-2,865	0,004	**
Coeficientes da Regressão Logística Simples (Confiança na empresa)					
$Y_{Tomd\ de\ Cred} = \alpha + \beta_{Conf\ Empresa} + \varepsilon$					
Variáveis	β	S.E.	z-valor	p-valor	
(Intercepto)	-0,9775	0,1903	5,137	0,000 ⁶⁷	
Conf Empresa	0,6146	0,3771	-1,630	0,103	
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1				

Fonte: Elaborado pela autora, 2021.

⁶⁵ p-valor = 0,000000414

⁶⁶ p-valor = 0,000000021

⁶⁷ p-valor = 0,000000028

Na tabela 60 observamos as razões de chances (OR – *odds ratio*) e os intervalos de confiança (I.C.) das regressões logísticas que avaliaram, individualmente, as predictoras candidatas que apresentaram significância estatística. Observamos que o aumento de uma unidade da variável independente *Conhec Score Cred* aumenta em 95% as chances na influência na formação do endividamento. De maneira similar, a falta de conhecimentos sobre o cálculo do *score* de crédito aumenta em 189% a probabilidade de influência na formação do endividamento.

Tabela 60: Razão de chances e intervalo de confiança das regressões logísticas simples das predictoras candidatas ao modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais

Razão de Chances da Regressão Logística Simples (Conhecimento do <i>score</i>)			
$Y_{Tomd\ de\ Cred} = \alpha + \beta_{Conhec\ Score\ Cred} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,34	0,2214	0,5137
Conhec Score Cred	1,95	1,0112	3,7641

Razão de Chances da Regressão Logística Simples (Cálculo do <i>score</i>)			
$Y_{Tomd\ de\ Cred} = \alpha + \beta_{Entend\ Calc\ Score\ Cred} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,33	0,2240	0,4864
Entend Calc Score Cred	2,89	1,3979	5,9556

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística múltipla do modelo reduzido são mostrados na tabela 61. A predictora *Conhec Score Cred*, além de possuir um p-valor de 0,025, apresentou coeficiente com sinal positivo, contribuindo para aumentar a influência na formação do endividamento.

Tabela 61: Coeficientes da regressão logística múltipla do modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais

Coeficientes do Modelo 6 (Reduzido)				
$Y_{Tomd\ de\ Cred} = \alpha + \beta_1\ Conhec\ Score\ Cred + \beta_2\ Entend\ Calc\ Score\ Cred + \varepsilon$				
Variáveis	B	S.E.	z-valor	p-valor
(Intercepto)	-1,1986	0,2239	-5,354	0,000 ⁶⁸ ***
Conhec Score Cred	0,3419	0,3723	0,918	0,359
Entend Calc Score Cred	0,9074	0,4042	2,245	0,025 *
Método: Enter	****' p < 0.001; ***' p < 0.01; '*' p < 0.05; '.' p < 0.1			

Fonte: Elaborado pela autora, 2021.

⁶⁸ p-valor = 0,0000000861

Na tabela 62 observamos as razões de chances (OR – *odds ratio*), juntamente com os intervalos de confiança (I.C.), da regressão logística do modelo reduzido. Observamos que a falta de conhecimentos sobre o cálculo do *score* aumenta em 148% a influência na formação do endividamento.

Tabela 62: Razão de chances e intervalo de confiança da regressão logística do modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais

Razão de Chances da Regressão Logística do Modelo 6 (Reduzido)			
$Y_{T\text{omd de Cred}} = \alpha + \beta_1 \text{Conhec Score Cred} + \beta_2 \text{Entend Calc Score Cred} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,30	0,1945	0,4677
Conhec Score Cred	1,41	0,6785	2,9199
Entend Calc Score Cred	2,48	1,1221	5,4720

Fonte: Elaborado pela autora, 2021.

A tabela 63 apresenta um comparativo entre o modelo proposto e o modelo reduzido. Embora o pseudo- R^2 (*Nagelkerke*) tenha sido maior no modelo proposto em comparação com o modelo reduzido, os valores do AIC⁶⁹, do BIC⁷⁰ e da acurácia geral não apresentaram variações expressivas. Paralelamente, o p-valor do teste qui-quadrado, que foi maior que 0,05, nos mostra que não existem diferenças estatisticamente significativas entre os dois modelos.

Tabela 63: Comparativo entre o modelo 6 e o modelo 6 reduzido para os respondentes que possuem somente dívidas pessoais

Comparativo entre os Modelos						
Modelo	Pseudo- R^2 (<i>Nagelkerke</i>)	df	AIC	BIC	χ^2 p-valor	Acurácia Geral
Modelo 6	0,09	8	222,8252	248,3395	0,7577	0,71
Modelo 6 reduzido	0,07	3	215,5096	225,0550		0,70

*** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1

Fonte: Elaborado pela autora, 2021.

De posse desses resultados, é possível aplicar o princípio da parcimônia, ou seja, podemos dizer que o modelo reduzido explica o evento de maneira mais simples, quando comparado ao modelo proposto. Podemos dizer que, no caso específico do modelo 6, o modelo com as covariáveis que avaliam a falta de conhecimentos sobre o *score* de crédito e a falta de

⁶⁹ Critério de Informação de Akaike (AIC).

⁷⁰ Critério de Informação de Bayesiano (BIC).

conhecimentos sobre o cálculo desse *score* consegue prever, de maneira similar ao modelo proposto, que esses dois fatores influenciam, positivamente, na formação do endividamento, embora a *falta de conhecimentos sobre o score de crédito* não possua significância estatística.

4.2.3 Dívidas da Empresa e Dívidas Pessoais

4.2.3.1 Modelo 1: Avaliação do uso de *score* de crédito sob o ponto de vista da empresa

Com relação aos respondentes que assinalaram possuir dívidas da empresa e dívidas pessoais (45 respondentes de acordo com a tabela 12), foi realizado um questionamento acerca do que mais parecia ser ruim no uso do *score* de crédito, tendo como opções de escolha: a agilidade, a assertividade, a segurança proporcionada, a credibilidade proporcionada, a facilidade e o preço do serviço. Os resultados podem ser vistos na tabela 64, na qual os respondentes afirmaram que o preço do serviço, com 71,11%, é o elemento que mais parece ser ruim no uso do *score*.

Tabela 64: Elementos que podem influenciar negativamente o uso do *score* de crédito para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	n	%
Agilidade	11	24,44
Assertividade	3	6,67
Segurança	2	4,44
Credibilidade	6	13,33
Facilidade	2	4,44
Preço do serviço	32	71,11

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 1 para lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, de acordo com a tabela 65, não foi possível obter observações conclusivas sobre as probabilidades relativas as covariáveis, uma vez que os seus p-valores não foram estatisticamente significativos.

Tabela 65: Coeficientes do modelo 1 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes do Modelo 1				
$Y_{Uso\ Score\ Cred} = \alpha + \beta_1 Agilidade + \beta_2 Assertividade + \beta_3 segurança + \beta_4 Credibilidade + \beta_5 Facilidade + \beta_6 Preço\ do\ Serviço + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,4765	1,1554	0,412	0,680
Agilidade	0,6394	1,1895	0,538	0,591

Assertividade	-18,8208	3756,1300	-0,005	0,996
Segurança	18,3435	4608,2694	0,004	0,997
Credibilidade	-0,7962	1,6740	-0,476	0,634
Facilidade	17,4501	4612,2021	0,004	0,997
Preço	0,4506	1,1911	0,378	0,705
Método: Enter	**** p < 0.001; *** p < 0.01; ** p < 0.05; . p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, que utiliza uma quantidade menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado, para os dados que versam sobre o que mais parece ser ruim no uso do *score* de crédito, são apresentadas, de forma compacta, na tabela 66. Enquanto a primeira linha apresenta a quantidade real de observações, a segunda linha apresenta a quantidade esperada de observações.

Tabela 66: Tabelas de contingência para os dados utilizados no modelo 1 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Usa_score	<u>Agilidade</u>		<u>Assertividade</u>		<u>Segurança</u>		<u>Credibilidade</u>		<u>Facilidade</u>		<u>Preço do Serviço</u>	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim (Real)	9	22	0	31	2	29	3	28	2	29	22	9
Não	2	12	3	11	0	14	3	11	0	14	10	4
Sim (Esp.)	7,6	23,4	2,1	28,9	1,4	29,6	4,1	26,9	1,4	29,6	22	9
Não	3,4	10,6	0,9	13,1	0,6	13,4	1,9	12,1	0,6	13,4	10	4

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 67, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Para as tabelas de contingência que apresentaram alguma casela com frequência esperada menor que 5, foram calculados, também, os p-valores pelo teste exato de *Fisher*, que neste caso, foram os valores considerados para a seleção das variáveis candidatas e para a verificação da significância estatística.

Apenas a variável *Assertividade* apresentou um p-valor menor que 0,20, sendo considerada candidata ao uso no modelo reduzido. Paralelo a isso, essa mesma variável foi significativa com um p-valor (*Fisher*) de 0,0257, indicando que a variável *Uso Score Cred* possuiu uma associação de dependência, quando realizadas as análises bivariadas.

Tabela 67: Teste de qui-quadrado aplicado às variáveis do modelo 1 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
-----------	---------------------------	----	---------	------------------

Agilidade	0,4775	1	0,4896		0,4578
Assertividade	4,0901	1	0,0431	*	0,0257 *
Segurança	0,0365	1	0,8485		1,0000
Credibilidade	0,3599	1	0,5486		0,3564
Facilidade	0,0365	1	0,8485		1,0000
Preço do Serviço	0,0000	1	1,0000		1,0000
‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1					

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística simples da única variável candidata e, por consequência, do modelo reduzido são mostrados na tabela 68. A variável independente *Assertividade* apresentou uma significância maior que 0,10 (por consequência não estatisticamente significativa), não sendo elegível ao uso no modelo reduzido.

Tabela 68: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 1 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes da Regressão Logística Simples (Assertividade)				
$Y_{Uso\ Score\ Cred} = \alpha + \beta_{Assertividade} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	1,0360	0,3510	2,952	0,003 **
Assertividade	-18,6020	2284,1020	-0,008	0,994
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Infelizmente, os modelos proposto e reduzido não possuem evidências suficientes para rejeitar a hipótese nula, não podendo inferir quaisquer probabilidades acerca dos eventos que envolvam as variáveis independentes com a variável dependente.

4.2.3.2 Modelo 2: Custo/benefício observado pela empresa no uso do score de crédito

Com relação aos respondentes que assinalaram já ter realizado alguma consulta ao *score* de crédito (31 respondentes conforme tabela 17), tendo em vista o custo-benefício no que diz respeito a utilização do *score*, foi indagado se era justo o valor pago pelo uso do serviço. Na sequência, foram realizados os seguintes questionamentos:

- Teve grandes gastos com treinamento de pessoal para usar o *score* de crédito?
- Desde que começou a utilizar o *score* de crédito, notou alguma redução na sua inadimplência?

- Desde que começou a utilizar o *score* de crédito, notou algum aumento nas vendas parceladas?

Os resultados podem ser vistos na tabela 69. Acha justo o valor pago pelo serviço, 19 respondentes (61,29%). Assinalaram que não tiveram grandes gastos com treinamento de pessoal, 24 respondentes (77,42%). Notaram alguma redução da sua inadimplência, 30 respondentes (96,77%). Todos os 31 respondentes (100%) notaram algum aumento nas vendas parceladas.

Tabela 69: Custo-benefício em relação ao uso do *score* de crédito

Variáveis	n	%
Acha justo o valor pago pelo uso do serviço de <i>score</i> de crédito?		
Sim	19	61,29
Não	12	38,71
Teve grandes gastos com treinamento de pessoal para usar o <i>score</i> de crédito?		
Sim	7	22,58
Não	24	77,42
Desde que começou a utilizar o <i>score</i> de crédito, notou alguma redução na sua inadimplência?		
Sim	30	96,77
Não	1	3,23
Desde que começou a utilizar o <i>score</i> de crédito, notou algum aumento nas vendas parceladas?		
Sim	31	100,00
Não	0	0,00

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 2 para lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, de acordo com a tabela 70, não foi possível obter observações conclusivas sobre as probabilidades relativas as covariáveis, uma vez que os seus p-valores não foram estatisticamente significativos. A covariável *Aum Vend Parcel*, por possuir um único valor (“sim”) para 100% dos casos, foi excluída da análise de regressão logística.

Tabela 70: Coeficientes do modelo 2 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes do Modelo 2				
$Y_{Custo \times Benef} = \alpha + \beta_1 Treinamento + \beta_2 Red Inadm + \beta_3 Aumt Vend Parcel + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	16,5661	2399,5447	0,007	0,994
Treinamento	-0,1542	0,8751	-0,176	0,860
Red Inadm	-16,1242	2399,5448	-0,007	0,995
Aumt Vend Parcel	-	-	-	-
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, com menos variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado de independência são apresentadas, de forma compacta, na tabela 71, cuja primeira linha apresenta a quantidade real de observações e a segunda linha apresenta a quantidade esperada de observações.

Tabela 71: Tabelas de contingência para os dados utilizados no modelo 2 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Custo x Benf	<u>Treinamento</u>		<u>Red Inadm</u>		<u>Aumt Vend Parcel</u>	
	Sim	Não	Sim	Não	Sim	Não
Sim (Real)	4	15	18	1	19	0
Não (Real)	3	9	12	0	12	0
Sim (Esp.)	4,3	14,7	18,4	0,6	19	0
Não (Esp.)	2,7	9,3	11,6	0,4	12	0

Fonte: Elaborado pela autora, 2021.

Os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência são apresentados na tabela 72. Também podemos ver os p-valores calculados pelo teste exato de *Fisher*, para as tabelas de contingência que apresentaram alguma casela com frequência esperada menor que 5. Neste caso, os p-valores do teste exato de *Fisher* foram os valores considerados para a seleção das variáveis candidatas e para verificar a significância estatística.

Nenhuma das variáveis apresentou um p-valor menor que 0,20 para ser considerada candidata ao uso no modelo reduzido. Ao mesmo tempo, nenhuma obteve um p-valor menor que 0,05, indicando que a variável *Custo x Benf* não possuiu correlação com as demais variáveis, quando das análises bivariadas.

Tabela 72: Teste de qui-quadrado aplicado às variáveis do modelo 2 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Treinamento	0,0000 ⁷¹	1	1,0000	1,0000
Red Inadm	0,0000 ⁷²	1	1,0000	1,0000
Aumt Vend Parcel	NaN ⁷³	1	NA ⁷⁴	1,0000

⁷¹ qui-quadrado = 0,000000000000000000000000000021132

⁷² qui-quadrado = 0,000000000000000000000000000046758

⁷³ NaN - *Not a number*

⁷⁴ NA – *Not available*

‘***’ $p < 0.001$; ‘**’ $p < 0.01$; ‘*’ $p < 0.05$; ‘.’ $p < 0.1$

Fonte: Elaborado pela autora, 2021.

Infelizmente, o modelo proposto não possui evidências suficientes para rejeitar a hipótese nula, não podendo inferir quaisquer probabilidades acerca dos eventos que envolvam as variáveis independentes com a variável dependente. Já para o modelo reduzido, não foi possível eleger nenhuma variável candidata para a sua construção.

4.2.3.3 Modelo 3: Avaliação da prática da LGPD sob o ponto de vista da empresa

Com relação aos respondentes que assinalaram possuir dívidas da empresa e dívidas pessoais (45 respondentes de acordo com a tabela 12), foi realizado um questionamento sobre a adequação da empresa em relação à LGPD. Em seguida, foi realizado um questionamento acerca do que pode inibir o uso das boas práticas da LGPD, tendo como opções de escolha os seguintes fatores:

- Uso de segurança para o cadastramento e armazenamento de dados;
- Treinamento de pessoal sobre o uso das boas práticas da LGPD;
- Investimentos em segurança de dados (programas seguros, banco de dados, nuvem etc.); e
- Problemas com vazamentos de dados.

Os resultados podem ser vistos na tabela 73, na qual os respondentes afirmaram que o investimento em segurança de dados é o fator que mais inibe o uso, com 62,22%. Na sequência, temos: problemas com vazamentos de dados, com 42,22%; treinamento de pessoas sobre o uso das boas práticas da LGPD, com 24,44%; e o uso de segurança para o cadastramento e armazenamento de dados, com 13,33%.

Tabela 73: Fatores que inibem o uso das boas práticas da LGPD para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	<i>n</i>	%
Uso de segurança para o cadastramento e armazenamento de dados	4	8,89
Treinamento de pessoas sobre o uso das boas práticas da LGPD	11	24,44
Investimentos em segurança de dados	28	62,22
Problemas com vazamentos de dados	19	42,22

Fonte: Elaborado pela autora, 2021

Avaliando o modelo 3 para lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, conforme tabela 74, não foi possível obter observações conclusivas sobre as probabilidades relativas as predictoras, uma vez que os seus p-valores não foram estatisticamente significativos.

Tabela 74: Coeficientes do modelo 3 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes do Modelo 3				
$Y_{Prática\ LGPD} = \alpha + \beta_1 Seg\ Dados\ Cad + \beta_2 Treinamento\ Func + \beta_3 Invest\ Seg\ Dados + \beta_4 Vazamt\ Dados + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	1,2583	0,7881	1,597	0,110
Seg Dados Cad	-0,1444	1,2425	-0,116	0,908
Treinamento Func	-0,9475	0,7733	-1,225	0,220
Invest Seg Dados	-0,9810	0,7430	-1,320	0,187
Vazamt Dados	-0,6528	0,7470	0,874	0,382
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, que utiliza uma quantidade menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado são apresentadas, de forma compacta, na tabela 75, na qual a primeira linha apresenta a quantidade real de observações e a segunda linha apresenta a quantidade esperada de observações.

Tabela 75: Tabelas de contingência para os dados utilizados no modelo 3 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Usa_LGPD	<u>Seg Dados Cad</u>		<u>Treinamento Func</u>		<u>Invest Seg Dados</u>		<u>Vazamt Dados</u>	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim	2	22	4	20	13	11	10	14
Não	2	19	7	14	15	6	9	12
Sim	2,1	21,9	5,9	18,1	14,9	9,1	10,1	13,9
Não	1,9	19,1	5,1	15,9	13,1	7,9	8,9	12,1

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 76, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. O p-valor do teste exato de Fisher foi calculado apenas entre a variável *Prática LGPD* e a variável *Seg Dados Cad*, visto que essa tabela de contingência foi a única a apresentar caselas com a frequência esperada menor que 5 (conforme visto na tabela 75).

Nenhuma das variáveis apresentou um p-valor menor que 0,20 para ser considerada candidata ao uso no modelo reduzido. Ao mesmo tempo, nenhuma obteve um p-valor menor que 0,05, indicando que a variável *Prática LGPD* não possui correlação com as demais variáveis, quando das análises bivariadas.

Tabela 76: Teste de qui-quadrado aplicado às variáveis do modelo 3 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Seg Dados Cad	0,0000 ⁷⁵	1	1,0000	1,0000
Treinamento Func	0,9029	1	0,3420	-
Invest Seg Dados	0,7804	1	0,3770	-
Vazant Dados	0,0000	1	1,0000	-

‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1

Fonte: Elaborado pela autora, 2021.

Infelizmente, o modelo proposto não possui evidências suficientes para rejeitar a hipótese nula, não podendo inferir quaisquer probabilidades acerca dos eventos que envolvam as variáveis independentes com a variável dependente. Já para o modelo reduzido, não foi possível eleger nenhuma variável candidata para a sua construção.

4.2.3.4 Modelo 4: Influência no uso de score de crédito sob o ponto de vista da empresa

Com relação aos respondentes que assinalaram possuir dívidas da empresa e dívidas pessoais (45 respondentes de acordo com a tabela 12), foram realizados três questionamentos:

- Na sua empresa, você já fez alguma consulta ao *score* de crédito?
- Sua empresa utiliza *Big Data* para questões envolvendo o *score* de crédito?
- Você já começou a adequar a sua empresa em relação à LGPD?

Os resultados podem ser vistos na tabela 77, na qual 31 respondentes (68,89%) afirmaram que já fizeram alguma consulta ao *score* de crédito. 24 respondentes (53,33%) afirmaram que utilizam *Big Data* em questões envolvendo o *score*. 24 respondentes (53,33%) afirmaram que já começaram a adequar a empresa em relação à LGPD.

⁷⁵ qui-quadrado = 0,00000000000000000000000000001981

Tabela 77: Informações sobre *Big Data* e LGPD e suas influências no uso do *score* de crédito para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	n	%
Na sua empresa, você faz alguma consulta ao <i>score</i> de crédito?		
Sim	31	68,89
Não	14	31,11
Sua empresa utiliza <i>Big Data</i> para questões envolvendo o <i>score</i> de crédito?		
Sim	24	53,33
Não	21	46,67
Você já começou a adequar a sua empresa em relação à LGPD?		
Sim	24	53,33
Não	21	46,67

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 4 para lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, não foi possível obter observações conclusivas sobre as probabilidades relativas as predictoras, uma vez que os seus p-valores não foram estatisticamente significativos, conforme exibido na tabela 78.

Tabela 78: Coeficientes do modelo 4 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes do Modelo 4				
$Y_{Score\ Cred} = \alpha + \beta_1 Big\ Data + \beta_2 LGPD + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,2231	0,4743	-0,470	0,638
Big Data	-17,4642	2254,4195	-0,007	0,994
LGPD	20,6831	2254,4197	0,009	0,993
Método: Enter	**** p < 0.001; *** p < 0.01; ** p < 0.05; . p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo reduzido, que utiliza um número menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado são apresentadas, de forma compacta, na tabela 79.

Tabela 79: Tabelas de contingência para os dados utilizados no modelo 4 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Score Cred	<u>Big Data</u>		<u>LGPD</u>	
	Sim	Não	Sim	Não
Sim (Real)	20	11	23	8
Não (Real)	4	10	1	13
Sim (Esp.)	16,5	14,5	16,5	14,5
Não (Esp.)	7,5	6,5	7,5	6,5

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 80, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Os p-valores do teste exato de *Fisher* não foram calculados, visto que as tabelas de contingência não apresentaram células com a frequência esperada menor que 5.

