

Modelo de Previsão de Falências nas Pequenas e Médias Empresas



Mestrado em Finanças e Fiscalidade

Dissertação para obtenção do grau de Mestre



Orientada por:

Professor Doutor António de Melo da Costa Cerqueira



Inês Castro Tavares | 2021

Faculdade de Economia do Porto



Nota Bibliográfica do Autor

Inês Castro Tavares nascida a 26 de abril de 1996, natural de Oliveira de Azeméis, Aveiro.

Mudou-se para o Porto em 2014 e licenciou-se em Gestão de Empresas pelo *Instituto Superior de Administração e Gestão* (ISAG) com média de 15 valores, em 2017. Concluiu no ano de 2018 a Pós-Graduação em Finanças e Fiscalidade na *Porto Business School*. Em 2019 iniciou o Mestrado em Finanças e Fiscalidade na faculdade de Economia no Porto.

Em janeiro de 2018 iniciou a sua atividade profissional no Grupo RAR em contas a pagar. Em outubro de 2020 abraçou um novo projeto profissional na área de *Business Service & Outsourcing*, na empresa KPMG.

Agradecimentos

Vários são os que estiveram no processo desta dissertação, e a todos, o meu muito obrigada.

Começo por agradecer aos meus pais, por todo o apoio ao longo de todos estes anos.

Ao meu orientador, Professor Doutor António Cerqueira, por toda a sua ajuda ao longo deste percurso.

Ao meu namorado, que com toda a sua paciência facilitou este processo.

Aos meus amigos e colegas de trabalho, que sempre me apoiaram e permitiram que conseguisse conciliar tudo.

Resumo

A previsão de falências é um tema que, até aos dias de hoje, tem vindo a ser largamente estudado por diversos autores, dada a necessidade de compreender o fenómeno da falência empresarial. Este fenómeno é caracterizado pelo risco envolvente e pela sua constante evolução e, por isso, tem originado diversos modelos até à data. Uma das grandes questões centra-se em descobrir quais as variáveis, amostra e métodos econométricos (desde os modelos estatísticos aos modelos desenvolvidos através de métodos computacionais) que, combinados, alcançam uma maior capacidade preditiva.

Assim, o principal objetivo deste trabalho é desenvolver um modelo de regressão logística que se revele eficaz na identificação de possíveis situações de insolvência, com base no modelo Z-score de Edward Altman, aplicado à realidade das PME's portuguesas que representam uma parte fundamental na estrutura económica do país. Neste sentido, a amostra utilizada foi de 80 empresas no período entre 2015 e 2019, divididas em dois grupos: 40 empresas falidas e 40 empresas ativas. Posteriormente, como segundo objetivo, foi introduzida uma variável adicional ao modelo principal com o propósito de melhorar a sua capacidade de previsão. Ambos os modelos foram estimados através das variáveis explicativas usadas por Altman, no modelo Z-Score.

De acordo com estudos já realizados, conclui-se que a regressão logística (*Ohlson, 1980*) apresenta uma maior eficácia face ao pretendido, tendo em conta a capacidade de classificar as observações da amostra, através da inferior percentagem de erro tipo I. Contudo, os rácios que mais contribuem para a explicação do modelo são os de *Altman (1968)*. Assim, através do modelo desenvolvido, a previsão de falências em PME's portuguesas, tem uma capacidade preditiva de 78,91%. A adição de uma sexta variável ao modelo permitiu aumentar a capacidade de previsão para 78,96%.

As principais conclusões foram que a regressão logística alcançou uma maior capacidade de previsão em comparação com a análise discriminante e que os modelos criados nas décadas de 1960 e 1980 ainda são relevantes para a realidade portuguesa. O desenvolvimento deste trabalho torna-se relevante uma vez que é desenvolvido com fatores de dois autores tomados como referência nesta temática.

Palavras-chave: Falência, Logit, Z-score, PME's

Abstract

Bankruptcy prediction is a topic that, to this day, has been widely studied by several authors, given the need to understand the phenomenon of corporate bankruptcy. This phenomenon is characterized by the risk involved and its constant evolution and, therefore, has originated several models to date. One of the big questions focuses on finding out which variables, which sample and which econometric methods (from statistical models to models developed through computational methods), combined, achieve a greater predictive capacity.

Thus, the main objective of this work is to develop a logistic regression model that proves to be effective in identifying possible insolvency situations, based on Edward Altman's Z-score model, applied to the reality of Portuguese SME's, which represent a fundamental part in the country's economic structure. In this sense, the sample used was 80 companies in the period between 2015 and 2019, divided into two groups: 40 bankrupt companies and 40 companies in the opposite situation. Subsequently, as a second objective, an additional variable was introduced to the main model with the purpose of improving its predictive ability. Both models were estimated using the explanatory variables used by Altman in the Z-Score model.

According to studies already conducted, it is concluded that logistic regression (Ohlson, 1980) is more effective than intended, given the ability to classify the observations of the sample, through the lower percentage of type I error. However, the ratios that contribute the most to the explanation of the model are those of Altman (1968). Thus, through the model developed, the prediction of bankruptcy in Portuguese SMEs, has a predictive ability of 78.91%. The addition of a sixth variable to the model allowed the predictive ability to be increased to 78.96%.

The main conclusions were that logistic regression achieved a higher predictive ability compared to discriminant analysis and that the models created in the 1960s and 1980s are still relevant to the Portuguese reality. The development of this work becomes relevant since it is developed with factors from two authors taken as reference in this theme.

Keywords: Bankruptcy, Logit, Z-score, SME's

Índice

Nota Bibliográfica do Autor.....	III
Agradecimentos.....	IV
Resumo.....	V
Abstract.....	VI
Índice.....	VII
Índice de Tabelas	VIII
Abreviaturas.....	IX
1. Introdução	1
2. Revisão da Bibliografia	4
2.1) Definição de Insolvência.....	4
2.2) Caracterização das PME's em Portugal	5
2.3) Modelos de previsão de falência e a sua evolução	7
3. Metodologia.....	13
3.1) Descrição dos dados	13
3.2) Amostra	13
3.3) Seleção das variáveis	14
3.3.1) Variáveis dependentes:.....	14
3.3.2) Variáveis independentes:.....	14
3.3.3) Introdução de uma nova variável (X6):.....	16
4. Hipóteses de Investigação	18
5. Estimação do modelo.....	19
5.1) Base de dados	19
5.1.1) Vantagens e limitações do modelo.....	19
5.2) Estimação	20
6. Resultados	22
7. Conclusão	30
8. Bibliografia	32
9. Anexos	34
Anexo 1) Estimação modelo Logit – cinco variáveis	34
Anexo 2) Estimação modelo Logit – seis variáveis	37

Índice de Tabelas

Tabela 1 – Principais indicadores das sociedades não financeiras, por dimensão (2018-2019)

Tabela 2 – Capacidade Preditiva do Modelo de Ohlson

Tabela 3 – Variáveis

Tabela 4 – Variáveis introduzidas no modelo de Altman e Sabato (2007)

Tabela 5 – Sinais esperados das variáveis utilizadas

Tabela 6 – Estatísticas Descritivas

Tabela 7 – Correlações de variáveis

Tabela 8 – Coeficientes da Regressão Logística

Tabela 9 – MANOVA

Tabela 10 – Capacidade Preditiva Erros tipo I e tipo II

Tabela 11 – Capacidade Preditiva – Final

Tabela 12 – Capacidade Preditiva - Análise comparativa

Abreviaturas

CIRE – Código da Insolvência e Recuperação de Empresas

INE – Instituto Nacional de Estatística

PME's – Pequenas e Médias Empresas