Ambas as variáveis apresentaram um p-valor menor que 0,20, sendo consideradas candidatas ao uso no modelo reduzido. Paralelo a isso, a variável *LGPD* foi estatisticamente significativa, com um p-valor de 0,0001, indicando que a variável *Score Cred* possuiu uma associação de dependência, quando realizadas as análises bivariadas.

Tabela 80: Teste de qui-quadrado aplicado às variáveis do modelo 4 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor		p-valor (Fisher)
Big Data	3,6665	1	0,0555	.	-
LGPD	14,8310	1	0,0001	**	-
‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1					

Fonte: Elaborado pela autora, 2021.

Os coeficientes das regressões logísticas simples das variáveis candidatas são mostrados na tabela 81. As preditoras *Big Data* e *LGPD*, quando estudadas de maneira individual, tiveram uma significância menor que 0,10, sendo utilizadas, portanto, na construção do modelo reduzido. Além disso, ambas foram significativas, apresentando sinais positivos nos coeficientes, indicando um aumento na probabilidade no uso do *score* de crédito.

Tabela 81: Coeficientes das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes da Regressão Logística Simples (Usa_Big_Data)				
$Y_{Score\ Cred} = \alpha + \beta_{Big\ Data} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,0953	0,4369	0,218	0.827
Big Data	1,5141	0,7006	2,161	0,031 *
Coeficientes da Regressão Logística Simples (Usa_LGPD)				
$Y_{Score\ Cred} = \alpha + \beta_{LGPD} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-0,4855	0,4494	-1,080	0.280 .
LGPD	3,6210	1,1158	3,245	0,001 **
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Na tabela 82 observamos as razões de chances (OR – *odds ratio*), juntamente com os intervalos de confiança (I.C.), das regressões logísticas que avaliaram, individualmente, as preditoras candidatas. Observamos que utilizar a *Big Data* aumenta em 355% a probabilidade do uso do *score* de crédito. De maneira similar, o uso da LGPD aumenta em 3638% as chances de utilizar o *score*.

Tabela 82: Razão de chances e intervalo de confiança das regressões logísticas simples das preditoras candidatas ao modelo 4 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais

Razão de Chances da Regressão Logística Simples (Usa_Big_Data)			
$Y_{Score\ Cred} = \alpha + \beta_{Big\ Data} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	1,10	0,4672	2,5901
Big Data	4,55	1,1513	17,9459
Razão de Chances da Regressão Logística Simples (Usa_LGPD)			
$Y_{Score\ Cred} = \alpha + \beta_{LGPD} + \varepsilon$			
Variável	OR	I.C. 95%	
		2,5%	97,5%
(Intercepto)	0,62	0,2551	1,4847
LGPD	37,38	4,1959	332,9211

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística múltipla do modelo reduzido são os mesmos do modelo proposto, haja vista que as covariáveis são as mesmas. Sendo assim, como os modelos são iguais, não é possível obter observações conclusivas. Contudo, as únicas informações sobre a influência da *Big Data* e da LGPD no uso do *score* de crédito são àquelas fornecidas através das análises das regressões logísticas simples e dos testes de associação bivariados.

4.2.3.5 Modelo 5: Avaliação do endividamento sob a ótica do tomador de crédito

Com relação aos respondentes que assinalaram possuir dívidas da empresa e dívidas pessoais (45 respondentes de acordo com a tabela 12), foi questionado sobre quais elementos influenciariam no aumento das dívidas, tendo como opções de escolha: a facilidade de crédito, a possibilidade de poder parcelar contas, o aumento salarial, a quantidade de cartões de crédito, a quantidade de contas no banco, já possuir dívidas e ter o nome limpo.

Os resultados podem ser vistos na tabela 83, na qual os respondentes afirmaram que a facilidade de crédito, com 93,33%, é o elemento que mais interfere no aumento das dívidas. Na sequência, foram assinalados: já possuir dívidas, com 91,11%; a quantidade de cartões de

crédito, com 64,44%; a quantidade de contas no banco, com 40%; a possibilidade de poder parcelar contas, com 22,22%; ter o nome limpo, com 15,56%; e o aumento salarial, com 11,11%.

Tabela 83: Elementos que podem influenciar no aumento das dívidas para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	n	%
Facilidade de crédito	42	93,33
Possibilidade de poder parcelar contas	10	22,22
Aumento salarial	5	11,11
Quantidade de cartões de crédito	29	64,44
Quantidade de contas no banco	18	40,00
Já possuir dívidas	41	91,11
Ter nome limpo	7	15,56

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 5 para os tomadores de crédito que possuem dívidas da empresa e dívidas pessoais, de acordo com a tabela 84, não foi possível obter observações conclusivas sobre as probabilidades relativas as covariáveis, uma vez que os p-valores não foram estatisticamente significativos.

Tabela 84: Coeficientes do modelo 5 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes do Modelo 5				
$Y_{Endividamento} = \alpha + \beta_1 Fac\ de\ Cred + \beta_2 Parcel\ de\ Cont + \beta_3 Aum\ Sal + \beta_4 Qnt\ Cart\ de\ Cred + \beta_5 Qnt\ Conta\ Banc + \beta_6 Qnt\ Dívida\ Ativa + \beta_7 Nome\ Limpo + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-17,8000	23519,9300	-0,001	0,999
Fac de Cred	17,7000	11063,0300	0,002	0,999
Parcel de Cont	-16,5600	19454,2900	-0,001	0,999
Aum Sal	-19,0900	11063,0300	-0,002	0,999
Qnt Cart de Cred	39,3300	11903,8600	0,003	0,997
Qnt Conta Banc	17,8000	12591,3400	0,001	0,999
Qnt Dívida Ativa	19,8800	23519,9300	0,001	0,999
Nome Limpo	-38,8800	13967,1400	-0,003	0,998
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo 5 reduzido, que utiliza menos variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado são apresentadas, de forma compacta, na tabela 85, cuja linha que apresenta a expressão (*Real*) exibe a quantidade real de observações e a linha que apresenta a expressão (*Esp.*) exibe a quantidade esperada de observações.

Tabela 85: Tabelas de contingência para os dados utilizados no modelo 5 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Endividamento	<u>Fac de Cred</u>		<u>Parcel de Cont</u>		<u>Aum Sal</u>		<u>Qnt Cart Cred</u>		<u>Qnt Conta Banc</u>		<u>Qnt Dívida Ativa</u>	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim (Real)	22	20	10	32	4	38	29	13	18	24	38	4
Não (Real)	2	1	0	3	1	2	0	3	0	3	3	0
Sim (Esp.)	22,4	19,6	9,3	32,7	4,7	37,3	27,1	14,9	16,8	25,2	38,3	3,7
Não (Esp.)	1,6	1,4	1	2,3	0,3	2,7	1,9	1,1	1,2	1,8	2,7	0,3
	<u>Nome Limpo</u>											
	Sim	Não										
Sim (Real)	6	36										
Não (Real)	1	2										
Sim (Esp.)	6,5	35,5										
Não (Esp.)	0,5	2,5										

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 86, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Também podemos ver os p-valores calculados pelo teste exato de Fisher, para as tabelas de contingência que apresentaram alguma casela com frequência esperada menor que 5. Neste caso, os p-valores do teste exato de Fisher foram os valores considerados para a seleção das variáveis candidatas e para verificar a significância estatística.

Apenas a variável *Qnt Cart Cred* apresentou um p-valor menor que 0,20, sendo considerada candidata ao uso no modelo reduzido. Paralelo a isso, essa mesma variável foi significativa com um p-valor (Fisher) de 0,0395, indicando que a variável *Endividamento* possuiu uma associação de dependência, quando realizadas as análises bivariadas.

Tabela 86: Teste de qui-quadrado aplicado às variáveis do modelo 5 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Fac de Cred	0,0000 ⁷⁶	1	1,0000	1,0000
Parcel de Cont	0,0574	1	0,8107	1,0000
Aum Sal	0,1005	1	0,7513	0,3037
Qnt Cart Cred	3,2022	1	0,0735	0,0395 *
Qnt Conta Banc	0,7292	1	0,3932	0,2636

⁷⁶ qui-quadrado = 0,00000000000000000000000000000023771

Qnt Dívida Ativa	0,0000 ⁷⁷	1	1,0000	1,0000
Nome Limpo	0,0030	1	0,9562	0,4055
‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1				

Fonte: Elaborado pela autora, 2021.

Os coeficientes da regressão logística simples da única variável candidata e, por consequência, do modelo reduzido são mostrados na tabela 87. A variável independente *Qnt Cart Cred* apresentou uma significância maior que 0,10 (por consequência não estatisticamente significativa), não sendo elegível ao uso no modelo reduzido.

Tabela 87: Coeficientes da regressão logística simples da preditora candidata ao modelo 5 reduzido para os respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes da Regressão Logística Simples (Facilidade de Crédito)				
$Y_{\text{Endividamento}} = \alpha + \beta_{\text{Fac de Cred}} + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	1,4663	0,6405	2,289	0.022 *
Qnt Cart Cred	19,0997	3292,4471	0,006	0,995
Método: Enter	‘***’ p < 0.001; ‘**’ p < 0.01; ‘*’ p < 0.05; ‘.’ p < 0.1			

Fonte: Elaborado pela autora, 2021.

Infelizmente, os modelos proposto e reduzido não possuem evidências suficientes para rejeitar a hipótese nula, não podendo inferir quaisquer probabilidades acerca dos eventos que envolvam as variáveis independentes com a variável dependente.

4.2.3.6 Modelo 6: Influência do endividamento sob a ótica do tomador de crédito

Com relação aos respondentes que assinalaram possuir dívidas da empresa e dívidas pessoais (45 respondentes de acordo com a tabela 12), foi inquirido quais fatores influenciariam na formação do endividamento, tendo como opções de escolha: a falta de conhecimentos sobre o *score* de crédito, a falta de conhecimentos sobre o cálculo do *score* de crédito, a confiança na empresa que vai parcelar a dívida/obter crédito, a confiança na segurança dos dados pessoais, a facilidade na obtenção de crédito, o valor das parcelas (possibilidade de pagamento do valor acordado) e ter o nome limpo.

Os resultados podem ser vistos na tabela 88, na qual os respondentes afirmaram que a facilidade de obtenção crédito, com 71,11%, é o fator que possui mais influência na formação

⁷⁷ qui-quadrado = 0,0000000000000000000000000000000038395

do endividamento. Na sequência, temos: o valor das parcelas (possibilidade de pagamento do valor acordado), com 40%; a falta de conhecimentos sobre o cálculo do *score* de crédito, com 37,78%; ter o nome limpo, com 26,67%; a falta de conhecimentos sobre o cálculo do *score* de crédito, com 17,78%; a confiança na empresa que vai parcelar a dívida/obter crédito, com 11,11%; e a confiança na segurança dos dados pessoais, com 6,67%.

Tabela 88: Fatores que podem influenciar na formação do endividamento para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	n	%
Falta de conhecimentos sobre o <i>score</i>	17	37,78
Falta de conhecimentos sobre o cálculo do <i>score</i>	8	17,78
Confiança na empresa que vai parcelar/obter	5	11,11
Confiança na segurança dos dados pessoais	3	6,67
Facilidade na obtenção de crédito	32	71,11
Valor das parcelas	18	40,00
Ter nome limpo	12	26,67

Fonte: Elaborado pela autora, 2021.

Avaliando o modelo 6 para os tomadores de crédito que possuem dívidas da empresa e dívidas pessoais, não foi possível obter observações conclusivas sobre as probabilidades relativas as variáveis independentes, conforme tabela 89, uma vez que os p-valores associados às covariáveis não foram estatisticamente significativos.

Tabela 89: Coeficientes do modelo 6 referentes aos respondentes que possuem dívidas da empresa e dívidas pessoais

Coeficientes do Modelo 6				
$Y_{Tomd\ de\ Cred} = \alpha + \beta_1\ Conhec\ Score\ Cred + \beta_2\ Entend\ Calc\ Score\ Cred + \beta_3\ Conf\ Empresa + \beta_4\ Conf\ Seg\ Dados + \beta_5\ Fac\ de\ Cred + \beta_6\ Valores\ Parcel + \beta_7\ Consulta\ Serasa/SPC + \varepsilon$				
Variáveis	β	S.E.	z-valor	p-valor
(Intercepto)	-19,3353	6651,6496	-0,003	0,998
Conhec Score Cred	0,6931	1,4577	0,475	0,634
Entend Calc Score Cred	-19,2113	9006,0617	-0,002	0,998
Conf Empresa	-19,0771	11222,3297	-0,002	0,999
Conf Seg Dados	-1,5288	17184,1020	0,000	1,000
Fac de Cred	17,9490	6651,6496	0,003	0,998
Valores Parcel	-16,7780	5335,4626	-0,003	0,997
Consulta Serasa/SPC	-19,6428	7449,8662	-0,003	0,998
Método: Enter	***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1			

Fonte: Elaborado pela autora, 2021.

Para a geração do modelo 6 reduzido, que utiliza uma quantidade menor de variáveis independentes, as tabelas de contingência utilizadas nos testes qui-quadrado de independência,

para os dados que cuidam sobre o que influenciaria na formação do endividamento, são apresentadas de forma compacta, na tabela 90, cuja linha que apresenta a expressão (*Real*) exibe a quantidade real de observações e a linha que apresenta a expressão (*Esp.*) exibe a quantidade esperada de observações.

Tabela 90: Tabelas de contingência para os dados utilizados no modelo 6 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Tomada de Crédito	<u>Conhec</u> <u>Score Cred</u>		<u>Entend Calc</u> <u>Score Cred</u>		<u>Conf</u> <u>Empresa</u>		<u>Conf Seg</u> <u>Dados</u>		<u>Fac de</u> <u>Cred</u>		<u>Valores</u> <u>Parcel</u>	
	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Sim	16	26	8	34	5	37	3	39	29	13	18	24
Não	1	2	0	3	0	3	0	3	3	0	0	3
Sim	15,9	26,1	7,5	34,5	4,7	37,3	2,8	39,2	29,9	12,1	16,8	25,2
Não	1,1	1,9	0,5	2,5	0,3	2,7	0,2	2,8	2,1	0,9	1,2	1,8
	<u>Consulta</u> <u>Serasa/SPC</u>											
	Sim	Não										
Sim	12	30										
Não	0	3										
Sim	11,2	30,8										
Não	0,8	2,2										

Fonte: Elaborado pela autora, 2021.

Podemos ver, na tabela 91, os valores do qui-quadrado e os p-valores calculados pelo teste qui-quadrado de independência. Também podemos ver os p-valores calculados pelo teste exato de *Fisher*, para as tabelas de contingência que apresentaram alguma casela com frequência esperada menor que 5. Neste caso, os p-valores do teste exato de *Fisher* foram os valores considerados para a seleção das variáveis candidatas e para verificar a significância estatística.

Nenhuma das variáveis apresentou um p-valor menor que 0,20 para ser considerada candidata ao uso no modelo reduzido. Ao mesmo tempo, nenhuma obteve um p-valor menor que 0,05, indicando que a variável *Tomd de Cred* não possuiu correlação com as demais variáveis, quando das análises bivariadas.

Tabela 91: Teste de qui-quadrado aplicado às variáveis do modelo 6 para os respondentes que possuem dívidas da empresa e dívidas pessoais

Variáveis	qui-quadrado (χ^2)	df	p-valor	p-valor (Fisher)
Conhec Score Cred	0,0000 ⁷⁸	1	1,0000	1,0000
Entend Calc Score Cred	0,0027	1	0,9584	1,0000
Conf Empresa	0,0000 ⁷⁹	1	1,0000	1,0000
Conf Seg Dados	0,0000 ⁸⁰	1	1,0000	1,0000
Fac de Cred	0,2337	1	0,6288	0,5456
Valores Parcel	0,7292	1	0,3932	0,2636
Consulta Serasa/SPC	0,1644	1	0,6852	0,5535
*** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.1				

Fonte: Elaborado pela autora, 2021.

Infelizmente, o modelo proposto não possui evidências suficientes para rejeitar a hipótese nula, não podendo inferir quaisquer probabilidades acerca dos eventos que envolvam as variáveis independentes com a variável dependente. Já para o modelo reduzido, não foi possível eleger nenhuma variável candidata para a sua construção.

4.3 Sumarização dos Resultados

A tabela 92 apresenta um resumo acerca dos resultados encontrados nas análises econométricas, isto é, as observações encontradas tanto nos modelos originais propostos quanto nos modelos reduzidos. Podemos perceber que, ao aplicar o princípio da parcimônia, por meio dos modelos reduzidos, otimizamos a avaliação das hipóteses ao apresentar modelos mais simples. Dada a complexidade dos modelos propostos, foi recorrente a ausência de evidências para rejeitar a hipótese nula, cuja premissa é que as predictoras não exerçam influência sobre a variável dependente. Contudo, os modelos simplificados, cujas observações tendem a ter maiores chances em explicar os eventos, apresentaram observações relevantes no que concerne as hipóteses formuladas.

Os modelos trataram de investigar:

- Modelo 1: A avaliação do uso de score de crédito sob o ponto de vista da empresa;

⁷⁸ qui-quadrado = 0,0000000000000000000000000027966

⁷⁹ qui-quadrado = 0,0000000000000000000000000049524

⁸⁰ qui-quadrado = 0,0000000000000000000000000022427

- Modelo 2: O custo/benefício observado pela empresa no uso do *score* de crédito;
- Modelo 3: A avaliação da prática da LGPD sob o ponto de vista da empresa;
- Modelo 4: A influência no uso de *score* de crédito sob o ponto de vista da empresa;
- Modelo 5: A avaliação do endividamento sob a ótica do tomador de crédito; e
- Modelo 6: A influência do endividamento sob a ótica do tomador de crédito.

Tabela 92: Sumarização dos resultados

Somente dívidas da empresa	
Modelo 1 - Proposto	Modelo 1 - Reduzido
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).	- Os coeficientes <i>Assertividade</i> e <i>Credibilidade</i> , embora não possuam significância estatística ($p < 0,05$), possuem sinais negativos, podendo contribuir, no caso de existir significância econômica ou prática, para diminuir a probabilidade em usar o <i>score</i> de crédito.
Modelo 2 - Proposto	Modelo 2 - Reduzido
- O gasto com treinamento de pessoal diminui em 85% o custo-benefício no uso do <i>score</i> de crédito.	- O modelo reduzido explica o evento de uma forma mais simples, com menos variáveis independentes. - O gasto com treinamento de pessoal diminui em 82% o custo-benefício no uso do <i>score</i> de crédito.
Modelo 3 - Proposto	Modelo 3 - Reduzido
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula). - Os coeficientes <i>Treinamento Func</i> e <i>Invest Seg Dados</i> , embora não possuam significância estatística ($p < 0,05$), possuem sinais negativos, podendo contribuir, no caso de existir significância econômica ou prática, para diminuir a probabilidade em usar a LGPD.	- Embora o modelo proposto não seja estatisticamente significativo, o modelo reduzido tenta explicar o evento de uma forma mais simples, com menos variáveis independentes. - O fato de ter de investir na segurança dos dados diminui em 83% a probabilidade no uso da LGPD.
Modelo 4 - Proposto	Modelo 4 - Reduzido
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).	- Todos os coeficientes foram selecionados, de forma que o modelo reduzido é igual ao modelo proposto. Logo, não existem evidências suficientes para rejeitar a H_0 (Hipótese nula). - Analisando as variáveis <i>Big Data</i> e LGPD, de maneira individual, observamos que utilizar a <i>Big Data</i> aumenta em 239% a probabilidade do uso do <i>score</i> de crédito, enquanto utilizar a LGPD aumenta em 1363% as chances de usar o <i>score</i> .
Somente dívidas pessoais	
Modelo 5 - Proposto	Modelo 5 - Reduzido
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).	- O coeficiente <i>Fac de Cred</i> , embora não possua significância estatística ($p < 0,05$), possui sinal positivo, podendo contribuir, no caso de existir significância econômica ou prática, para ampliar a probabilidade de influenciar o aumento das dívidas. - Os coeficientes <i>Parcel de Contas</i> e <i>Aum Sal</i> , embora não possuam significância estatística ($p < 0,05$), possuem sinais negativos, indicando, no caso de existir significância econômica ou prática, uma possível atenuação no aumento das dívidas.

Modelo 6 - Proposto		Modelo 6 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula). - O coeficiente <i>Entend Calc Score Cred</i> , embora não possua significância estatística ($p < 0,05$), possui sinal positivo, podendo contribuir, no caso de existir significância econômica ou prática, para aumentar a influência na formação do endividamento.		- Embora o modelo proposto não seja estatisticamente significativo, o modelo reduzido tenta explicar o evento de uma forma mais simples, com menos variáveis independentes. - A falta de conhecimentos sobre o cálculo do <i>score</i> de crédito aumenta em 148% a influência na formação do endividamento.	
Dívidas da empresa e dívidas pessoais			
Modelo 1 - Proposto		Modelo 1 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).		- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).	
Modelo 2 - Proposto		Modelo 2 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).		- Não foi possível eleger nenhuma variável candidata a construção do modelo reduzido.	
Modelo 3 - Proposto		Modelo 3 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).		- Não foi possível eleger nenhuma variável candidata a construção do modelo reduzido.	
Modelo 4 - Proposto		Modelo 4 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).		- Todos os coeficientes foram selecionados, de forma que o modelo reduzido é igual ao modelo proposto. Logo, não existem evidências suficientes para rejeitar a H_0 (Hipótese nula). - Analisando as variáveis <i>Big Data</i> e LGPD, de maneira individual, observamos que utilizar a <i>Big Data</i> aumenta em 355% a probabilidade do uso do <i>score</i> de crédito, enquanto utilizar a LGPD aumenta em 3638% as chances de usar o <i>score</i> .	
Modelo 5 - Proposto		Modelo 5 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).		- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).	
Modelo 6 - Proposto		Modelo 6 - Reduzido	
- Sem evidências suficientes para rejeitar a H_0 (Hipótese nula).		- Não foi possível eleger nenhuma variável candidata a construção do modelo reduzido.	

Fonte: Elaborado pela autora, 2021.

Para os respondentes que assinalaram possuir apenas dívidas da empresa, dos 04 modelos originais propostos, mais complexos, apenas o modelo 2 apresentou significância estatística. Ao aplicar o princípio da simplificação aos modelos originais, os modelos 2 e 3 foram significantes estatisticamente. Em relação ao modelo 2, estatisticamente significativo tanto no modelo original quanto no modelo simplificado, o princípio da parcimônia foi fundamental para tornar o modelo mais enxuto e, conseqüentemente, com maiores chances de acerto.

Para os respondentes que assinalaram possuir apenas dívidas pessoais, dos 02 modelos originais propostos, mais complexos, nenhum apresentou significância estatística. Aplicando o princípio da simplificação, o modelo 6 foi estatisticamente significativo. Por meio da aplicação

do princípio da parcimônia, o modelo 6 foi otimizado, apresentando resultados para a análise científica.

Para os respondentes que assinalaram possuir tanto dívidas da empresa quanto dívidas pessoais, dos 06 modelos originais propostos, mais complexos, nenhum apresentou significância estatística. Mesmo aplicando o princípio da simplificação, nenhum modelo desse grupo de respondentes conseguiu ser otimizado ao ponto de apresentar resultados estatisticamente significantes.

5 CONCLUSÃO

A presente pesquisa objetivou analisar como o sistema de *score* de crédito pode ser impactado pelo uso da *Big Data* e da nova LGPD, bem como suas influências junto às empresas que operam com crédito e aos tomadores de crédito.

Para tanto foram propostos e aplicados seis novos modelos, dos quais quatro, sob o ponto de vista dos lojistas/empresas, avaliaram: o uso de *score* de crédito (modelo 1); o custo-benefício no uso do *score* de crédito (modelo 2); o uso da LGPD (modelo 3); e a relação entre *score* de crédito, Big Data e LGPD (modelo 4). Os outros dois, relacionados aos tomadores de crédito, examinaram: o que pode potencializar o endividamento (modelo 5); e o que pode influenciar na tomada de crédito (modelo 6).

Ao avaliar o modelo 1, tanto para os lojistas/empresas que possuem somente dívidas das empresas quanto para os que possuem dívidas da empresa e dívidas pessoais, os dados mostraram que mais de 90% dos respondentes já ouviram falar do *score* de crédito e que mais de 60% já realizaram alguma consulta ao serviço. Ainda assim, a grande maioria dos respondentes nunca buscou verificar como se realiza o cálculo do *score* e, apesar disso, quase 100% da amostra não acha justo ocultar quais informações são efetivamente utilizadas nesse cálculo e quão relevantes elas são.

O elemento que mais contribui para a utilização do serviço de *score* de crédito, ultrapassando a marca de 70% dos respondentes, é a *credibilidade proporcionada*, seguidas pela *assertividade*, *agilidade* e *segurança proporcionada*. Na ponta oposta, com mais de $\frac{3}{4}$ dos

respondentes, o elemento que mais inibe a (possível) utilização é o *preço do serviço*. Possíveis problemas de *assertividade*, *agilidade* e *credibilidade* também podem inibir o uso.

Com relação as análises dos modelos propriamente ditos, ou seja, da equação proposta e da equação simplificada, tanto para os lojistas/empresas que possuem somente dívidas da empresa quanto para os que possuem dívidas da empresa e dívidas pessoais, não foi possível chegar a uma conclusão definitiva, embora problemas com a *credibilidade* e a *assertividade* possam reduzir as chances de uso do *score*, quando das análises individuais para os lojistas/empresas que possuem somente dívidas da empresa.

No que se refere ao modelo 2, foi verificada uma dissidência entre os lojistas/empresas que possuem somente dívidas das empresas e os que possuem dívidas da empresa e dívidas pessoais, com relação ao custo-benefício no uso do *score* de crédito: enquanto mais de 60% do primeiro grupo não acha justo o valor pago pelo serviço, no outro grupo, mais de 60% consideram justo o valor pago. Ainda que exista essa divergência, a maioria dos respondentes dos dois grupos tiveram grandes gastos com treinamento de pessoal e mais de 90% tanto perceberam redução na sua inadimplência quanto no aumento das suas vendas parceladas.

Examinando o modelo proposto para lojistas/empresas que possuem apenas dívidas da empresa, o gasto com treinamento de pessoal reduz em 85% a relação custo-benefício em se utilizar o *score* de crédito. Já no modelo simplificado para esse mesmo grupo, que analisou o gasto com treinamento de pessoal, de forma individual, a redução foi de 83%. Para os lojistas/empresas que possuem tanto dívidas da empresa quanto dívidas pessoais não foi possível chegar a um desfecho definitivo tanto no modelo proposto quanto no modelo reduzido.

Acerca do modelo 3, tanto para os lojistas/empresas que possuem somente dívidas das empresas quanto para os que possuem dívidas da empresa e dívidas pessoais, mais da metade dos respondentes já começou a adequar a empresa às boas práticas da LGPD. Concomitantemente, mais de 50% também relataram problemas com vazamentos de dados.

O fator *investimentos em segurança de dados (programas seguros, banco de dados, nuvem etc.)* influencia tanto positivamente quanto negativamente, no uso das boas práticas da LGPD, para mais de 60% dos respondentes. Resta evidente que a amostra está dividida, com relação à adequação da LGPD, em considerar ora um investimento ora um custo para a empresa.

Com relação à análise do modelo proposto, isto é, o uso das boas práticas da LGPD depender dos seguintes fatores, a saber: *uso de segurança para o cadastramento e armazenamento de dados, treinamento de pessoas sobre o uso das boas práticas da LGPD, investimentos em segurança de dados e problemas com vazamentos de dados*, tanto para os lojistas/empresas que possuem somente dívidas da empresa quanto aqueles que possuem dívidas da empresa e dívidas pessoais não foi possível apresentar inferências válidas. Para o modelo simplificado, apenas com os fatores *treinamento de pessoas sobre o uso das boas práticas da LGPD e investimentos em segurança de dados*, para os lojistas/empresas que possuem somente dívidas da empresa, os investimentos em segurança de dados podem diminuir em 83% o uso das boas práticas da LGPD.

No modelo 4 proposto, que relaciona o que poderia influenciar o *score* de crédito, comparando com o uso da *Big Data* e o uso da LGPD, temos que, tanto para os lojistas/empresas que possuem somente dívidas da empresa quanto aqueles que possuem dívidas da empresa e dívidas pessoais, mais de 60% dos respondentes já fizeram alguma consulta ao *score*. Paralelamente, mais de 50% utilizam a *Big Data* no uso do *score*, assim como já começaram a adequar a empresa às boas práticas da LGPD.

Com relação ao exame do modelo proposto, tanto para os lojistas/empresas que possuem somente dívidas da empresa quanto para os que possuem dívidas da empresa e dívidas pessoais, não foi possível chegar a uma conclusão definitiva. Contudo, esses fatores influenciam, possivelmente, de maneira positiva, visto que, nas análises individuais das variáveis, para os lojistas/empresas que possuem apenas dívidas da empresa, o uso da *Big Data* pode influenciar em 239% o uso do *score* de crédito. Já a utilização da LGPD pode influenciar em 1363% o uso do *score*. Para os lojistas/empresas que possuem tanto dívidas da empresa quanto dívidas pessoais, a influência no uso do *score* foi de 355% para o uso da *Big Data* e 3638% para o uso da LGPD.

Avaliando o modelo 5, que visa verificar o que pode potencializar o endividamento, o elemento que mais contribuiu, para os tomadores de crédito, para potencializar o endividamento é a *quantidade de cartões de crédito*, sendo assinalado por mais de 60% dos respondentes, seguida pela *quantidade de contas bancárias*. Para os lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, a *facilidade de crédito* é o elemento que mais contribuiu, com mais de 90% dos respondentes, seguido por *já possuir dívidas*.

Com relação ao exame do modelo proposto, tanto para os tomadores de crédito quanto para os lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, não foi possível chegar a uma conclusão definitiva. Contudo, para os tomadores de crédito, na análise individual da *facilidade de crédito*, embora a significância tenha ultrapassado o valor limite de 5%, este elemento talvez possa potencializar positivamente o endividamento.

Ao avaliar o modelo 6, tanto para os tomadores de crédito quanto para os lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, o fator que mais influencia a tomada de crédito é a *facilidade de crédito*, com mais de 40% dos respondentes para os tomadores de crédito e mais de 70% para os lojistas/empresas que possuem dívidas da empresa e dívidas pessoais.

Com relação à análise da equação proposta, ou seja, o que pode influenciar a tomada de crédito, para o grupo dos tomadores de crédito, tanto a *falta de conhecimentos sobre o score de crédito*, a *falta de conhecimentos sobre o cálculo do score* e a *confiança na empresa que vai parcelar/obter* pode influenciar, de maneira positiva, quando são avaliadas de forma individual. No modelo reduzido, que avalia *falta de conhecimentos sobre o score de crédito* e a *falta de conhecimentos sobre o cálculo do score*, vemos que o não entendimento do cálculo do *score* influencia em 148% a tomada de crédito. Já para os lojistas/empresas que possuem dívidas da empresa e dívidas pessoais, não foi possível chegar a uma conclusão definitiva.

Além da análise dos modelos matemáticos, foram ponderados os seguintes questionamentos, para uma reflexão de toda uma sociedade: O *credit score* realmente teria transparência frente ao titular dos dados? Até que ponto o *credit score*, estará apto a aplicar o direito de revisões automatizadas? Esse cadastro, cuja base de dados é objeto de tratamento, poderia provocar dificuldades para a obtenção de crédito para aqueles que, porventura, já estiveram classificados como “maus pagadores”?

Dentro deste contexto, este trabalho tentou apresentar como a nossa sociedade evoluiu e caminhou para uma sociedade algorítmica, na qual os dados pessoais, fornecidos de forma natural por inúmeros usuários, constituíram-se em grandes ativos de empresas da área de TIC⁸¹. Em consonância, com a exploração da *Big Data* e das aplicações de algoritmos nas áreas de

⁸¹ TIC – Tecnologia da Informação e Comunicação.

ciências de dados, inteligência artificial e *machine learning*, que por sua natureza robótica, automatizaram, de maneira ímpar, a coleta e a análise dessas informações.

Sobre os questionamentos anteriormente citados, examinou-se por normas mais contemporâneas e pertinentes à proteção de dados e aos direitos fundamentais, que possuem em seu arcabouço o intuito de proteger os titulares de dados. Neste sentido, o Brasil vem alcançando patamares igualitários às principais democracias do mundo no que diz respeito a proteção de dados, por meio da LGPD, cuja interação com a Lei do Cadastro Positivo para a proteção de crédito se dá de maneira direta.

Sendo assim, a transparência do *credit score*, frente ao titular de dados, vem da combinação dos artigos 7º e 20º da LGPD, com seus pontos relevantes, quanto ao foco no tratamento de dados pessoais. Neles constam que, a respeito do tratamento de dados pessoais cujo acesso é público, deve prevalecer a boa-fé e o interesse público que justifiquem sua disponibilização.

Posto isso, não basta apenas a criação de leis para regulamentar a coleta de informações e a análise de crédito. Para vias de mitigação da inadimplência, faz-se necessário investir na educação financeira da população, abarcando uma consciência sobre os valores da economia e melhoria nas condições de emprego e qualidade de vida, evitando assim o superendividamento. Com certeza estes são os alicerces para distanciar os indivíduos de uma provável negativação e até mesmo da criação do *credit score* que os prejudiquem.

Ao mesmo tempo prevê-se um certo conflito, visto que por um lado objetiva-se assegurar a privacidade e a garantia do controle dos dados pessoais conforme a lei, porém, por outro prisma, é assegurado a autorização do tratamento de dados pessoais de forma automatizada, sem o consentimento expresso dos indivíduos frente a lei do Cadastro Positivo. Essa não é uma discussão simples, pois engloba aspectos econômicos, sociais, políticos e tecnológicos, sob a perspectiva de algoritmos de pontuação baseados em dados pessoais. Há necessidade de um despertar para tais práticas, já que estas podem evidenciar que os *scores* obtidos talvez reproduzam vieses de desigualdades, que já são bastante evidentes.

Portanto, que por meio da LGPD, ao menos, o titular dos dados obtenha o direito de respostas transparentes e a possibilidade de contestar as decisões discordantes que são tomadas pelos algoritmos de forma automatizada. Com o advento da LGPD, os indivíduos terão a

oportunidade de acompanhar o amadurecimento da legislação, a busca de seus direitos e, principalmente, testemunhar e consolidar uma mudança de cultura de dados no Brasil.

Como sugestões para futuras pesquisas, no que tange como o sistema de *Score* de crédito pode ser impactado pelo uso de *Big data* e pela nova LGPD, uma maior quantidade de indivíduos respondentes que possam refinar e aperfeiçoar o cálculo das possíveis *odds ratio* dos modelos matemáticos propostos, aumentando o poder estatístico da amostra, associado ao estudo de serviços incipientes, observados por meio do Pix⁸² e do *Open Bank*⁸³, em que o grande desafio será criar os meios adequados para a coleta e a gestão do consentimento para o tratamento de dados pessoais, poderá promover uma significativa contribuição no desenvolvimento desse conteúdo.

⁸² Pix – Serviço de pagamento instantâneo gerenciado pelo Banco Central do Brasil.

⁸³ *Open Bank* - Consiste em um conjunto de regras e tecnologias voltadas para simplificar o compartilhamento de dados entre diferentes instituições financeiras, de forma gratuita e segura.

REFERÊNCIAS

- ABDOU, H. A.; POINTON, J. Credit scoring, statistical techniques and evaluation criteria: a review of the literature: credit scoring, techniques & evaluation criteria: a literature review. **Intelligent Systems in Accounting, Finance and Management**, v. 18, n. 2–3, p. 59–88, 2011.
- ACSP. Cadastro positivo: como ele pode beneficiar a sua empresa. **Associação Comercial de São Paulo**. Disponível em: <https://acsp.com.br/publicacao/s/cadastro-positivo-como-ele-pode-beneficiar-a-sua-empresa>, Acesso em: 10 de Outubro de 2020.
- AGRAWAL, D. Analytics based decision making. **Journal of Indian Business Research**, v. 6, n. 4, p. 332–340, 2014.
- AKERKAR, R. Analytics on big aviation data: Turning Data into Insights. **International Journal of Computer Science and Applications**, v. 11, n.3, p.116-127; 2014.
- ALI, A.; GHOURI, S. P. Global Crisis and Credit Risk Management by Banks: a comparative study of banks in pakistan. **International Journal Of Business And Economics Research**, v. 2, n. 6, p. 158-168, 2013.
- ALMEIDA, P. H. S. A.; SANTOS, L. P. G.; ALMEIDA, L. A. L.; ROCHA, J. S.; FREITAS, S. C. Utilização de Algoritmo Genético na Parametrização de Redes Neurais Artificiais para Aplicação na Elaboração de Orçamento de Vendas. XXVIII Encontro Nacional dos Programas de Pós-Graduação em Administração – EnANPAD, **Anais**, Curitiba. De 25 a 29 de setembro de 2004.
- ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. **The Journal of Finance**, v. 23, n. 4, p. 589–609, 1968.
- ALTMAN, E. I.; HARTZELL, J.; PECK, M. Emerging market corporate bonds—A scoring system. In: **Emerging market capital flows**. Springer, Boston: MA, 1998.
- ALTMAN, E. I. Predicting financial distress of companies: revisiting the Z-score and ZETA® models. In: **Handbook of research methods and applications in empirical finance**. Edward Elgar Publishing, 2013.

ANALYTICS10. **Top 10 Big Data Trends of 2020**. 2020. Disponível em: <https://www.analyticsinsight.net/top-10-big-data-trends-2020/>. Acesso em: 22 de Dezembro de 2020.

ANBIMA. ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS. **Basiléia III: Novos desafios para a adequação da regulação bancária**. Rio de Janeiro: Anbima, 2010.

ANDERSON, R. **The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation**. New York: Oxford University Press, 2007.

ARRUDA, V. C. G. **Big Data, dados pessoais e propriedade intelectual: limites e tutela jurídica das bases de dados**, 2019. 98 f. Dissertação (Direito Político e Econômico) – Universidade Presbiteriana Mackenzie, São Paulo.

AZAMBUJA, A. J. G.; GRANVILLE, L. Z.; SARMENTO, A. G. M. A privacidade, a segurança da informação e a proteção de dados no Big Data. **Parcerias Estratégicas**, v. 24, n. 48, p. 09-32, 2020.

BACEN. **Estudos Especiais do Banco Central**. 2020. Disponível em: https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/EE080_Indicadores_de_endividamento_de_risco_e_perfil_do_tomador_de_credito.pdf. Acesso em: 15 dez. 2020.

BACEN. **Endividamento das famílias com o Sistema Financeiro Nacional em relação à renda acumulada dos últimos doze meses**. 2021. Disponível em: <https://dadosabertos.bcb.gov.br/dataset/19882-endividamento-das-familias-com-o-sistema-financeiro-nacional-em-relacao-a-renda-acumulada-dos>. Acesso em: 30 jun. 2021.

BAGLEY, A. W.; BROWN, J. S. Limited Consumer Privacy Protections Against the Layers of Big Data. **Santa Clara Computer & High Tech. LJ**, v. 31, n. 1, p. 483, 2014.

BANASIK, J.; CROOK, J.; THOMAS, L. Sample selection bias in credit scoring models. **Journal of the Operational Research Society**, v. 54, n. 8, p. 822-832, 2003.

BANCO DE PORTUGAL. MAR - Modelo de Avaliação de Riscos. **Departamento de Supervisão Bancária - Boletim Econômico**, 2007. Disponível em:

https://www.bportugal.pt/pt-PT/OBancoeoEurosistema/ConsultasPublicas/Documents/Consulta_BP_2_07_MAR.pdf. Acesso em: 12 de Dezembro de 2020.

BASLER AUSSCHUSS FÜR BANKENAUF SICH. Range of Practice in Banks, Internal Rating Systems. **Diskussionspapier**, Publication, n. 66, 2000.

BEAVER, W. H. Financial ratios predictors of failure, **Journal of Accounting Research**, v. 4, n. 1, p. 71–111, 1966.

BELLOTTI, T.; CROOK, J. Modelling and estimating Loss Given Default for credit cards. Credit Research Centre Working Paper, **WP 08/1**, University of Edinburg, 2008.

BESSA, L. **Cadastro Positivo: Comentários à Lei 12.414 de 09 de junho de 2011**. 1ª ed. São Paulo: Revista dos Tribunais, 2011.

BESSIS, J. **Risk management in banking**. 30 ed. Jon Wiley & Sons, Chichester: UK, 2011.

BLATT, A. **Avaliação de risco e decisão de crédito: um enfoque prático**. São Paulo: Nobel, 1999.

BOAVISTA. **Sobre a BoaVista SCPC**. 2020. Disponível em: <https://www.boavistaservicos.com.br/sobre-a-boa-vista-scpc/>. Acesso em: 15 dez. 2020.

BONILLA, M.; OLMEDA, I.; PUERTAS, R. Modelos paramétricos y no paramétricos en problemas de credit scoring. **Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad**, v. 32, n. 118, p. 833-869, 2003.

BOTSMAN, R. Big data meets Big Brother as China moves to rate its citizens. **Wired UK**, v.21, n. 1, p. 1-11, 2017.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. **Classification and regression trees**. CRC press, 1984.

BRITO FILHO, D. A. **Aplicação de árvores de regressão aditivas bayesianas no desenvolvimento de modelos de escore de crédito no Brasil**, 2016. 65 f. Dissertação (Mestrado em Economia) - INSPER – Programa de Mestrado Profissional em Economia, São Paulo.

BUGNION, E.; NIEH, J.; TSAFRIR, D. Hardware and software support for virtualization. **Synthesis Lectures on Computer Architecture**, v. 12, n. 1, p. 1-206, 2017.

BULOS, U. L. Curso de direito constitucional. rev. e atual. de acordo com a Emenda Constitucional n. 76/2013. **São Paulo: Saraiva**, 2014.

CABRERA-SÁNCHEZ, J. P.; VILLAREJO-RAMOS, Á. F. Fatores que afetam a adoção de análises de *Big Data* em empresas. **Revista de Administração de Empresas**, v. 59, n. 6, p. 415-429, 2019.

CAOUCETTE, J. B.; ALTMAN, E. I.; NARAYANAN, P. **Managing credit risk: the next great financial challenge**. John Wiley & Sons, 1998.

CAVALCANTI, P. P. **Privacidade e proteção de dados pessoais: uma análise comparativa dos quadros regulatórios brasileiro e europeu**. 2018. 62 f. TCC (graduação em Direito) - Universidade Federal de Pernambuco - UFPE – Recife.

CERVANTES, V.; RODRIGUES, D. F. Big data e proteção de dados: o desafio está lançado, I Seminário Internacional de Concorrência e Inovação – FDRT/USP/NCI-PUC/SP, **Anais**, São Paulo, 2018.

CHAIA, A. J. **Modelos de gestão do risco de crédito e sua aplicabilidade ao mercado brasileiro**. 2003. 121 f. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo.

CHATTERJEE, S.; CORBAE, D.; DEMPSEY, K. P.; RÍOS-RULL, J. V. A quantitative theory of the credit score. **NBER Working Paper**, n. w27671, 2020.

CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. **MIS quarterly**, v. 35, n. 4, p. 1105-1155, 2012.

CHEN, Y.; CHEUNG, A. S. Y. The transparent self under big data profiling: Privacy and Chinese Legislation on the Social Credit System. **The Journal of Comparative Law**, v. 12, n. 2, p. 356–378 2017.

CHEREGATI, J. W. S. **Determinantes do score de crédito e tempo até inadimplência para empréstimos comerciais a pessoas físicas**. 2008. 131 f. Dissertação (Mestrado em Economia do Setor Público) - Universidade de Brasília, Brasília, 2008.

CHIAVENATO, J. J. **Ética globalizada & sociedade de consumo**. 2 ed.reform. São Paulo: Editora Moderna Ltda, 2004.

CONTADOR, J. L.; SENNE, E. L. F. Testes não paramétricos para pequenas amostras de variáveis não categorizadas: um estudo. **Gestão & Produção**, v. 23, p. 588-599, 2016.

COONEY, M. G. **how big trends in security, mobile, big data and cloud computing will change IT: a quick roundup of IT trends, from Android adoption to cloud security**. Network World, 2012. Disponível em: <https://www.networkworld.com/article/2161069/lan-wan/gartner--how-big-trends-in-security--mobile--big-data-and-cloud-computing-will-change-it.html>. Acesso em: 16 de Dezembro de 2020.

CORRÊA, M. F.; Machado, M. A. S. Construção de um Modelo de Credit Scoring em Redes Neurais para Previsão de Inadimplência na Concessão de Microcrédito. Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração, EnANPAD, **Anais**, Curitiba. De 25 a 29 de setembro de 2004.

CORTAZIO, R. S. Banco de dados no Brasil: Uma análise do sistema credit scoring à luz da lei n. 13.709/2018 (LGPD). **Revista Eletrônica da Procuradoria Geral do Estado do Rio de Janeiro**, v. 2 n. 3, p.23, 2019.

CORTES, F. P. **Gestão de Risco nas Instituições Financeiras: Uma Análise do Novo Acordo de Basiléia e Apresentação de Conceitos para Desenvolvimento de um Sistema de Informações Gerenciais**. Brasília: Fundação Getúlio Vargas, 2004.

COVELLO, S. C. Notas sobre os contratos bancários. **Revista de Direito Civil**, v. 45, n. 1, p. 110-123, 1998.

CPBR6. Campus Party. **Big Data e computação em nuvem**. 2013. Disponível em: https://www.youtube.com/watch?v=dOR3hGS_IZI. Acesso em: 10 de out de 2020.

CRAIG, G. J. O impacto da crise econômica asiática na Tailândia, **Gerencial Finance**, vol. 26, n. 4, p. 39-48, 2000.

CUNHA, M. E. P. F. A. **Os dados pessoais como o novo petróleo: a abordagem jurídica no tratamento de dados na era da informação**. 2018. 66 f. Trabalho de conclusão de curso (Bacharel em Direito) - Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro.

DA SILVA FILHO, F. G. L. Aplicação de Redes Neurais na Análise e na Concessão de Crédito ao Consumidor, **Revista de Administração**, v.44, n.1, p.34-45, 2009.

D'ACQUISTO, G.; DOMINGO-FERRER, J.; KIKIRAS, P.; TORRA, V.; MONTJOYE, Y. A.; BOURKA, A. **Privacy by Design in Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics**, European Union Agency for Network and Information Security (ENISA) Report, December 2015.

D'ALESSIO, G.; IEZZI, S. Household over-indebtedness definition and measurement with Italian data. **Bank of Italy Occasional Paper**, n.149, 2013.

D'ALESSIO, G.; IEZZI, S. Over-Indebtedness in Italy: How Widespread and Persistent Is It? **Bank of Italy Occasional Paper**, n.319, 2016.

DAVENPORT, T. H.; BARTH, P.; BEAN, R. How 'big data' is different. **MIT Sloan Management Review**, v. 54, n. 1, p. 22-24, 2012.

DEAN CAIRE, C. F. A. **Building Credit Scorecards for Small Business Lending in Developing Markets**, Bannock Consulting, 2004. Disponível em: http://www.microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf. Acesso em: 10 de out de 2020.

DE CARVALHO, A. L. L.; VILHENA, M. M. R. Proteção De Dados: Nova Dimensão Da Função Social Da Empresa. **Governança E Direitos Fundamentais**, 1ª ed., p. 84-92, 2020.

DE CNUUDE, S.; MOEYERSOMS, J.; STANKOVA, M.; TOBBACK, E.; JAVALY, V.; MARTENS, D. What does your Facebook profile reveal about your creditworthiness? Using alternative data for microfinance. **Journal of the Operational Research Society**, v. 70, n. 3, p. 353-363, 2019.

DEMCHENKO, Y.; GROSSO, P.; DE LAAT, C.; MEMBREY, P. Addressing big data issues in scientific data infrastructure. In: **2013 International Conference on Collaboration Technologies and Systems (CTS)**, IEEE, p.48-55, 2013.

DÓRIA, F. G. **Utilização de big data para a construção de features no desenvolvimento de modelos preditivos em análise de crédito**. 2019. 53 f. Trabalho de Conclusão de Curso (Graduação em Engenharia de produção Mecânica) - Universidade Federal de Santa Catarina, Santa Catarina.

DOYLE, A. C. **Sherlock Holmes: the complete novels and stories volume I**. New York: Bantam classics, 2013.

DUMBILL, E. Volume, Velocity, Variety: What You Need to Know About Big Data. **Forbes**. v. 19, 2012. Disponível em: <http://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/>. Acesso em: 10 de out de 2020.

DURAND, D. **Risk elements in consumer installment financing**. 1ª ed. New York: National Bureau of Economic Research, 1941.

ECONOMIST, T. **The world's most valuable resource is no longer oil, but data**. The Economist: New York, 2017.

FERGUSON, G. A. **Statistical analysis in psychology and education**. 6.ed. New York: McGraw-Hill book, 1989

FERREIRA, T. S. V.; COSTA, F. J. Big Data: Reflexões epistemológicas e impactos nos estudos de finanças e mercado de capitais. **Revista de Educação e Pesquisa em Contabilidade**, v. 11, n. 4, p. 396-407, 2017.

FIGUEIREDO, P. H. **Regressão logística usando o SPSS**. 2017. Disponível em: <http://lafiexufvjm.blogspot.com/2017/06/regressao-logistica.html>. Acesso em: 23 de jun de 2021.

FIGUEIREDO FILHO, D. B.; SILVA JUNIOR, J. A. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). **Revista Política Hoje**, v. 18, n. 1, p. 115-146, 2009.

FILHO, D. R. A diretiva europeia sobre proteção de dados pessoais - uma análise de seus Aspectos Gerais. **Lex magister**. 2013. Disponível em: http://www.lex.com.br/doutrina_24316822_A_DIRETIVA_EUROPEIA_SOBRE_PROTEC

AO_DE_DADOS_PESSOAIS_UMA_ANALISE_DE_SEUS_ASPECTOS_GERAIS.aspx.

Acesso em: 07 de Novembro de 2020.

FILOTTO U.; GIANNASCA. C. Credito al consumo: qualità del credito e gestione del rischio.

Banche e Banchieri, v.3, n.3, p. 241-250, 1996.

FLICK, U. **Introdução à metodologia de pesquisa: um guia para iniciantes**. Penso Editora, p. 167, 2013.

FREEDMAN, S.; JIN, G. Z. The information value of online social networks: Lessons from peer-to-peer lending. *International Journal of Industrial Organization*, v. 51, p. 185–222, mar. 2017.

FREIRE, Y. M. B. **Pontuação de crédito com uso de Big Data: 10 anos de revisão sistemática mediante enfoque meta analítico**. 2019. 44 f. Trabalho de Conclusão de Curso (Bacharelado em Administração) - Universidade de Brasília, Brasília.

FREITAS, B. S. J. A LGPD e os impactos sobre o tratamento de dados pessoais de adolescentes nas relações de trabalho. **Direito UNIFACS–Debate Virtual**, v. 1, n. 246, p. 1-15, 2020.

FREITAS, C. O. A.; MAFFINI, M. A proteção dos dados pessoais no crédito bancário e a Lei Geral de Proteção de Dados frente ao Cadastro Positivo. **Revista Jurídica Cesumar: Mestrado**, v. 20, n. 1, p. 11-28, 2020.

GALDINO, N. Big Data: Ferramentas e Aplicabilidade. In. Congresso de Engenharia, XIII SEGET - Simpósio de Excelência Em Gestão e Tecnologia, 65, **Anais**, São Paulo, 2016.

GARCIA, V. S. Gestão de risco de crédito e regulamentação: uma reflexão sobre o caso brasileiro. 2002. 105 f. Dissertação (Mestrado em Administração) - Fundação Getúlio Vargas – Escola de Administração de Empresas de São Paulo, São Paulo.

GESTEL, T. V. Bayesian kernel based classification for financial distress detection. **European Journal of Operational Research**, v. 172, n.3, p. 979-1003, 2006.

GOH. R. Y.; LEE, L. S. Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches. **Advances in Operation Research**, v. 2019, 2019.

GOLDMAN, A. KON, F.; JUNIOR, F. P.; POLATO, I.; PEREIRA, R. F. Apache hadoop: conceitos teóricos e práticos, evolução e novas possibilidades. XXXI Jornadas de atualizações em informática, **Anais**, Curitiba, de 16 a 19 de julho de 2012.

GOMES, H. S. **Lei Geral de Proteção de Dados (LGPD): Uma análise dos impactos da lei na cultura e tratamento de dados no Brasil**. 2019. 28 f. TCC (Tecnólogo em Análise e Desenvolvimento de Sistemas) – UNISUL, Universidade do Sul de Santa Catarina, Florianópolis.

GONZALEZ, R. C.; WOODS, R. E. Digital Image Processing Addison-Wesley. **Reading, Ma**, v. 2, 1992.

GOULART, G. Por uma visão renovada dos arquivos de consumo: Databrokers e score de crédito. **Revista de Direito do Consumidor**, v. 107, p. 1-26, 2016.

GREENE, W. Sample selection in credit-scoring models. **Japan and the world economy**, v. 10, n. 3, p. 299-316, 1998.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. Por ed. [S.l.]: [s.n.], 2011.

HAND D. J.; HENLEY, W. E. Can Reject inference even work? **IMA Journal Mathematics Applied in Business**, v. 5, n. 1, p. 45-55, 1993.

HAND, D. J.; MCCONWAY, K. J.; STANGHELLINI, E. Graphical models of applicants for credit. **IMA Journal of Management Mathematics**, v. 8, n. 2, p. 143-155, 1997.

HAYKIN, S. **Redes neurais: princípios e prática**. 2ª ed. São Paulo: Bookman, 2007.

HELFRICH, M. **Kreditscoring und Scorewertbildung der SCHUFA**. 1ª ed. Germany: Nomos, 2010.

HOFFMANN, R. **Análise de regressão uma introdução à econometria**. 5. ed. Piracicaba: Portal de livros abertos da USP, 393p. 2016.

HOUAISS A.; VILLAR, M. S.; FRANCO, F. M. M. **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro: Objetiva, 2004.

HUANG, C. L.; CHEN, M. C.; WANG, C. J. Credit scoring with a data mining approach based on support vector machines. **Expert systems with applications**, v. 33, n. 4, p. 847-856, 2007.

HURLEY, M.; ADEBAYO, J. **Credit scoring in the era of big data**. **Yale JL & Tech.**, v. 18, n. 1, p. 1-70, 2017.

HURWITZ, J. S.; NUGENT, A.; HALPER, F.; KAUFMAN, M. **Big data for dummies**. New Jersey: John Wiley & Sons, 2013.

IBM. **What is Big Data analytics?** 2008. Disponível em <http://www.ibm.com/big-data/us/en/>. Acesso em 18 de Dezembro de 2020.

JIMÉNEZ, G.; SAURINA, J. Credit Cycles, Credit Risk, and Prudential Regulation. **International Journal of Central Banking**, v. 2, n. 2, p. 65-98, 2006.

JUNIOR, E. C. B.; DE ALBUQUERQUE, A. C. A busca por ferramentas de controle terminológico na esfera da Arte em ambiente web: sua valia para unidades de informação. In: X Encontro Internacional de Informação, Conhecimento e Ação. **Anais**, São Paulo, de 18 a 23 de junho de 2018.

KAY, J. W.; TITTERINGTON, D. M. **Statistics and neural networks: advances at the interface**. New York: Oxford University Press, 1999.

KHASHMAN, A. Credit risk evaluation using neural networks: Emotional versus conventional models. **Applied Soft Computing**, v. 11, n. 8, p. 5477-5484, 2011.

KOMORAD, K. **On credit scoring estimation**. Master of Science thesis, Institute for Statistics and Econometrics, Humboldt University, Berlin, 2002.

LACHTERMACHER, G.; ESPENCHITT, D. G. Previsão de falência de empresas: estudo de generalização de redes neurais. In: XXV Encontro da Associação de Pós-Graduação em Administração. **Anais**, Campinas, 2001.

LEE, T. S.; CHEN, I. F. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. **Expert Systems with Applications**, v. 28, n. 4, p. 743-752, 2005.

LEE, T. S.; CHIU, C. C.; LU, C. J.; CHEN, I. F. Credit scoring using the hybrid neural discriminant technique, **Expert Systems with Applications**, v. 23, n. 3, p. 245-254, 2002.

LEMOS, G. S. **Alinhamento de Sequências na Avaliação de Resultados de Teste de Robustez**. 2012. 125 f. Tese (Doutorado Ciência da Computação) - Universidade Estadual de Campinas – UNICAMP, Campinas.

LENARD, M. J.; ALAM, P.; MADEY, G. R. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. **Decision Sciences**, v. 26, n. 2, p. 209-227, 1995.

LENSBERG, T.; EILIFSEN, A.; MCKEE, T. E. Bankruptcy theory development and classification via genetic programming. **European Journal of Operational Research**, v. 169, n. 2, p. 677-697, 2006.

LEOPARDI, M. T. Metodologia da pesquisa na saúde. 2ª edição. **Florianópolis: UFSC/Pósgraduação em Enfermagem**, 2002.

LERMAN, J. Big Data and Its Exclusions. **Stanford Law Review**. v. 66, n. 1, p. 55, 2013.

LIBERATI, C.; CAMILLO, F.; SAPORTA, G. Advances in credit scoring: combining performance and interpretation in kernel discriminant analysis. **Advances in Data Analysis and Classification**, v. 11, n. 1, p. 121-138, 2017.

LIMA, F. G.; KIMURA H. Aplicação de redes neurais na análise e na concessão de crédito ao consumidor. **Revista de Administração - RAUSP**, v. 44, n. 1, p. 34-45, 2009.

LIN M. P. Julgando os mutuários pela empresa que mantêm: redes de amizade e informações assimetria de informação em empréstimos online ponto a ponto. **Management Sci**, v. 59, n. 1 p.17–35, 2013.

MALHOTRA, N. K. **Pesquisa de Marketing: Uma orientação aplicada**. (7ª. ed.). Porto Alegre: Bookman Editora, 2019.

MALHOTRA, R.; DAVINDER, K. Differentiating between good creditsand badcredits using neuro-fuzzy systems. **European Journal of Operational Research**, v. 136, n. 1, p.190-211, 2002.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. **Big data: The next frontier for innovation, competition, and productivity**. Washington: McKinsey Global Institute, 2011.

MARINHO, F. **A Adequação da LGPD nas empresas**. 2019. Disponível em: <https://www.linkedin.com/pulse/adequa%C3%A7%C3%A3o-da-lgpd-nas-empresas-fernando-marinho/>, Acesso em: 19 de Dezembro de 2020.

MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. Explorando o comportamento dos classificadores básicos em conjuntos de pontuação de crédito. **Sistemas especialistas com aplicativos**, v. 39, n. 11, p. 10244-10250, 2012.

MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. A literature review on the application of evolutionary computing to credit scoring. **Journal of the Operational Research Society**, v. 64, n. 9, p. 1384-1399, 2013.

MARTINS, M. G.; JORGETTO, L. F. M. R. G.; SUTTI, A. C. A. Big Data e a proteção do direito à privacidade no contexto da sociedade da informação. **Revista Jurídica Cesumar: Mestrado**, v. 19, n. 3, p. 705-725, 2019.

MATOS, O. C. **Econometria básica: teoria e aplicações**. 3 ed. São Paulo: Atlas, 2000.

MCAFEE, A.; BRYNJOLFSSON, E. Big data: The management revolution. **Harvard Business Review**, v. 90, n. 10, p. 60-68, 2012.

MEEAMPOL. S. LERSKULLAWAT, P.; WONGSORNTHAM, A.; SRINAMMUANG, P.; RODPETCH, V.; NOONOI, R. Applying Emerging Market Z-Score Model To Predict Bankruptcy: A Case Study Of Listed Companies In The Stock Exchange Of Thailand. In: Management, **Knowledge na Learning, International Conference**, Slovenia: Toknowpress, 2014.

MENDES, C.M. A pesquisa online: potencialidades da pesquisa qualitativa no ambiente virtual. **Hipertextus Revista Digital**. n.2, 2009.

MENDES, G. F.; BRANCO, P. G. G. **Curso de Direito Constitucional**, 12^a ed. São Paulo: Saraiva, 2017.

MEYER, M. **Lei 13.709/2018: Lei de Proteção de Dados Pessoais**, 2018. Disponível em: https://www.machadomeyer.com.br/images/publicacoes/PDFs/Lei_Protecao_de_Dados_ebook_18.pdf. Acesso em: 19 de dezembro de 2020.

MONTEIRO, R. L. Existe um direito à explicação na Lei Geral de Proteção de Dados do Brasil. **Artigo Estratégico**, v. 39, n. 1, p. 1-14, 2018.

MUCELIN, G. Peers Inc.: a nova estrutura da relação de consumo na economia do compartilhamento. **Revista de Direito do Consumidor**. vol. 118. ano 27. p. 77-126, 2018.

NADER, P. **Curso de Direito Civil: Contratos**. 5 ed. Rio de Janeiro: Forense; 2010.

NEOWAY. **O que é Big Data e qual a importância de implementá-lo na empresa**. 2019. Disponível em: <https://www.neoway.com.br/o-que-e-big-data/>. Acesso em: 15 de Novembro de 2020.

NETTO, A. S. D. P.; MORO, E. P.; FERREIRA, F. F. **Big Data e suas influências sobre a estratégia das empresas**. Curso de Engenharia de Controle e Automação, Universidade Federal do Rio de Janeiro, UFRJ, 2014. Disponível em: https://www.gta.ufrj.br/grad/15_1/bigdata/index.html. Acesso em: 21 de Novembro de 2020.

NEVES, J. C.; SILVA, J. A. **Análise do Risco de Incumprimento: na Perspectiva da Segurança Social**. Segurança portuguesa, 1998.

NOGUEIRA, L. C. B. **Big Data: ranking dos cidadãos como instrumento de controle social**. 2020. 97 f. Dissertação (Mestrado em Ciências da comunicação) – Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas, Lisboa.

NASCIMENTO, P. S.; KOMATI, K. S.; ANDRADE, J. O. Modelo Preditivo para Avaliação de Crédito em Empréstimos Pessoais. Em: Escola Regional de Informática de Goiás (ERI-GO), 7, Goiânia. VII Escola Regional de Informática de Goiás. **Anais...** de 22 a 23 de novembro de 2019.

OLIVA, M. D; VIÉGAS, F. A. Tratamento de dados para a concessão de crédito. In: FRAZÃO, A.; TEPEDINO, G.; OLIVA, M. D. **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thonson Reuters Brasil, 2019.

OLIVEIRA, D. D. P. R. D. **Sistemas, organização e métodos: Uma Abordagem Gerencial**. 13ª ed. São Paulo: Atlas, 2002.

O'NEIL, C. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. New York: Crown, 2016.

ONG, M. K. **Credit Ratings: methodologies, rationale and default risk**. London: Risk Books, 2002.

ORGLER, Y. E. Evaluation of bank consumer loans with credit scoring models. **Journal of Bank Research**, v. 2, n. 1, p. 31-37, 1971.

PALMUTI, C. S.; PICHIATTI, D. Mensuração do risco de crédito por meio de análise estatística multivariada, **Revista Economia Ensaios**, v. 26, n. 2, p. 7-22, 2004.

PASQUALE, F. A. **The black box society: the secret algorithms that control money and information**. Cambridge: Harvard University Press, 2016.

PENNONI. L. F, STANGHELLINI. E. Metodi Statistici Multivariati Applicati All'Analisi del Comportamento dei Titotalari di Carta di Credito di Tipo Revolving, 1999. 122 f. Tesi (Specializzazione Economia e Commercio) - **Università Degli Studi di Perugia**, Facoltà di Economia Corso di Laurea in Economia e Commercio, Itália.

PEREIRA, A. R.; PALMUTI, C. S.; JUNIOR, P. S. P.; SILVA, S. W.; ALVES, A. F. A influência dos índices setoriais da Bolsa de Valores sobre o índice Ibovespa: uma análise estatística-econométrica para o período 2015 e 2016. **Revista Debate Econômico**, v. 6, n. 2, p. 63-85, 2018.

PLANALTO. **LEI Nº 12.965, DE 23 DE ABRIL DE 2014**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Acesso em: 15 de dezembro de 2020.

PLANALTO. **LEI Nº 13.709, DE 14 DE AGOSTO DE 2018**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 15 de dezembro de 2020.

POINTON, J. Intelligent Systems in Accounting. **Finance and Management**, v.18, n.2-3, p. 59-88, 2011.

POLETTINI, M. R. N. F. A LGPD e os impactos nas relações de consumo. **Revista JurisFIB**, v. 10, n. 2, P. 131-138, 2020.

POSPIECH, M.; FELDEN, C. A descriptive big data model using grounded theory. **Computational Science and Engineering (CSE), IEEE 16th International Conference on** 878–885, 2013.

QUOD. **Sobre nós**. 2020. Disponível em: <https://www.quod.com.br/a-quod>. Acesso em: 15 dez. 2020.

QUOD. **Conheça como é a análise de crédito da Quod por dentro**. 2019. Disponível em: <https://www.quod.com.br/lp/analise-credito#planos>. Acesso em: 15 dez. 2020.

RAMOS, R. V. R. **Modelo de análise de risco de crédito**. 2016. 39 f. Relatório de estágio (Mestrado em Finanças) - Faculdade de Economia do Porto, Portugal.

RAPÔSO, C. F. L.; LIMA, H. M.; JUNIOR, W. F. O.; SILVA, P. A. F.; BARROS, E. E. S. LGPD - Lei Geral de Proteção de Dados Pessoais em Tecnologia da Informação: Revisão Sistemática. **RACE - Revista de Administração do Cesmac**, v. 4, n. 1, p. 58-67, 2019.

REGMI, P.R; et. al. Guide to the design and application of online questionnaire surveys; **Nepal Journal of Epidemiology**. v.6 n.4, p.640- 644. 2016

REGULAMENTO GERAL DE PROTEÇÃO DE DADOS (UE) 2016/679 do **Parlamento Europeu e do Conselho**, de 27 de abril de 2016, JO L 119 de 4.5.2016, p. 1—88. Disponível em: <http://data.europa.eu/eli/reg/2016/679/oj>. Acesso em: 01 de Janeiro de 2021.

RIVES-LANGE, J. L. **“Droit bancaire”**. 6. ed. Paris: Dalloz, 1995.

ROCHA, S. O. N.; MAIA, N. A.; JUNIOR, A. B. C.; DIAS, M. P. A.; VELOSO, R. R. Utilização de redes neurais para estimativa da resistência à compressão do concreto simples. **Brazilian Journal of Development**, v. 6, n. 10, p. 79910-79922, 2020.

ROSA, P. **Modelos de Credit scoring: Regressão Logística, CHAID e REAL**. São Paulo, 2000. Tese de Doutorado. Dissertação de Mestrado (Mestrado em Estatística).

SAMARAKOON, L.; HASAN, T. Altman's Z-Score Models of Predicting Corporate Distress: Evidence from the Emerging Sri Lankan Stock Market. **Journal Finance Academy**, v. 1, n. 1, p. 119-125, 2003.

SANTIAGO, V. C.; TAMBA, D. H. Proteção de dados no Brasil: novo marco regulatório. **Migalhas de Peso**, 2018. Disponível em: <https://www.migalhas.com.br/depeso/290866/protecao-de-dados-no-brasil--novo-marco-regulatorio>. Acesso em: 16 de Dezembro de 2020.

SANTOS, I. C. J. **Desenvolvimento infantil e parentalidade: conhecimento de gestantes cadastradas em Unidades Básicas de Saúde**. Diamantina, 2021.

SAUNDERS, A. **Medindo o risco de crédito: novas abordagens para value at risk e outros paradigmas**. Rio de Janeiro: Qualitymark, 2000.

SARTORIS, A. **Estatística e introdução à econometria**. 1ª Ed. São Paulo: Saraiva, 2003.

SCHIMINKOSKI, S.; TAGLIAPIETRA, O. M.; BERTOLIN, G. R. F. As influências da responsabilidade social praticada pela Serasa Experian na vida dos ex-colaboradores da agência de Cascavel, 4º Fórum Internacional Inovar, **Anais...** Rio Grande do Sul, de 26 a 28 de agosto de 2015.

SEBBEN, A. N. **O direito à proteção de dados pessoais e o mercado de crédito no Brasil: análise à luz da lei nº 13.709/2018 (lgpd) e da lei complementar nº 166/2019**. 2020. 143 f. Dissertação (Mestrado em Direito) - Universidade de Marília, UNIMAR, Rio de Janeiro.

SECURATO, J. R. **Crédito: Análise e Avaliação de Risco – Pessoas Físicas e Jurídicas**. 2ªed, São Paulo: Saint Paul, 2012.

SILVA, F. A. S. **Credit Scoring as an Asset for Decision Making in Intelligent Support Systems**. 2011. 102 f. Dissertação (Mestrado em Engenharia Informática) - Universidade do Minho, Portugal.

SILVA, J. P. **Gestão e análise de risco de crédito**. São Paulo: Atlas, 2000.

SILVA, W. M.; NAKAMURA, W. T.; MORAES, D. C. Comportamento de risco de cartão de crédito em campus universitários: evidências do Brasil. **BAR-Revista da Administração Brasileira**, v. 9, n. 3, p. 351-373, 2012.

SILVEIRA, S. A. **Tudo sobre tod@s: redes digitais, privacidade e venda de dados pessoais**. São Paulo: Edições Sesc SP, 2017.

SILVEIRA, M.; MARCOLIN, C. B.; FREITAS, H. O big data e seu uso corporativo: uma revisão de literatura. IV SINGEP - Simpósio Internacional de Gestão de Projetos, Inovação e Sustentabilidade, **Anais...**São Paulo, de 8 a 10 de novembro de 2015.

SIMON, P. **Too Big to Ignore: The business case for Big Data**. Hoboken: John Wiley & sons Inc, 2013.

SIVARAJAH, U.; KAMAL, M. M.; IRANI, Z.; WEERAKKODY, V. Critical analysis of Big Data challenges and analytical methods. **Journal of Business Research**, v. 70, n. 1, p. 263-286, 2017.

SOUZA, V. B. P.; MUCHERONI, M. L.; MASSENA, C. A. Evolução do tratamento e coleta de dados na ciência: E-science, Big Data e Big Science. XIX Encontro Nacional De Pesquisa Em Ciência Da Informação – ENANCIB, **Anais...** Paraná, de 22 a 26 de outubro de 2018.

SPC BRASIL. **Conheça a Loja SPC Brasil**. 2020. Disponível em: <https://loja.spcbrasil.org.br/sobre>. Acesso em: 15 dez. 2020.

STEENACKERS, A., GOOVAERTS, Marc. A credit scoring model for personal loans. **Insurance: Mathematics and Economics**, v. 8, n. 1, p.31-34, 1989.

STJ. **Recurso Especial: Nº 1.457.199 - RS (2014/0126130-2)**. Relator: Ministro Paulo de Tarso Sanseverino. DJ: 17/12/2014. **JusBrasil**, 2014. Disponível em: <https://bdjur.stj.jus.br/jspui/bitstream/2011/114173/REsp1457199.pdf>. Acesso em: 25 dez. 2020.

SZINVELSKI, M. M.; ARCENO, T. S.; FRANCISCO, L. B. Perspectivas jurídicas da relação entre big data e proteção de dados. **Perspectivas em Ciência da Informação**, v. 24, n. 4, p. 132-144, 2019.

TAURION, C. **Big Data**. Rio de Janeiro: Brasport, 2013.

TELES, A. C. **Pontuação de crédito com base em Big Data**. 2017. 62 f. Trabalho de conclusão de curso (Bacharelado em Direito) Universidade de Brasília, Brasília, 2017.

THOMAS, L.; CROOK, J.; EDELMAN, D. **Credit Scoring and its Applications**. 2ª ed. Philadelphia: Society for Industrial and Applied Mathematics; 2017.

TIMOSSI, L. S.; FRANCISCO, A. C.; JUNIOR, G. S.; XAVIER, A. A. P. Análise da qualidade de vida no trabalho de colaboradores com diferentes níveis de instrução através de uma análise de correlações. **Production**, v. 20, n. 3, p. 471-480, 2010.

TOLE, A. A. “Big Data Challenges”. **Database Systems Journal**, v. 4, n. 3, p. 31-40, 2013.

TRIOLA, M. F. **Introdução à estatística**. 9. ed. Rio de Janeiro: LTC, 2005.

TRIPPI, R. R.; TURBAN, E. **Neural networks in finance and investing: Using artificial intelligence to improve real world performance**. McGraw-Hill, Inc., 1992.

URAPEEPATANAPONG, K.; SETHSATHIRA, S.; OKANURAK, C. Nova lei de falências para impulsionar a economia tailandesa. **Revisão do Direito Financeiro Internacional**, v. 17, n. 4, p. 33, 1998.

VAN DER AALST, W. M. P. Data scientist: The engineer of the future. In: **Enterprise interoperability VI**. Springer, Cham, 2014.

VARIAN, H. R. **Microeconomia - Princípios Básicos: Uma abordagem moderna**. Rio de Janeiro: Elsevier Brasil, 2006.

VICENTE, J. I. C. **Modelo de Scoring para o crédito à Habitação: Um estudo de caso em Portugal**. 2015. 77 f. Dissertação (Mestrado em Administração) - Instituto Superior de Contabilidade e Administração de Lisboa, Portugal.

VILLELA, T. C. M. **Lei Geral de Proteção de Dados e a atividade bancária: base legal do interesse legítimo e sua aplicação**. 2019. 30 f. TCC (Pós-Graduação em Direito empresarial) – Instituto de Ensino Superior e Pesquisa – INSPER, São Paulo.

VON ZUBEN, F. J. **Modelos paramétricos e não paramétricos de redes neurais artificiais e aplicações**. 1996. 243 f. Dissertação (Doutorado em Engenharia Elétrica) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica, Departamento de Engenharia de Computação e Automação Industrial, São Paulo.

WANG, C. J.; CHUN, N. G. Y.; BROOK, R. H. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. **Jama**, v. 323, n. 14, p. 1341-1342, 2020.

WASSLER, P.; TOLKACH, D. Orwellian Tourism 2020? China's Social Credit Score. In: **2019 TTRA Europe Chapter Conference**, 8-10 April 2019, Bournemouth University, UK.

WESTON, J. F.; BRIGHAM, E. F. **Fundamentos da administração financeira**. 10ªed. São Paulo: Makron Books, 2000.

YANAKA, G. M.; HOLLAND, M. Basileia II e exigência de capital para risco de crédito dos bancos no Brasil. **Revista brasileira de finanças**, v. 8, n. 2, p. 167-195, 2010.

YANG Y. Adaptive credit scoring with kernel learning methods. **European Journal of Operational Research**, v. 183, n. 3, p. 1521-1536, 2007.

ZEKIC-SUSAC, M.; SARLIJA, N.; BENSIC, M. Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models. In: **26th International Conference on Information Technology Interfaces, IEEE**. June 7 – 10, 2004

ZENDERSKY, H. C. **Gerenciamento de resultados em instituições financeiras no Brasil – 2000 a 2004**. 2005. 134 f. Dissertação (Mestrado em Ciências Contábeis) - Universidade de Brasília, Brasília, 2005.

ZIKOPOULOS, P.; LIGHTSTONE, S.; HURAS, M.; SACHEDINA, A. New dynamic in-memory analytics for the era of big data. **IBM Data Management Magazine**, v. 4, n. 1, p. 1-47, 2013.

ZULIANI, M. S. S. **Score de crédito só pode ser judicializado se empréstimo foi negado**. Tribunal de Justiça do Distrito Federal e dos Territórios, TJDFT, Brasília, BSB. 2018. Disponível em: <https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2017/score-de-credito-so-pode-ser-judicializado-se-emprestimo-foi-negado-juiz-matheus-stamillo-santarelli-zuliani>. Acesso em: 11 de Outubro de 2020.

APÊNDICE A – CÓDIGO-FONTE R (MODELO 1)

```

### Regressão Logística Binária ###
#####

### Modelo 1:

###  $Y(\text{Uso Score Cred}) = A + B1(\text{Agilidade}) + B2(\text{Assertividade}) +$ 
###  $B3(\text{Segurança}) + B4(\text{Credibilidade}) + B5(\text{Facilidade}) +$ 
###  $B6(\text{Preço do Serviço}) + E$ 

### Carregamento dos pacotes utilizados
#####

if(!require(pacman)) install.packages("pacman")
library(pacman)
pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

options(scipen = 999)

### Carregamento da base de dados
#####

# Respondentes com apenas dívidas da empresa

dados_PJ <- read.csv2('./Dados/PJ/Mod1/mod1_PJ.csv',
                     stringsAsFactors = TRUE)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF <- read.csv2('./Dados/PJ_PF/Mod1/mod1_PJ_PF.csv',
                          stringsAsFactors = TRUE)

### Análise exploratória
#####

# Respondentes com apenas dívidas da empresa

View(dados_PJ)
table(dados_PJ$usa_score)
summary(dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

View(dados_PJ_PF)
table(dados_PJ_PF$usa_score)
summary(dados_PJ_PF)

### Categorias de referência
#####

# Respondentes com apenas dívidas da empresa

levels(dados_PJ$usa_score)
levels(dados_PJ$agilidade)
levels(dados_PJ$assertividade)
levels(dados_PJ$seguranca)

```

```

levels(dados_PJ$credibilidade)
levels(dados_PJ$facilidade)
levels(dados_PJ$preco)

# Respondentes com dívidas da empresa e dívidas pessoais

levels(dados_PJ_PF$usa_score)
levels(dados_PJ_PF$agilidade)
levels(dados_PJ_PF$assertividade)
levels(dados_PJ_PF$seguranca)
levels(dados_PJ_PF$credibilidade)
levels(dados_PJ_PF$facilidade)
levels(dados_PJ_PF$preco)

### Construção do modelo com todas as preditoras
#####

# Respondentes com apenas dívidas da empresa

modelo_PJ <- glm(usa_score ~ .,
                family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo_PJ_PF <- glm(usa_score ~ .,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Construção dos modelos com as preditoras candidatas/selecionadas
#####
###

# Respondentes com apenas dívidas da empresa

modelo1_PJ <- glm(usa_score ~ agilidade,
                 family = binomial(link = 'logit'), data = dados_PJ)

modelo2_PJ <- glm(usa_score ~ assertividade,
                 family = binomial(link = 'logit'), data = dados_PJ)

modelo3_PJ <- glm(usa_score ~ credibilidade,
                 family = binomial(link = 'logit'), data = dados_PJ)

modelo4_PJ <- glm(usa_score ~ assertividade + credibilidade,
                 family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo1_PJ_PF <- glm(usa_score ~ assertividade,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Pressupostos para a Regressão Logística:
#####

##1. (variável dependente dicotômica - mutuamente exclusiva)
##2. (independência das observações - sem medidas repetidas)

##3. (ausência de outliers - limites dentro de [-3, 3])

# Respondentes com apenas dívidas da empresa

```

```

summary(stdres(modelo_PJ))
summary(stdres(modelo1_PJ))
summary(stdres(modelo2_PJ))
summary(stdres(modelo3_PJ))
summary(stdres(modelo4_PJ))

# Respondentes com dívidas da empresa e dívidas pessoais

summary(stdres(modelo_PJ_PF))
summary(stdres(modelo1_PJ_PF))

##4. (ausência de multicolinearidade entre as covariáveis - [VIF < 10])

# Respondentes com apenas dívidas da empresa

vif(modelo_PJ)
vif(modelo4_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

vif(modelo_PJ_PF)

### Análise do modelo

##(efeitos gerais)

# Respondentes com apenas dívidas da empresa

Anova(modelo_PJ, type = "II", test = "Wald")
Anova(modelo1_PJ, type = "II", test = "Wald")
Anova(modelo2_PJ, type = "II", test = "Wald")
Anova(modelo3_PJ, type = "II", test = "Wald")
Anova(modelo4_PJ, type = "II", test = "Wald")

# Respondentes com dívidas da empresa e dívidas pessoais

Anova(modelo_PJ_PF, type = "II", test = "Wald")
Anova(modelo1_PJ_PF, type = "II", test = "Wald")

##(efeitos específicos)

# Respondentes com apenas dívidas da empresa

summary(modelo_PJ)
summary(modelo1_PJ)
summary(modelo2_PJ)
summary(modelo3_PJ)
summary(modelo4_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

summary(modelo_PJ_PF)
summary(modelo1_PJ_PF)

### Razão de chance com IC 95% (usando erro padrão)

# Respondentes com apenas dívidas da empresa

exp(cbind(Odds_Ratio = coef(modelo2_PJ), confint.default(modelo2_PJ)))
exp(cbind(Odds_Ratio = coef(modelo3_PJ), confint.default(modelo3_PJ)))

```

```
# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

### Avaliação da qualidade e comparação entre modelos
##(pseudo R-quadrado)

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

##(comparação de modelos - AIC e BIC)

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

##(qui-quadrado - modelos são diferentes com sig. < 0,05)

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

##(tabela de classificação)

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo
```


APÊNDICE B – CÓDIGO-FONTE R (MODELO 2)

```

### Regressão Logística Binária ###
#####

### Modelo 2:

###  $Y(\text{Custo} \times \text{Benf}) = A + B1(\text{Treinamento}) + B2(\text{Red Inadm}) +$ 
###  $B3(\text{Aumt Vend Parcel}) + E$ 

### Carregamento dos pacotes utilizados
#####

if(!require(pacman)) install.packages("pacman")
library(pacman)
pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

options(scipen = 999)

### Carregamento da base de dados
#####

# Respondentes com apenas dívidas da empresa

dados_PJ <- read.csv2('./Dados/PJ/Mod2/mod2_PJ.csv', stringsAsFactors =
TRUE)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF <- read.csv2('./Dados/PJ_PF/Mod2/mod2_PJ_PF.csv',
stringsAsFactors = TRUE)

### Análise exploratória
#####

# Respondentes com apenas dívidas da empresa

View(dados_PJ)
table(dados_PJ$custo_x_beneficio)
summary(dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

View(dados_PJ_PF)
table(dados_PJ_PF$custo_x_beneficio)
summary(dados_PJ_PF)

### Categorias de referência
#####

# Respondentes com apenas dívidas da empresa

levels(dados_PJ$custo_x_beneficio)
levels(dados_PJ$treinamento)
levels(dados_PJ$red_inadimplencia)
levels(dados_PJ$aum_vend_parcelada)

```

```

# Respondentes com dívidas da empresa e dívidas pessoais

levels(dados_PJ_PF$custo_x_beneficio)
levels(dados_PJ_PF$treinamento)
levels(dados_PJ_PF$red_inadimplencia)
levels(dados_PJ_PF$aum_vend_parcelada)

### Construção do modelo com todas as preditoras
#####

# Respondentes com apenas dívidas da empresa

modelo_PJ <- glm(custo_x_beneficio ~ .,
                 family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo_PJ_PF <- glm(custo_x_beneficio ~ treinamento + red_inadimplencia,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Construção dos modelos com as preditoras (sig.) avaliadas pelo qui-
quadrado
#####
###

# Respondentes com apenas dívidas da empresa

modelo1_PJ <- glm(custo_x_beneficio ~ treinamento,
                 family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

### Pressupostos para a Regressão Logística:
#####

##1. (variável dependente dicotômica - mutuamente exclusiva)
##2. (independência das observações - sem medidas repetidas)

##3. (ausência de outliers - limites dentro de [-3, 3])

# Respondentes com apenas dívidas da empresa

summary(stdres(modelo_PJ))
summary(stdres(modelo1_PJ))

# Respondentes com dívidas da empresa e dívidas pessoais

summary(stdres(modelo_PJ_PF))

##4. (ausência de multicolinearidade entre as VIs - [VIF < 10])

# Respondentes com apenas dívidas da empresa

vif(modelo_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

vif(modelo_PJ_PF)

```

```

##5. (relação linear entre cada VI contínua e o logito da VD)
##(teste de Box-Tidwell: interação entre a VI contínua e seu log - [sig. >
0.05])

#--> não é o caso neste modelo

### Análise do modelo

##(efeitos gerais)

# Respondentes com apenas dívidas da empresa

Anova(modelo_PJ, type = "II", test = "Wald")
Anova(modelo1_PJ, type = "II", test = "Wald")

# Respondentes com dívidas da empresa e dívidas pessoais

Anova(modelo_PJ_PF, type = "II", test = "Wald")

##(efeitos específicos)

# Respondentes com apenas dívidas da empresa

summary(modelo_PJ)
summary(modelo1_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

summary(modelo_PJ_PF)

### Razão de chance com IC 95% (usando erro padrão)

# Respondentes com apenas dívidas da empresa

exp(cbind(Odds_Ratio = coef(modelo_PJ), confint.default(modelo_PJ)))
exp(cbind(Odds_Ratio = coef(modelo1_PJ), confint.default(modelo1_PJ)))

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

### Avaliação da qualidade e comparação entre modelos

##(pseudo R-quadrado - quanto de variação da variável dependente está sendo
explicada pelo modelo)

# Respondentes com apenas dívidas da empresa

PseudoR2(modelo_PJ, which = "Nagelkerke")
PseudoR2(modelo1_PJ, which = "Nagelkerke")

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(comparação de modelos - AIC e BIC - diferença de pelo menos 10 [quanto
menor, melhor])

# Respondentes com apenas dívidas da empresa

```

```
AIC(modelo_PJ, modelol_PJ)
BIC(modelo_PJ, modelol_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(qui-quadrado - modelos são diferentes com [sig. < 0,05])

# Respondentes com apenas dívidas da empresa
anova(modelol_PJ, modelo_PJ, test = "Chisq")

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(tabela de classificação)

# Respondentes com apenas dívidas da empresa

ClassLog(modelo_PJ, dados_PJ$custo_x_beneficio)
ClassLog(modelol_PJ, dados_PJ$custo_x_beneficio)

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo
```

APÊNDICE C – CÓDIGO-FONTE R (MODELO 3)

```

### Regressão Logística Binária ###
#####

### Modelo 3:

###  $Y(\text{Prática LGPD}) = A + B1(\text{Seg Dados Cad}) + B2(\text{Treinamento Func}) +$ 
###  $B3(\text{Invest Seg Dados}) + B4(\text{Vazamt Dados}) + E$ 

### Carregamento dos pacotes utilizados
#####

if(!require(pacman)) install.packages("pacman")
library(pacman)
pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

options(scipen = 999)

### Carregamento da base de dados
#####

# Respondentes com apenas dívidas da empresa

dados_PJ <- read.csv2('./Dados/PJ/Mod3/mod3_PJ.csv', stringsAsFactors =
TRUE)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF <- read.csv2('./Dados/PJ_PF/Mod3/mod3_PJ_PF.csv',
stringsAsFactors = TRUE)

### Análise exploratória
#####

# Respondentes com apenas dívidas da empresa

View(dados_PJ)
table(dados_PJ$ usa_LGPD)
summary(dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

View(dados_PJ_PF)
table(dados_PJ_PF$ usa_LGPD)
summary(dados_PJ_PF)

### Categorias de referência
#####

# Respondentes com apenas dívidas da empresa

levels(dados_PJ$usa_LGPD)
levels(dados_PJ$segurança)
levels(dados_PJ$treinamento)
levels(dados_PJ$investimentos)
levels(dados_PJ$vazamentos)

```

```

# Respondentes com dívidas da empresa e dívidas pessoais

levels(dados_PJ_PF$usa_LGPD)
levels(dados_PJ_PF$segurança)
levels(dados_PJ_PF$treinamento)
levels(dados_PJ_PF$investimentos)
levels(dados_PJ_PF$vazamentos)

### Construção do modelo com todas as preditoras
#####

# Respondentes com apenas dívidas da empresa

modelo_PJ <- glm(usa_LGPD ~ .,
                 family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo_PJ_PF <- glm(usa_LGPD ~ .,
                    family = binomial(link = 'logit'), data = dados_PJ_PF)

### Construção dos modelos com as preditoras (sig.) avaliadas pelo qui-
quadrado
#####
###

# Respondentes com apenas dívidas da empresa

modelo1_PJ <- glm(usa_LGPD ~ treinamento,
                  family = binomial(link = 'logit'), data = dados_PJ)

modelo2_PJ <- glm(usa_LGPD ~ investimentos,
                  family = binomial(link = 'logit'), data = dados_PJ)

modelo3_PJ <- glm(usa_LGPD ~ vazamentos,
                  family = binomial(link = 'logit'), data = dados_PJ)

modelo4_PJ <- glm(usa_LGPD ~ treinamento + investimentos,
                  family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

### Pressupostos para a Regressão Logística:
#####

##1. (variável dependente dicotômica - mutuamente exclusiva)
##2. (independência das observações - sem medidas repetidas)

##3. (ausência de outliers - limites dentro de [-3, 3])

# Respondentes com apenas dívidas da empresa

summary(stdres(modelo_PJ))
summary(stdres(modelo1_PJ))
summary(stdres(modelo2_PJ))
summary(stdres(modelo3_PJ))
summary(stdres(modelo4_PJ))

```

```

# Respondentes com dívidas da empresa e dívidas pessoais

summary(stdres(modelo_PJ_PF))

##4. (ausência de multicolinearidade entre as VIs - [VIF < 10])

# Respondentes com apenas dívidas da empresa

vif(modelo_PJ)
vif(modelo4_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

vif(modelo_PJ_PF)

##5. (relação linear entre cada VI contínua e o logito da VD)
##(teste de Box-Tidwell: interação entre a VI contínua e seu log - [sig. >
0.05])

#--> não é o caso neste modelo

### Análise do modelo

##(efeitos gerais)

# Respondentes com apenas dívidas da empresa

Anova(modelo_PJ, type = "II", test = "Wald")
Anova(modelo1_PJ, type = "II", test = "Wald")
Anova(modelo2_PJ, type = "II", test = "Wald")
Anova(modelo3_PJ, type = "II", test = "Wald")
Anova(modelo4_PJ, type = "II", test = "Wald")

# Respondentes com dívidas da empresa e dívidas pessoais

Anova(modelo_PJ_PF, type = "II", test = "Wald")

##(efeitos específicos)

# Respondentes com apenas dívidas da empresa

summary(modelo_PJ)
summary(modelo1_PJ)
summary(modelo2_PJ)
summary(modelo3_PJ)
summary(modelo4_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

summary(modelo_PJ_PF)

### Razão de chance com IC 95% (usando erro padrão)

# Respondentes com apenas dívidas da empresa

exp(cbind(Odds_Ratio = coef(modelo2_PJ), confint.default(modelo2_PJ)))
exp(cbind(Odds_Ratio = coef(modelo3_PJ), confint.default(modelo3_PJ)))

# Respondentes com dívidas da empresa e dívidas pessoais

```

```
#--> não é o caso neste modelo

### Avaliação da qualidade e comparação entre modelos

##(pseudo R-quadrado - quanto de variação da variável dependente está sendo
explicada pelo modelo)

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(comparação de modelos - AIC e BIC - diferença de pelo menos 10 [quanto
menor, melhor])

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(qui-quadrado - modelos são diferentes com [sig. < 0,05])

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(tabela de classificação)

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo
```


APÊNDICE D – CÓDIGO-FONTE R (MODELO 4)

```

### Regressão Logística Binária ###
#####

### Modelo 4:

###  $Y(\text{Score Cred}) = A + B1(\text{Big Data}) + B2(\text{LGPD}) + E$ 

### Carregamento dos pacotes utilizados
#####

if(!require(pacman)) install.packages("pacman")
library(pacman)
pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

options(scipen = 999)

### Carregamento da base de dados
#####

# Respondentes com apenas dívidas da empresa

dados_PJ <- read.csv2('./Dados/PJ/Mod4/mod4_PJ.csv', stringsAsFactors =
TRUE)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF <- read.csv2('./Dados/PJ_PF/Mod4/mod4_PJ_PF.csv',
stringsAsFactors = TRUE)

### Análise exploratória
#####

# Respondentes com apenas dívidas da empresa

View(dados_PJ)
table(dados_PJ$usa_score)
summary(dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

View(dados_PJ_PF)
table(dados_PJ_PF$usa_score)
summary(dados_PJ_PF)

### Categorias de referência
#####

# Respondentes com apenas dívidas da empresa

levels(dados_PJ$usa_score)
levels(dados_PJ$usa_Big_Data)
levels(dados_PJ$usa_LGPD)

# Respondentes com dívidas da empresa e dívidas pessoais

```

```

levels(dados_PJ_PF$usa_score)
levels(dados_PJ_PF$usa_Big_Data)
levels(dados_PJ_PF$usa_LGPD)

### Construção do modelo com todas as preditoras
#####

# Respondentes com apenas dívidas da empresa

modelo_PJ <- glm(usa_score ~ .,
                family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo_PJ_PF <- glm(usa_score ~ .,
                  family = binomial(link = 'logit'), data = dados_PJ_PF)

### Construção dos modelos com as preditoras (sig.) avaliadas pelo qui-
quadrado
#####
###

# Respondentes com apenas dívidas da empresa

modelo1_PJ <- glm(usa_score ~ usa_Big_Data,
                family = binomial(link = 'logit'), data = dados_PJ)

modelo2_PJ <- glm(usa_score ~ usa_LGPD,
                family = binomial(link = 'logit'), data = dados_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo1_PJ_PF <- glm(usa_score ~ usa_Big_Data,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

modelo2_PJ_PF <- glm(usa_score ~ usa_LGPD,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Pressupostos para a Regressão Logística:
#####

##1. (variável dependente dicotômica - mutuamente exclusiva)
##2. (independência das observações - sem medidas repetidas)

##3. (ausência de outliers - limites dentro de [-3, 3])

# Respondentes com apenas dívidas da empresa

summary(stdres(modelo_PJ))
summary(stdres(modelo1_PJ))
summary(stdres(modelo2_PJ))

# Respondentes com dívidas da empresa e dívidas pessoais

summary(stdres(modelo_PJ_PF))
summary(stdres(modelo1_PJ_PF))
summary(stdres(modelo2_PJ_PF))

##4. (ausência de multicolinearidade entre as VIs - [VIF < 10])

```

```

# Respondentes com apenas dívidas da empresa

vif(modelo_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

vif(modelo_PJ_PF)

##5. (relação linear entre cada VI contínua e o logito da VD)
##(teste de Box-Tidwell: interação entre a VI contínua e seu log - [sig. >
0.05])

#--> não é o caso neste modelo

### Análise do modelo

##(efeitos gerais)

# Respondentes com apenas dívidas da empresa

Anova(modelo_PJ, type = "II", test = "Wald")
Anova(modelo1_PJ, type = "II", test = "Wald")
Anova(modelo2_PJ, type = "II", test = "Wald")

# Respondentes com dívidas da empresa e dívidas pessoais

Anova(modelo_PJ_PF, type = "II", test = "Wald")
Anova(modelo1_PJ_PF, type = "II", test = "Wald")
Anova(modelo2_PJ_PF, type = "II", test = "Wald")

##(efeitos específicos)

# Respondentes com apenas dívidas da empresa

summary(modelo_PJ)
summary(modelo1_PJ)
summary(modelo2_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

summary(modelo_PJ_PF)
summary(modelo1_PJ_PF)
summary(modelo2_PJ_PF)

### Razão de chance com IC 95% (usando erro padrão)

# Respondentes com apenas dívidas da empresa

exp(cbind(Odds_Ratio = coef(modelo1_PJ), confint.default(modelo1_PJ)))
exp(cbind(Odds_Ratio = coef(modelo2_PJ), confint.default(modelo2_PJ)))

# Respondentes com dívidas da empresa e dívidas pessoais

exp(cbind(Odds_Ratio = coef(modelo1_PJ_PF), confint.default(modelo1_PJ_PF)))
exp(cbind(Odds_Ratio = coef(modelo2_PJ_PF), confint.default(modelo2_PJ_PF)))

### Avaliação da qualidade e comparação entre modelos

##(pseudo R-quadrado - quanto de variação da variável dependente está sendo
explicada pelo modelo)

```

```
# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

##(comparação de modelos - AIC e BIC - diferença de pelo menos 10 [quanto
menor, melhor])

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

##(qui-quadrado - modelos são diferentes com [sig. < 0,05])

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo

##(tabela de classificação)

# Respondentes com apenas dívidas da empresa
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais
#--> não é o caso neste modelo
```

APÊNDICE E – CÓDIGO-FONTE R (MODELO 5)

```

### Regressão Logística Binária ###
#####

### Modelo 5:

###  $Y(\text{Endividamento}) = A + B1(\text{Fac de Cred}) + B2(\text{Parcel de Cont}) +$ 
###  $B3(\text{Aum Sal}) + B4(\text{Qnt Cart Cred}) + B5(\text{Qnt Conta Banc}) +$ 
###  $B6(\text{Qnt Dívida Ativa}) + B7(\text{Nome Limpo}) + E$ 

### Carregamento dos pacotes utilizados
#####

if(!require(pacman)) install.packages("pacman")
library(pacman)
pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

options(scipen = 999)

### Carregamento da base de dados
#####

# Respondentes com apenas dívidas da empresa

dados_PF <- read.csv2('./Dados/PF/Mod5/mod5_PF.csv', stringsAsFactors =
TRUE)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF <- read.csv2('./Dados/PJ_PF/Mod5/mod5_PJ_PF.csv',
stringsAsFactors = TRUE)

### Análise exploratória
#####

# Respondentes com apenas dívidas da empresa

View(dados_PF)
table(dados_PF$ouviu_score)
summary(dados_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

View(dados_PJ_PF)
table(dados_PJ_PF$ouviu_score)
summary(dados_PJ_PF)

### Categorias de referência
#####

# Respondentes com apenas dívidas da empresa

levels(dados_PF$ouviu_score)
levels(dados_PF$facilidade)
levels(dados_PF$parcelamento)
levels(dados_PF$aum_salarial)

```

```

levels(dados_PF$cart_credito)
levels(dados_PF$cont_bancária)
levels(dados_PF$dívidas)
levels(dados_PF$nome_limpo)

# Respondentes com dívidas da empresa e dívidas pessoais

levels(dados_PJ_PF$ouviu_score)
levels(dados_PJ_PF$facilidade)
levels(dados_PJ_PF$parcelamento)
levels(dados_PJ_PF$aum_salarial)
levels(dados_PJ_PF$cart_credito)
levels(dados_PJ_PF$cont_bancária)
levels(dados_PJ_PF$dívidas)
levels(dados_PJ_PF$nome_limpo)

### Construção do modelo com todas as preditoras
#####

# Respondentes com apenas dívidas da empresa

modelo_PF <- glm(ouviu_score ~ .,
                 family = binomial(link = 'logit'), data = dados_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo_PJ_PF <- glm(ouviu_score ~ .,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Construção dos modelos com as preditoras (sig.) avaliadas pelo qui-
quadrado
#####
###

# Respondentes com apenas dívidas da empresa

modelol_PF <- glm(ouviu_score ~ facilidade,
                 family = binomial(link = 'logit'), data = dados_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

modelol_PJ_PF <- glm(ouviu_score ~ cart_credito,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Pressupostos para a Regressão Logística:
#####

##1. (variável dependente dicotômica - mutuamente exclusiva)
##2. (independência das observações - sem medidas repetidas)

##3. (ausência de outliers - limites dentro de [-3, 3])

# Respondentes com apenas dívidas da empresa

summary(stdres(modelo_PF))
summary(stdres(modelol_PF))

# Respondentes com dívidas da empresa e dívidas pessoais

summary(stdres(modelo_PJ_PF))

```

```

summary(stdres(modelo1_PJ_PF))

##4. (ausência de multicolinearidade entre as VIs - [VIF < 10])

# Respondentes com apenas dívidas da empresa

vif(modelo_PJ)

# Respondentes com dívidas da empresa e dívidas pessoais

vif(modelo_PJ_PF)

##5. (relação linear entre cada VI contínua e o logito da VD)
##(teste de Box-Tidwell: interação entre a VI contínua e seu log - [sig. >
0.05])

#--> não é o caso neste modelo

### Análise do modelo

##(efeitos gerais)

# Respondentes com apenas dívidas da empresa

Anova(modelo_PF, type = "II", test = "Wald")
Anova(modelo1_PF, type = "II", test = "Wald")

# Respondentes com dívidas da empresa e dívidas pessoais

Anova(modelo_PJ_PF, type = "II", test = "Wald")
Anova(modelo1_PJ_PF, type = "II", test = "Wald")

##(efeitos específicos)

# Respondentes com apenas dívidas da empresa

summary(modelo_PF)
summary(modelo1_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

summary(modelo_PJ_PF)
summary(modelo1_PJ_PF)

### Razão de chance com IC 95% (usando erro padrão)

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

### Avaliação da qualidade e comparação entre modelos

##(pseudo R-quadrado - quanto de variação da variável dependente está sendo
explicada pelo modelo)

# Respondentes com apenas dívidas da empresa

```

```
#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(comparação de modelos - AIC e BIC - diferença de pelo menos 10 [quanto
menor, melhor])

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(qui-quadrado - modelos são diferentes com [sig. < 0,05])

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(tabela de classificação)

# Respondentes com apenas dívidas da empresa

#--> não é o caso neste modelo

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo
```


APÊNDICE F – CÓDIGO-FONTE R (MODELO 6)

```

### Regressão Logística Binária ###
#####

### Modelo 6:

###  $Y(\text{Tomd de Cred}) = A + B1(\text{Conhec Score Cred}) +$ 
###  $B2(\text{Entend Calc Score Cred}) + B3(\text{Conf Empresa}) +$ 
###  $B4(\text{Conf Seg Dados}) + B5(\text{Fac de Cred}) + B6(\text{Valores Parcel}) +$ 
###  $B7(\text{Consulta Serasa/SPC}) + E$ 

### Carregamento dos pacotes utilizados
#####

if(!require(pacman)) install.packages("pacman")
library(pacman)
pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

options(scipen = 999)

### Carregamento da base de dados
#####

# Respondentes com apenas dívidas da empresa

dados_PF <- read.csv2('./Dados/PF/Mod6/mod6_PF.csv', stringsAsFactors =
TRUE)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF <- read.csv2('./Dados/PJ_PF/Mod6/mod6_PJ_PF.csv',
stringsAsFactors = TRUE)

### Análise exploratória
#####

# Respondentes com apenas dívidas da empresa

View(dados_PF)
table(dados_PF$ouviu_score)
summary(dados_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

View(dados_PJ_PF)
table(dados_PJ_PF$ouviu_score)
summary(dados_PJ_PF)

### Categorias de referência
#####

# Respondentes com apenas dívidas da empresa

dados_PF$ouviu_score <- relevel(dados_PF$ouviu_score, ref = "sim")

levels(dados_PF$conhec_score)

```

```

levels(dados_PF$entend_score)
levels(dados_PF$confiança_emp)
levels(dados_PF$confiança_seg_dados)
levels(dados_PF$facilidade)
levels(dados_PF$parcelas)
levels(dados_PF$nome_limpo)

# Respondentes com dívidas da empresa e dívidas pessoais

dados_PJ_PF$ouviu_score <- relevel(dados_PJ_PF$ouviu_score, ref = "sim")

levels(dados_PJ_PF$ouviu_score)
levels(dados_PJ_PF$conhec_score)
levels(dados_PJ_PF$entendimento_score)
levels(dados_PJ_PF$confiança_emp)
levels(dados_PJ_PF$confiança_seg_dados)
levels(dados_PJ_PF$facilidade)
levels(dados_PJ_PF$parcelas)
levels(dados_PJ_PF$nome_limpo)

### Construção do modelo com todas as preditoras
#####

# Respondentes com apenas dívidas da empresa

modelo_PF <- glm(ouviu_score ~ .,
                 family = binomial(link = 'logit'), data = dados_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

modelo_PJ_PF <- glm(ouviu_score ~ .,
                   family = binomial(link = 'logit'), data = dados_PJ_PF)

### Construção dos modelos com as preditoras (sig.) avaliadas pelo qui-
quadrado
#####
###

# Respondentes com apenas dívidas da empresa

modelo1_PF <- glm(ouviu_score ~ conhec_score,
                 family = binomial(link = 'logit'), data = dados_PF)

modelo2_PF <- glm(ouviu_score ~ entend_score,
                 family = binomial(link = 'logit'), data = dados_PF)

modelo3_PF <- glm(ouviu_score ~ confiança_emp,
                 family = binomial(link = 'logit'), data = dados_PF)

modelo4_PF <- glm(ouviu_score ~ conhec_score + entend_score,
                 family = binomial(link = 'logit'), data = dados_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

### Pressupostos para a Regressão Logística:
#####

#1. (variável dependente dicotômica - mutuamente exclusiva)

```

```

##2. (independência das observações - sem medidas repetidas)

##3. (ausência de outliers - limites dentro de [-3, 3])

# Respondentes com apenas dívidas da empresa

summary(stdres(modelo_PF))
summary(stdres(modelo1_PF))
summary(stdres(modelo2_PF))
summary(stdres(modelo3_PF))
summary(stdres(modelo4_PF))

# Respondentes com dívidas da empresa e dívidas pessoais

summary(stdres(modelo_PJ_PF))

##4. (ausência de multicolinearidade entre as VIs - [VIF < 10])

# Respondentes com apenas dívidas da empresa

vif(modelo_PF)
vif(modelo4_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

vif(modelo_PJ_PF)

##5. (relação linear entre cada VI contínua e o logito da VD)
##(teste de Box-Tidwell: interação entre a VI contínua e seu log - [sig. >
0.05])

#--> não é o caso neste modelo

### Análise do modelo

##(efeitos gerais)

# Respondentes com apenas dívidas da empresa

Anova(modelo_PF, type = "II", test = "Wald")
Anova(modelo1_PF, type = "II", test = "Wald")
Anova(modelo2_PF, type = "II", test = "Wald")
Anova(modelo3_PF, type = "II", test = "Wald")
Anova(modelo4_PF, type = "II", test = "Wald")

# Respondentes com dívidas da empresa e dívidas pessoais

Anova(modelo_PJ_PF, type = "II", test = "Wald")

##(efeitos específicos)

# Respondentes com apenas dívidas da empresa

summary(modelo_PF)
summary(modelo1_PF)
summary(modelo2_PF)
summary(modelo3_PF)
summary(modelo4_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

```

```

summary(modelo_PJ_PF)

### Razão de chance com IC 95% (usando erro padrão)

# Respondentes com apenas dívidas da empresa

exp(cbind(Odds_Ratio = coef(modelo1_PF), confint.default(modelo1_PF)))
exp(cbind(Odds_Ratio = coef(modelo2_PF), confint.default(modelo2_PF)))
exp(cbind(Odds_Ratio = coef(modelo4_PF), confint.default(modelo4_PF)))

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

### Avaliação da qualidade e comparação entre modelos

##(pseudo R-quadrado - quanto de variação da variável dependente está sendo
explicada pelo modelo)

# Respondentes com apenas dívidas da empresa

PseudoR2(modelo_PF, which = "Nagelkerke")
PseudoR2(modelo4_PF, which = "Nagelkerke")

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(comparação de modelos - AIC e BIC - diferença de pelo menos 10 [quanto
menor, melhor])

# Respondentes com apenas dívidas da empresa

AIC(modelo_PF, modelo4_PF)
BIC(modelo_PF, modelo4_PF)

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(qui-quadrado - modelos são diferentes com [sig. < 0,05])

# Respondentes com apenas dívidas da empresa

anova(modelo4_PF, modelo_PF, test = "Chisq")

# Respondentes com dívidas da empresa e dívidas pessoais

#--> não é o caso neste modelo

##(tabela de classificação)

# Respondentes com apenas dívidas da empresa

ClassLog(modelo_PF, dados_PF$ouviu_score)
ClassLog(modelo4_PF, dados_PF$ouviu_score)

# Respondentes com dívidas da empresa e dívidas pessoais

```

#--> não é o caso neste modelo

APÊNDICE F – CÓDIGO-FONTE R (QUI-QUADRADO)

```
#### Qui-quadrado de independência ###
#####

#### Passo 1: Carregar os pacotes que serão usados

if(!require(dplyr)) install.packages("dplyr")
library(dplyr)
if(!require(rstatix)) install.packages("rstatix")
library(rstatix)
if(!require(psych)) install.packages("psych")
library(psych)

options(scipen = 999)

#### Passo 2: Carregamento do banco de dados
#(banco já no formato de tabela de contingência)
#####

# Referente aos respondentes com apenas dívidas da empresa

#Modelo 1

dados1_PJ_Mod1 <- read.csv2('./Dados/PJ/Mod1/mod1_tab-
contig_agilidade.csv', row.names = 1)
dados2_PJ_Mod1 <- read.csv2('./Dados/PJ/Mod1/mod1_tab-
contig_assertividade.csv', row.names = 1)
dados3_PJ_Mod1 <- read.csv2('./Dados/PJ/Mod1/mod1_tab-
contig_seguranca.csv', row.names = 1)
dados4_PJ_Mod1 <- read.csv2('./Dados/PJ/Mod1/mod1_tab-
contig_credibilidade.csv', row.names = 1)
dados5_PJ_Mod1 <- read.csv2('./Dados/PJ/Mod1/mod1_tab-
contig_facilidade.csv', row.names = 1)
dados6_PJ_Mod1 <- read.csv2('./Dados/PJ/Mod1/mod1_tab-contig_preco.csv',
row.names = 1)

View(dados1_PJ_Mod1)
glimpse(dados1_PJ_Mod1)

View(dados2_PJ_Mod1)
glimpse(dados2_PJ_Mod1)

View(dados3_PJ_Mod1)
glimpse(dados3_PJ_Mod1)

View(dados4_PJ_Mod1)
glimpse(dados4_PJ_Mod1)

View(dados5_PJ_Mod1)
glimpse(dados5_PJ_Mod1)

View(dados6_PJ_Mod1)
glimpse(dados6_PJ_Mod1)

#Modelo 2
```

```
dados1_PJ_Mod2 <- read.csv2('./Dados/PJ/Mod2/mod2_tab-
contig_treinamento.csv', row.names = 1)
dados2_PJ_Mod2 <- read.csv2('./Dados/PJ/Mod2/mod2_tab-
contig_red_inadimplencia.csv', row.names = 1)
dados3_PJ_Mod2 <- read.csv2('./Dados/PJ/Mod2/mod2_tab-
contig_aum_vend_parcelada.csv', row.names = 1)

View(dados1_PJ_Mod2)
glimpse(dados1_PJ_Mod2)

View(dados2_PJ_Mod2)
glimpse(dados2_PJ_Mod2)

View(dados3_PJ_Mod2)
glimpse(dados3_PJ_Mod2)

#Modelo 3

dados1_PJ_Mod3 <- read.csv2('./Dados/PJ/Mod3/mod3_tab-
contig_seguranca.csv', row.names = 1)
dados2_PJ_Mod3 <- read.csv2('./Dados/PJ/Mod3/mod3_tab-
contig_treinamento.csv', row.names = 1)
dados3_PJ_Mod3 <- read.csv2('./Dados/PJ/Mod3/mod3_tab-
contig_investimentos.csv', row.names = 1)
dados4_PJ_Mod3 <- read.csv2('./Dados/PJ/Mod3/mod3_tab-
contig_vazamentos.csv', row.names = 1)

View(dados1_PJ_Mod3)
glimpse(dados1_PJ_Mod3)

View(dados2_PJ_Mod3)
glimpse(dados2_PJ_Mod3)

View(dados3_PJ_Mod3)
glimpse(dados3_PJ_Mod3)

View(dados4_PJ_Mod3)
glimpse(dados4_PJ_Mod3)

#Modelo 4

dados1_PJ_Mod4 <- read.csv2('./Dados/PJ/Mod4/mod4_tab-contig_Big_Data.csv',
row.names = 1)
dados2_PJ_Mod4 <- read.csv2('./Dados/PJ/Mod4/mod4_tab-contig_LGPD.csv',
row.names = 1)

View(dados1_PJ_Mod4)
glimpse(dados1_PJ_Mod4)

View(dados2_PJ_Mod4)
glimpse(dados2_PJ_Mod4)

# Referente aos respondentes com apenas dívidas pessoais

#Modelo 5

dados1_PJ_Mod5 <- read.csv2('./Dados/PJ/Mod5/mod5_tab-
contig_facilidade.csv', row.names = 1)
```

```
dados2_PF_Mod5 <- read.csv2('./Dados/PF/Mod5/mod5_tab-
contig_parcelamento.csv', row.names = 1)
dados3_PF_Mod5 <- read.csv2('./Dados/PF/Mod5/mod5_tab-
contig_aum_salarial.csv', row.names = 1)
dados4_PF_Mod5 <- read.csv2('./Dados/PF/Mod5/mod5_tab-
contig_cart_credito.csv', row.names = 1)
dados5_PF_Mod5 <- read.csv2('./Dados/PF/Mod5/mod5_tab-
contig_cont_bancaria.csv', row.names = 1)
dados6_PF_Mod5 <- read.csv2('./Dados/PF/Mod5/mod5_tab-contig_dividas.csv',
row.names = 1)
dados7_PF_Mod5 <- read.csv2('./Dados/PF/Mod5/mod5_tab-
contig_nome_limpo.csv', row.names = 1)

View(dados1_PF_Mod5)
glimpse(dados1_PF_Mod5)

View(dados2_PF_Mod5)
glimpse(dados2_PF_Mod5)

View(dados3_PF_Mod5)
glimpse(dados3_PF_Mod5)

View(dados4_PF_Mod5)
glimpse(dados4_PF_Mod5)

View(dados5_PF_Mod5)
glimpse(dados5_PF_Mod5)

View(dados6_PF_Mod5)
glimpse(dados6_PF_Mod5)

View(dados7_PF_Mod5)
glimpse(dados7_PF_Mod5)

#Modelo 6

dados1_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-
contig_conhec_score.csv', row.names = 1)
dados2_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-
contig_entend_score.csv', row.names = 1)
dados3_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-
contig_confianca_emp.csv', row.names = 1)
dados4_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-
contig_confianca_seg_dados.csv', row.names = 1)
dados5_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-
contig_facilidade.csv', row.names = 1)
dados6_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-contig_parcelas.csv',
row.names = 1)
dados7_PF_Mod6 <- read.csv2('./Dados/PF/Mod6/mod6_tab-
contig_nome_limpo.csv', row.names = 1)

View(dados1_PF_Mod6)
glimpse(dados1_PF_Mod6)

View(dados2_PF_Mod6)
glimpse(dados2_PF_Mod6)

View(dados3_PF_Mod6)
glimpse(dados3_PF_Mod6)
```



```
View(dados4_PF_Mod6)
glimpse(dados4_PF_Mod6)

View(dados5_PF_Mod6)
glimpse(dados5_PF_Mod6)

View(dados6_PF_Mod6)
glimpse(dados6_PF_Mod6)

View(dados7_PF_Mod6)
glimpse(dados7_PF_Mod6)

# Referente aos respondentes com dívidas da empresa e dívidas pessoais

#Modelo 1

dados1_PJ_PF_Mod1 <- read.csv2('./Dados/PJ_PF/Mod1/mod1_tab-
contig_agilidade.csv', row.names = 1)
dados2_PJ_PF_Mod1 <- read.csv2('./Dados/PJ_PF/Mod1/mod1_tab-
contig_assertividade.csv', row.names = 1)
dados3_PJ_PF_Mod1 <- read.csv2('./Dados/PJ_PF/Mod1/mod1_tab-
contig_seguranca.csv', row.names = 1)
dados4_PJ_PF_Mod1 <- read.csv2('./Dados/PJ_PF/Mod1/mod1_tab-
contig_credibilidade.csv', row.names = 1)
dados5_PJ_PF_Mod1 <- read.csv2('./Dados/PJ_PF/Mod1/mod1_tab-
contig_facilidade.csv', row.names = 1)
dados6_PJ_PF_Mod1 <- read.csv2('./Dados/PJ_PF/Mod1/mod1_tab-
contig_preco.csv', row.names = 1)

View(dados1_PJ_PF_Mod1)
glimpse(dados1_PJ_PF_Mod1)

View(dados2_PJ_PF_Mod1)
glimpse(dados2_PJ_PF_Mod1)

View(dados3_PJ_PF_Mod1)
glimpse(dados3_PJ_PF_Mod1)

View(dados4_PJ_PF_Mod1)
glimpse(dados4_PJ_PF_Mod1)

View(dados5_PJ_PF_Mod1)
glimpse(dados5_PJ_PF_Mod1)

View(dados6_PJ_PF_Mod1)
glimpse(dados6_PJ_PF_Mod1)

#Modelo 2

dados1_PJ_PF_Mod2 <- read.csv2('./Dados/PJ_PF/Mod2/mod2_tab-
contig_treinamento.csv', row.names = 1)
dados2_PJ_PF_Mod2 <- read.csv2('./Dados/PJ_PF/Mod2/mod2_tab-
contig_red_inadimplencia.csv', row.names = 1)
dados3_PJ_PF_Mod2 <- read.csv2('./Dados/PJ_PF/Mod2/mod2_tab-
contig_aum_vend_parcelada.csv', row.names = 1)

View(dados1_PJ_PF_Mod2)
glimpse(dados1_PJ_PF_Mod2)

View(dados2_PJ_PF_Mod2)
```

```
glimpse(dados2_PJ_PF_Mod2)

View(dados3_PJ_PF_Mod2)
glimpse(dados3_PJ_PF_Mod2)

#Modelo 3

dados1_PJ_PF_Mod3 <- read.csv2('./Dados/PJ_PF/Mod3/mod3_tab-
contig_seguranca.csv', row.names = 1)
dados2_PJ_PF_Mod3 <- read.csv2('./Dados/PJ_PF/Mod3/mod3_tab-
contig_treinamento.csv', row.names = 1)
dados3_PJ_PF_Mod3 <- read.csv2('./Dados/PJ_PF/Mod3/mod3_tab-
contig_investimentos.csv', row.names = 1)
dados4_PJ_PF_Mod3 <- read.csv2('./Dados/PJ_PF/Mod3/mod3_tab-
contig_vazamentos.csv', row.names = 1)

View(dados1_PJ_PF_Mod3)
glimpse(dados1_PJ_PF_Mod3)

View(dados2_PJ_PF_Mod3)
glimpse(dados2_PJ_PF_Mod3)

View(dados3_PJ_PF_Mod3)
glimpse(dados3_PJ_PF_Mod3)

View(dados4_PJ_PF_Mod3)
glimpse(dados4_PJ_PF_Mod3)

#Modelo 4

dados1_PJ_PF_Mod4 <- read.csv2('./Dados/PJ_PF/Mod4/mod4_tab-
contig_Big_Data.csv', row.names = 1)
dados2_PJ_PF_Mod4 <- read.csv2('./Dados/PJ_PF/Mod4/mod4_tab-
contig_LGPD.csv', row.names = 1)

View(dados1_PJ_PF_Mod4)
glimpse(dados1_PJ_PF_Mod4)

View(dados2_PJ_PF_Mod4)
glimpse(dados2_PJ_PF_Mod4)

#Modelo 5

dados1_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_facilidade.csv', row.names = 1)
dados2_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_parcelamento.csv', row.names = 1)
dados3_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_aum_salarial.csv', row.names = 1)
dados4_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_cart_credito.csv', row.names = 1)
dados5_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_cont_bancaria.csv', row.names = 1)
dados6_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_dividas.csv', row.names = 1)
dados7_PJ_PF_Mod5 <- read.csv2('./Dados/PJ_PF/Mod5/mod5_tab-
contig_nome_limpo.csv', row.names = 1)

View(dados1_PJ_PF_Mod5)
glimpse(dados1_PJ_PF_Mod5)
```

```

View(dados2_PJ_PF_Mod5)
glimpse(dados2_PJ_PF_Mod5)

View(dados3_PJ_PF_Mod5)
glimpse(dados3_PJ_PF_Mod5)

View(dados4_PJ_PF_Mod5)
glimpse(dados4_PJ_PF_Mod5)

View(dados5_PJ_PF_Mod5)
glimpse(dados5_PJ_PF_Mod5)

View(dados6_PJ_PF_Mod5)
glimpse(dados6_PJ_PF_Mod5)

View(dados7_PJ_PF_Mod5)
glimpse(dados7_PJ_PF_Mod5)

#Modelo 6

dados1_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_conhec_score.csv', row.names = 1)
dados2_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_entend_score.csv', row.names = 1)
dados3_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_confianca_emp.csv', row.names = 1)
dados4_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_confianca_seg_dados.csv', row.names = 1)
dados5_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_facilidade.csv', row.names = 1)
dados6_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_parcelas.csv', row.names = 1)
dados7_PJ_PF_Mod6 <- read.csv2('./Dados/PJ_PF/Mod6/mod6_tab-
contig_nome_limpo.csv', row.names = 1)

View(dados1_PJ_PF_Mod6)
glimpse(dados1_PJ_PF_Mod6)

View(dados2_PJ_PF_Mod6)
glimpse(dados2_PJ_PF_Mod6)

View(dados3_PJ_PF_Mod6)
glimpse(dados3_PJ_PF_Mod6)

View(dados4_PJ_PF_Mod6)
glimpse(dados4_PJ_PF_Mod6)

View(dados5_PJ_PF_Mod6)
glimpse(dados5_PJ_PF_Mod6)

View(dados6_PJ_PF_Mod6)
glimpse(dados6_PJ_PF_Mod6)

View(dados7_PJ_PF_Mod6)
glimpse(dados7_PJ_PF_Mod6)

#### Passo 3: Realização do teste de Qui-quadrado
#(em tabelas 2x2: "correct = TRUE" ativa a correção de Yates)
#####

```

```
# Referente aos respondentes com apenas dívidas da empresa

#Modelo 1

quiqual_PJ_Mod1 <- chisq.test(dados1_PJ_Mod1, correct = TRUE)
quiqua2_PJ_Mod1 <- chisq.test(dados2_PJ_Mod1, correct = TRUE)
quiqua3_PJ_Mod1 <- chisq.test(dados3_PJ_Mod1, correct = TRUE)
quiqua4_PJ_Mod1 <- chisq.test(dados4_PJ_Mod1, correct = TRUE)
quiqua5_PJ_Mod1 <- chisq.test(dados5_PJ_Mod1, correct = TRUE)
quiqua6_PJ_Mod1 <- chisq.test(dados6_PJ_Mod1, correct = TRUE)

quiqual_PJ_Mod1
quiqua2_PJ_Mod1
quiqua3_PJ_Mod1
quiqua4_PJ_Mod1
quiqua5_PJ_Mod1
quiqua6_PJ_Mod1

#Modelo 2

quiqual_PJ_Mod2 <- chisq.test(dados1_PJ_Mod2, correct = TRUE)
quiqua2_PJ_Mod2 <- chisq.test(dados2_PJ_Mod2, correct = TRUE)
quiqua3_PJ_Mod2 <- chisq.test(dados3_PJ_Mod2, correct = TRUE)

quiqual_PJ_Mod2
quiqua2_PJ_Mod2
quiqua3_PJ_Mod2

#Modelo 3

quiqual_PJ_Mod3 <- chisq.test(dados1_PJ_Mod3, correct = TRUE)
quiqua2_PJ_Mod3 <- chisq.test(dados2_PJ_Mod3, correct = TRUE)
quiqua3_PJ_Mod3 <- chisq.test(dados3_PJ_Mod3, correct = TRUE)
quiqua4_PJ_Mod3 <- chisq.test(dados4_PJ_Mod3, correct = TRUE)

quiqual_PJ_Mod3
quiqua2_PJ_Mod3
quiqua3_PJ_Mod3
quiqua4_PJ_Mod3

#Modelo 4

quiqual_PJ_Mod4 <- chisq.test(dados1_PJ_Mod4, correct = TRUE)
quiqua2_PJ_Mod4 <- chisq.test(dados2_PJ_Mod4, correct = TRUE)

quiqual_PJ_Mod4
quiqua2_PJ_Mod4

# Referente aos respondentes com apenas dívidas pessoais

#Modelo 5

quiqual_PF_Mod5 <- chisq.test(dados1_PF_Mod5, correct = TRUE)
quiqua2_PF_Mod5 <- chisq.test(dados2_PF_Mod5, correct = TRUE)
quiqua3_PF_Mod5 <- chisq.test(dados3_PF_Mod5, correct = TRUE)
quiqua4_PF_Mod5 <- chisq.test(dados4_PF_Mod5, correct = TRUE)
quiqua5_PF_Mod5 <- chisq.test(dados5_PF_Mod5, correct = TRUE)
quiqua6_PF_Mod5 <- chisq.test(dados6_PF_Mod5, correct = TRUE)
quiqua7_PF_Mod5 <- chisq.test(dados7_PF_Mod5, correct = TRUE)
```

```

quiqua1_PJ_Mod5
quiqua2_PJ_Mod5
quiqua3_PJ_Mod5
quiqua4_PJ_Mod5
quiqua5_PJ_Mod5
quiqua6_PJ_Mod5
quiqua7_PJ_Mod5

#Modelo 6

quiqua1_PJ_Mod6 <- chisq.test(dados1_PJ_Mod6, correct = TRUE)
quiqua2_PJ_Mod6 <- chisq.test(dados2_PJ_Mod6, correct = TRUE)
quiqua3_PJ_Mod6 <- chisq.test(dados3_PJ_Mod6, correct = TRUE)
quiqua4_PJ_Mod6 <- chisq.test(dados4_PJ_Mod6, correct = TRUE)
quiqua5_PJ_Mod6 <- chisq.test(dados5_PJ_Mod6, correct = TRUE)
quiqua6_PJ_Mod6 <- chisq.test(dados6_PJ_Mod6, correct = TRUE)
quiqua7_PJ_Mod6 <- chisq.test(dados7_PJ_Mod6, correct = TRUE)

quiqua1_PJ_Mod6
quiqua2_PJ_Mod6
quiqua3_PJ_Mod6
quiqua4_PJ_Mod6
quiqua5_PJ_Mod6
quiqua6_PJ_Mod6
quiqua7_PJ_Mod6

# Referente aos respondentes com dívidas da empresa e dívidas pessoais

#Modelo 1

quiqua1_PJ_Mod1 <- chisq.test(dados1_PJ_Mod1, correct = TRUE)
quiqua2_PJ_Mod1 <- chisq.test(dados2_PJ_Mod1, correct = TRUE)
quiqua3_PJ_Mod1 <- chisq.test(dados3_PJ_Mod1, correct = TRUE)
quiqua4_PJ_Mod1 <- chisq.test(dados4_PJ_Mod1, correct = TRUE)
quiqua5_PJ_Mod1 <- chisq.test(dados5_PJ_Mod1, correct = TRUE)
quiqua6_PJ_Mod1 <- chisq.test(dados6_PJ_Mod1, correct = TRUE)

quiqua1_PJ_Mod1
quiqua2_PJ_Mod1
quiqua3_PJ_Mod1
quiqua4_PJ_Mod1
quiqua5_PJ_Mod1
quiqua6_PJ_Mod1

#Modelo 2

quiqua1_PJ_Mod2 <- chisq.test(dados1_PJ_Mod2, correct = TRUE)
quiqua2_PJ_Mod2 <- chisq.test(dados2_PJ_Mod2, correct = TRUE)
quiqua3_PJ_Mod2 <- chisq.test(dados3_PJ_Mod2, correct = TRUE)

quiqua1_PJ_Mod2
quiqua2_PJ_Mod2
quiqua3_PJ_Mod2

#Modelo 3

quiqua1_PJ_Mod3 <- chisq.test(dados1_PJ_Mod3, correct = TRUE)
quiqua2_PJ_Mod3 <- chisq.test(dados2_PJ_Mod3, correct = TRUE)
quiqua3_PJ_Mod3 <- chisq.test(dados3_PJ_Mod3, correct = TRUE)

```

```

quiqua4_PJ_PF_Mod3 <- chisq.test(dados4_PJ_PF_Mod3, correct = TRUE)

quiqual_PJ_PF_Mod3
quiqua2_PJ_PF_Mod3
quiqua3_PJ_PF_Mod3
quiqua4_PJ_PF_Mod3

#Modelo 4

quiqual_PJ_PF_Mod4 <- chisq.test(dados1_PJ_PF_Mod4, correct = TRUE)
quiqua2_PJ_PF_Mod4 <- chisq.test(dados2_PJ_PF_Mod4, correct = TRUE)

quiqual_PJ_PF_Mod4
quiqua2_PJ_PF_Mod4

#Modelo 5

quiqual_PJ_PF_Mod5 <- chisq.test(dados1_PJ_PF_Mod5, correct = TRUE)
quiqua2_PJ_PF_Mod5 <- chisq.test(dados2_PJ_PF_Mod5, correct = TRUE)
quiqua3_PJ_PF_Mod5 <- chisq.test(dados3_PJ_PF_Mod5, correct = TRUE)
quiqua4_PJ_PF_Mod5 <- chisq.test(dados4_PJ_PF_Mod5, correct = TRUE)
quiqua5_PJ_PF_Mod5 <- chisq.test(dados5_PJ_PF_Mod5, correct = TRUE)
quiqua6_PJ_PF_Mod5 <- chisq.test(dados6_PJ_PF_Mod5, correct = TRUE)
quiqua7_PJ_PF_Mod5 <- chisq.test(dados7_PJ_PF_Mod5, correct = TRUE)

quiqual_PJ_PF_Mod5
quiqua2_PJ_PF_Mod5
quiqua3_PJ_PF_Mod5
quiqua4_PJ_PF_Mod5
quiqua5_PJ_PF_Mod5
quiqua6_PJ_PF_Mod5
quiqua7_PJ_PF_Mod5

#Modelo 6

quiqual_PJ_PF_Mod6 <- chisq.test(dados1_PJ_PF_Mod6, correct = TRUE)
quiqua2_PJ_PF_Mod6 <- chisq.test(dados2_PJ_PF_Mod6, correct = TRUE)
quiqua3_PJ_PF_Mod6 <- chisq.test(dados3_PJ_PF_Mod6, correct = TRUE)
quiqua4_PJ_PF_Mod6 <- chisq.test(dados4_PJ_PF_Mod6, correct = TRUE)
quiqua5_PJ_PF_Mod6 <- chisq.test(dados5_PJ_PF_Mod6, correct = TRUE)
quiqua6_PJ_PF_Mod6 <- chisq.test(dados6_PJ_PF_Mod6, correct = TRUE)
quiqua7_PJ_PF_Mod6 <- chisq.test(dados7_PJ_PF_Mod6, correct = TRUE)

quiqual_PJ_PF_Mod6
quiqua2_PJ_PF_Mod6
quiqua3_PJ_PF_Mod6
quiqua4_PJ_PF_Mod6
quiqua5_PJ_PF_Mod6
quiqua6_PJ_PF_Mod6
quiqua7_PJ_PF_Mod6

#### Passo 4: Análise das frequências esperadas
#####

# Referente aos respondentes com apenas dívidas da empresa

#Modelo 1

quiqual_PJ_Mod1$expected
quiqua2_PJ_Mod1$expected

```

quiqua3_PJ_Mod1\$expected
quiqua4_PJ_Mod1\$expected
quiqua5_PJ_Mod1\$expected
quiqua6_PJ_Mod1\$expected

#Modelo 2

quiqua1_PJ_Mod2\$expected
quiqua2_PJ_Mod2\$expected
quiqua3_PJ_Mod2\$expected

#Modelo 3

quiqua1_PJ_Mod3\$expected
quiqua2_PJ_Mod3\$expected
quiqua3_PJ_Mod3\$expected
quiqua4_PJ_Mod3\$expected

#Modelo 4

quiqua1_PJ_Mod4\$expected
quiqua2_PJ_Mod4\$expected

Referente aos respondentes com apenas dívidas pessoais

#Modelo 5

quiqua1_PF_Mod5\$expected
quiqua2_PF_Mod5\$expected
quiqua3_PF_Mod5\$expected
quiqua4_PF_Mod5\$expected
quiqua5_PF_Mod5\$expected
quiqua6_PF_Mod5\$expected
quiqua7_PF_Mod5\$expected

#Modelo 6

quiqua1_PF_Mod6\$expected
quiqua2_PF_Mod6\$expected
quiqua3_PF_Mod6\$expected
quiqua4_PF_Mod6\$expected
quiqua5_PF_Mod6\$expected
quiqua6_PF_Mod6\$expected
quiqua7_PF_Mod6\$expected

Referente aos respondentes com dívidas da empresa e dívidas pessoais

#Modelo 1

quiqua1_PJ_PF_Mod1\$expected
quiqua2_PJ_PF_Mod1\$expected
quiqua3_PJ_PF_Mod1\$expected
quiqua4_PJ_PF_Mod1\$expected
quiqua5_PJ_PF_Mod1\$expected
quiqua6_PJ_PF_Mod1\$expected

#Modelo 2

quiqua1_PJ_PF_Mod2\$expected
quiqua2_PJ_PF_Mod2\$expected

```
quiqua3_PJ_PF_Mod2$expected
```

```
#Modelo 3
```

```
quiqua1_PJ_PF_Mod3$expected
```

```
quiqua2_PJ_PF_Mod3$expected
```

```
quiqua3_PJ_PF_Mod3$expected
```

```
quiqua4_PJ_PF_Mod3$expected
```

```
#Modelo 4
```

```
quiqua1_PJ_PF_Mod4$expected
```

```
quiqua2_PJ_PF_Mod4$expected
```

```
#Modelo 5
```

```
quiqua1_PJ_PF_Mod5$expected
```

```
quiqua2_PJ_PF_Mod5$expected
```

```
quiqua3_PJ_PF_Mod5$expected
```

```
quiqua4_PJ_PF_Mod5$expected
```

```
quiqua5_PJ_PF_Mod5$expected
```

```
quiqua6_PJ_PF_Mod5$expected
```

```
quiqua7_PJ_PF_Mod5$expected
```

```
#Modelo 6
```

```
quiqua1_PJ_PF_Mod6$expected
```

```
quiqua2_PJ_PF_Mod6$expected
```

```
quiqua3_PJ_PF_Mod6$expected
```

```
quiqua4_PJ_PF_Mod6$expected
```

```
quiqua5_PJ_PF_Mod6$expected
```

```
quiqua6_PJ_PF_Mod6$expected
```

```
quiqua7_PJ_PF_Mod6$expected
```

```
### Realização do teste exato de Fisher
```

```

#(pressuposto: [n(exp) < 5 numa casela] OU [n(exp) < 5 em mais de 20% das
caselas])

```

```
#####
#####
```

```
# Referente aos respondentes com apenas dívidas da empresa
```

```
#Modelo 1
```

```
fisher1_PJ_Mod1 <- fisher.test(dados1_PJ_Mod1)
```

```
fisher2_PJ_Mod1 <- fisher.test(dados2_PJ_Mod1)
```

```
fisher3_PJ_Mod1 <- fisher.test(dados3_PJ_Mod1)
```

```
fisher4_PJ_Mod1 <- fisher.test(dados4_PJ_Mod1)
```

```
fisher5_PJ_Mod1 <- fisher.test(dados5_PJ_Mod1)
```

```
fisher6_PJ_Mod1 <- fisher.test(dados6_PJ_Mod1)
```

```
fisher1_PJ_Mod1
```

```
fisher2_PJ_Mod1
```

```
fisher3_PJ_Mod1
```

```
fisher4_PJ_Mod1
```

```
fisher5_PJ_Mod1
```

```
fisher6_PJ_Mod1
```

```
#Modelo 2
```



```
fisher2_PJ_Mod2 <- fisher.test(dados2_PJ_Mod2)
fisher3_PJ_Mod2 <- fisher.test(dados3_PJ_Mod2)

fisher2_PJ_Mod2
fisher3_PJ_Mod2

#Modelo 3

fisher1_PJ_Mod3 <- fisher.test(dados1_PJ_Mod3)

fisher1_PJ_Mod3

# Referente aos respondentes com dívidas da empresa e dívidas pessoais

#Modelo 1

fisher1_PJ_PF_Mod1 <- fisher.test(dados1_PJ_PF_Mod1)
fisher2_PJ_PF_Mod1 <- fisher.test(dados2_PJ_PF_Mod1)
fisher3_PJ_PF_Mod1 <- fisher.test(dados3_PJ_PF_Mod1)
fisher4_PJ_PF_Mod1 <- fisher.test(dados4_PJ_PF_Mod1)
fisher5_PJ_PF_Mod1 <- fisher.test(dados5_PJ_PF_Mod1)
fisher6_PJ_PF_Mod1 <- fisher.test(dados6_PJ_PF_Mod1)

fisher1_PJ_PF_Mod1
fisher2_PJ_PF_Mod1
fisher3_PJ_PF_Mod1
fisher4_PJ_PF_Mod1
fisher5_PJ_PF_Mod1
fisher6_PJ_PF_Mod1

#Modelo 2

fisher1_PJ_PF_Mod2 <- fisher.test(dados1_PJ_PF_Mod2)
fisher2_PJ_PF_Mod2 <- fisher.test(dados2_PJ_PF_Mod2)
fisher3_PJ_PF_Mod2 <- fisher.test(dados3_PJ_PF_Mod2)

fisher1_PJ_PF_Mod2
fisher2_PJ_PF_Mod2
fisher3_PJ_PF_Mod2

#Modelo 3

fisher1_PJ_PF_Mod3 <- fisher.test(dados1_PJ_PF_Mod3)

fisher1_PJ_PF_Mod3

#Modelo 5

fisher1_PJ_PF_Mod5 <- fisher.test(dados1_PJ_PF_Mod5)
fisher2_PJ_PF_Mod5 <- fisher.test(dados2_PJ_PF_Mod5)
fisher3_PJ_PF_Mod5 <- fisher.test(dados3_PJ_PF_Mod5)
fisher4_PJ_PF_Mod5 <- fisher.test(dados4_PJ_PF_Mod5)
fisher5_PJ_PF_Mod5 <- fisher.test(dados5_PJ_PF_Mod5)
fisher6_PJ_PF_Mod5 <- fisher.test(dados6_PJ_PF_Mod5)
fisher7_PJ_PF_Mod5 <- fisher.test(dados7_PJ_PF_Mod5)

fisher1_PJ_PF_Mod5
fisher2_PJ_PF_Mod5
fisher3_PJ_PF_Mod5
fisher4_PJ_PF_Mod5
```

```

fisher5_PJ_PF_Mod5
fisher6_PJ_PF_Mod5
fisher7_PJ_PF_Mod5

#Modelo 6

fisher1_PJ_PF_Mod6 <- fisher.test(dados1_PJ_PF_Mod6)
fisher2_PJ_PF_Mod6 <- fisher.test(dados2_PJ_PF_Mod6)
fisher3_PJ_PF_Mod6 <- fisher.test(dados3_PJ_PF_Mod6)
fisher4_PJ_PF_Mod6 <- fisher.test(dados4_PJ_PF_Mod6)
fisher5_PJ_PF_Mod6 <- fisher.test(dados5_PJ_PF_Mod6)
fisher6_PJ_PF_Mod6 <- fisher.test(dados6_PJ_PF_Mod6)
fisher7_PJ_PF_Mod6 <- fisher.test(dados7_PJ_PF_Mod6)

fisher1_PJ_PF_Mod6
fisher2_PJ_PF_Mod6
fisher3_PJ_PF_Mod6
fisher4_PJ_PF_Mod6
fisher5_PJ_PF_Mod6
fisher6_PJ_PF_Mod6
fisher7_PJ_PF_Mod6

#### Passo 5: Análise dos resíduos padronizados ajustados (SPSS)
#(resíduo padronizado ajustado < -1,96 ou > 1,96 para ser significativo
0.05)
#####
##

# Referente aos respondentes com apenas dívidas da empresa

#Modelo 1

quiqua1_PJ_Mod1$stdres
quiqua2_PJ_Mod1$stdres
quiqua3_PJ_Mod1$stdres
quiqua4_PJ_Mod1$stdres
quiqua5_PJ_Mod1$stdres
quiqua6_PJ_Mod1$stdres

#Modelo 2

quiqua1_PJ_Mod2$stdres
quiqua2_PJ_Mod2$stdres
quiqua3_PJ_Mod2$stdres

#Modelo 3

quiqua1_PJ_Mod3$stdres
quiqua2_PJ_Mod3$stdres
quiqua3_PJ_Mod3$stdres
quiqua4_PJ_Mod3$stdres

#Modelo 4

quiqua1_PJ_Mod4$stdres
quiqua2_PJ_Mod4$stdres

# Referente aos respondentes com apenas dívidas pessoais

```

#Modelo 5

quiqua1_PF_Mod5\$stdres
quiqua2_PF_Mod5\$stdres
quiqua3_PF_Mod5\$stdres
quiqua4_PF_Mod5\$stdres
quiqua5_PF_Mod5\$stdres
quiqua6_PF_Mod5\$stdres
quiqua7_PF_Mod5\$stdres

#Modelo 6

quiqua1_PF_Mod6\$stdres
quiqua2_PF_Mod6\$stdres
quiqua3_PF_Mod6\$stdres
quiqua4_PF_Mod6\$stdres
quiqua5_PF_Mod6\$stdres
quiqua6_PF_Mod6\$stdres
quiqua7_PF_Mod6\$stdres

Referente aos respondentes com dívidas da empresa e dívidas pessoais

#Modelo 1

quiqua1_PJ_PF_Mod1\$stdres
quiqua2_PJ_PF_Mod1\$stdres
quiqua3_PJ_PF_Mod1\$stdres
quiqua4_PJ_PF_Mod1\$stdres
quiqua5_PJ_PF_Mod1\$stdres
quiqua6_PJ_PF_Mod1\$stdres

#Modelo 2

quiqua1_PJ_PF_Mod2\$stdres
quiqua2_PJ_PF_Mod2\$stdres
quiqua3_PJ_PF_Mod2\$stdres

#Modelo 3

quiqua1_PJ_PF_Mod3\$stdres
quiqua2_PJ_PF_Mod3\$stdres
quiqua3_PJ_PF_Mod3\$stdres
quiqua4_PJ_PF_Mod3\$stdres

#Modelo 4

quiqua1_PJ_PF_Mod4\$stdres
quiqua2_PJ_PF_Mod4\$stdres

#Modelo 5

quiqua1_PJ_PF_Mod5\$stdres
quiqua2_PJ_PF_Mod5\$stdres
quiqua3_PJ_PF_Mod5\$stdres
quiqua4_PJ_PF_Mod5\$stdres
quiqua5_PJ_PF_Mod5\$stdres
quiqua6_PJ_PF_Mod5\$stdres
quiqua7_PJ_PF_Mod5\$stdres

#Modelo 6

quiqua1_PJ_PF_Mod6\$stdres
quiqua2_PJ_PF_Mod6\$stdres
quiqua3_PJ_PF_Mod6\$stdres
quiqua4_PJ_PF_Mod6\$stdres
quiqua5_PJ_PF_Mod6\$stdres
quiqua6_PJ_PF_Mod6\$stdres
quiqua7_PJ_PF_Mod6\$stdres

ANEXO A – QUESTIONÁRIO SCORE DE CRÉDITO, *BIG DATA* E LGPD

Questionário Score de Crédito, Big data e LGPD

Olá! Me chamo Elisa, sou aluna de mestrado em Administração do PPGA da UNIGRANRIO.

Este questionário busca entender um pouco mais a respeito das questões de endividamento de indivíduos e empresas em relação aos fatores do score de crédito, da Lei Geral de Proteção de Dados e da Big Data.

Por questões de segurança e privacidade, nenhuma informação que possa identificar o usuário, de maneira direta ou indireta, será solicitada.

Por fim, para que o estudo reflita um resultado o mais próximo possível da realidade, peço para que as respostas sejam as mais exatas possíveis.

Desde já agradeço pela sua participação neste trabalho!

*Obrigatório



Informações Sócio-Demográficas

1. Qual o seu sexo? *

Esta pergunta visa estudar se existem diferenças significativas entre as opções abaixo, sabendo-se que a questão de gênero não é limitada em Feminino e Masculino.

Marcar apenas uma oval.

- Feminino
 Masculino

2. Qual a sua etnia? *

Esta pergunta visa estudar se existem diferenças significativas entre as opções abaixo, utilizando-se a classificação adotada pelo PNAD - 2019.

Marcar apenas uma oval.

- Branco
 Preto
 Pardo
 Amarelo
 Indígena

3. Qual a sua faixa etária? *

Marcar apenas uma oval.

- Entre 18 anos e 24 anos
- Entre 25 anos e 29 anos
- Entre 30 anos e 34 anos
- Entre 35 anos e 39 anos
- Entre 40 anos e 44 anos
- Entre 45 anos e 49 anos
- Entre 50 anos e 54 anos
- Entre 55 anos e 59 anos
- Entre 60 anos e 64 anos
- Igual ou maior que 65 anos

4. Qual a sua escolaridade atual? *

Marcar apenas uma oval.

- Ensino fundamental incompleto
- Ensino fundamental completo
- Ensino médio incompleto
- Ensino médio completo
- Ensino superior incompleto
- Ensino superior completo
- Pós-graduação (MBA, Mestrado, Doutorado) incompleto
- Pós-graduação (MBA, Mestrado, Doutorado) completo

5. Qual o seu estado civil? *

Marcar apenas uma oval.

- Solteiro(a)
- Casado(a)
- Divorciado(a)
- União Estável
- Viúvo(a)

6. Qual a sua renda mensal bruta? *

Marcar apenas uma oval.

- Igual ou menor que R\$ 2.200,00 (Igual ou menor que 2 salários mínimos)
- Entre R\$ 2.200,01 e R\$ 4.400,00 (Entre 2 e 4 salários mínimos)
- Entre R\$ 4.400,01 e R\$ 6.600,00 (Entre 4 e 6 salários mínimos)
- Entre R\$ 6.600,01 e R\$ 8.800,00 (Entre 6 e 8 salários mínimos)
- Entre R\$ 8.800,01 e R\$ 11.000,00 (Entre 8 e 10 salários mínimos)
- Entre R\$ 11.000,01 e R\$ 22.000,00 (Entre 10 e 20 salários mínimos)
- Maior que R\$ 22.000,00 (Maior que 20 salários mínimos)

Informações sobre Tomada de Crédito e Endividamento

7. Atualmente você possui dívidas (em dia ou não)? *

As dívidas podem ser pessoais e/ou da sua empresa

Marcar apenas uma oval.

- Sim
- Não *Pular para a seção 16 (Obrigada pela participação!)*

Informações sobre Tomada de Crédito e Endividamento

8. Suas dívidas são: *

Marcar apenas uma oval.

- Apenas dívidas pessoais *Pular para a pergunta 21*
- Apenas dívidas da empresa (caso você seja sócio/proprietário de alguma empresa)
- Dívidas pessoais e dívidas da empresa (caso você seja sócio/proprietário de alguma empresa)

Informações sobre Tomada de Crédito e Endividamento referentes a sócios/proprietários de empresas

9. Na sua empresa, você faz alguma consulta ao score de crédito? *

Marcar apenas uma oval.

- Sim
- Não *Pular para a pergunta 15*

Informações sobre Tomada de Crédito e Endividamento referentes a sócios/proprietários de empresas

10. Na sua percepção, ao utilizar o score de crédito, o que mais lhe parece ser interessante no uso? (você pode marcar mais de uma opção se desejar) *

Marque todas que se aplicam.

- Agilidade
- Assertividade
- Segurança proporcionada
- Credibilidade proporcionada
- Facilidade
- Preço do serviço

Informações sobre Tomada de Crédito e Endividamento referentes a sócios/proprietários de empresas

Em relação ao custo-benefício com a utilização do score de crédito, responda às seguintes questões:

11. Acha justo o valor pago pelo uso do serviço de score de crédito? *

Marcar apenas uma oval.

- Sim
- Não

12. Teve grandes gastos com treinamento de pessoal para usar o score de crédito? *

Marcar apenas uma oval.

- Sim
- Não

13. Desde que começou a utilizar o score de crédito, notou alguma redução na sua inadimplência? *

Marcar apenas uma oval.

- Sim
- Não

14. Desde que começou a utilizar o score de crédito, notou algum aumento nas vendas parceladas? *

*

Marcar apenas uma oval.

Sim

Não

Informações sobre Tomada de Crédito e Endividamento referentes a sócios/proprietários de empresas

15. Na sua percepção, o que mais lhe parece ser ruim no uso do score de crédito? (você pode marcar mais de uma opção se desejar)

Marque todas que se aplicam.

Agilidade

Assertividade

Segurança proporcionada

Credibilidade proporcionada

Facilidade

Preço do serviço

Informações sobre à LGPD

A LGPD (Lei Geral de Proteção de Dados) entrou em vigor recentemente e trouxe mudanças para a gestão das informações por parte das empresas. Com base nisso, responda à seguinte questão:

16. Você já começou a adequar a sua empresa em relação à LGPD? *

Marcar apenas uma oval.

Sim

Não *Pular para a pergunta 18*

Informações sobre à LGPD

A LGPD (Lei Geral de Proteção de Dados) entrou em vigor recentemente e trouxe mudanças para a gestão das informações por parte das empresas. Com base nisso, responda à seguinte questão:

17. No seu negócio, você acredita que o uso das boas práticas da LGPD pode influenciar em quê? (você pode marcar mais de uma opção se desejar) *

Marque todas que se aplicam.

- Uso de segurança para o cadastramento e armazenamento de dados
 Treinamento de pessoal sobre o uso das boas práticas da LGPD
 Investimentos em segurança de dados (programas seguros, banco de dados, nuvem, etc)
 Problemas com vazamento de dados

Outro: _____

Informações sobre à LGPD

A LGPD (Lei Geral de Proteção de Dados) entrou em vigor recentemente e trouxe mudanças para a gestão das informações por parte das empresas. Com base nisso, responda à seguinte questão:

18. No seu negócio, o que você acredita que possa inibir o uso das boas práticas da LGPD?(você pode marcar mais de uma opção se desejar)

Marque todas que se aplicam.

- Uso de segurança para o cadastramento e armazenamento de dados
 Treinamento de pessoal sobre o uso das boas práticas da LGPD
 Investimentos em segurança de dados (programas seguros, banco de dados, nuvem, etc)
 Problemas com vazamento de dados

Outro: _____

Informações sobre à LGPD

A LGPD (Lei Geral de Proteção de Dados) entrou em vigor recentemente e trouxe mudanças para a gestão das informações por parte das empresas. Com base nisso, responda à seguinte questão:

19. Já houve algum vazamento de dados na sua empresa? *

Marcar apenas uma oval.

Sim

Não

Informações sobre o uso da BIG DATA

A Big Data é a área do conhecimento que estuda como tratar, analisar e obter informações a partir de grandes volumes de dados, cuja as análises não são feitas por sistemas tradicionais. Com base nisso, responda à seguinte questão:

20. Sua empresa utiliza Big Data para questões envolvendo o score de crédito? *

Uso de dados de mídias sociais, moradia, renda familiar, se tem casa própria, etc

Marcar apenas uma oval.

Sim

Não

Informações sobre Tomada de Crédito e Endividamento referentes à análise do Endividamento Pessoal

21. Quantas contas bancárias você possui atualmente? *

Marcar apenas uma oval.

0	1	2	3	4	5	6	7	8	9	10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

22. Quantos cartões de crédito você possui atualmente? *

Marcar apenas uma oval.

0	1	2	3	4	5	6	7	8	9	10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

23. Você acha que ter o nome limpo é importante? *

Marcar apenas uma oval.

Sim

Não

24. Você já ouviu falar sobre o score de crédito? *

Marcar apenas uma oval.

Sim

Não

25. Você já buscou verificar como se calcula o score de crédito? *

Marcar apenas uma oval.

Sim

Não

26. Na sua opinião, você acha correto as empresas responsáveis pelo score de crédito não informarem, para a população em geral, quais informações são utilizadas e que importância (peso) possuem para calcular o score de crédito (as empresas alegam sigilo empresarial)? *

Por exemplo: pagar contas em dia teria importância média; pagar contas adiantado teria importância alta; etc

Marcar apenas uma oval.

Sim

Não

27. Na sua opinião, o que você acredita que pode gerar mais influência no aumento das dívidas? (você pode marcar mais de uma opção se desejar) *

Marque todas que se aplicam.

Facilidade de crédito

Possibilidade de poder parcelar contas

Aumento salarial

Quantidade de cartões de crédito

Quantidade de contas no banco

Já possuir dívidas

Ter nome limpo

Outro: _____

28. Na sua opinião, o que você acredita que mais possa influenciar na formação desse endividamento pessoal? (você pode marcar mais de uma opção se desejar) *

Marque todas que se aplicam.

Falta de conhecimentos sobre o score de crédito

Falta de entendimento sobre o cálculo do score de crédito

Confiança na empresa que vai parcelar a dívida/obter crédito

Confiança na segurança dos dados pessoais

Facilidade na obtenção de crédito

O valor das parcelas (possibilidade de pagamento do valor acordado)

Ter o nome limpo

Outro: _____

Informações sobre Tomada de Crédito e Endividamento referentes à análise do Endividamento Pessoal

29. Na sua opinião, o que você acredita que pode gerar mais influência no aumento das dívidas? (você pode marcar mais de uma opção se desejar) *

Marque todas que se aplicam.

- Facilidade de crédito
- Possibilidade de poder parcelar contas
- Aumento salarial
- Quantidade de cartões de crédito
- Quantidade de contas no banco
- Já possuir dívidas
- Ter nome limpo

Outro: _____

Obrigada pela participação!

A pesquisa terminou.

Gostaria de agradecer sua colaboração!

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários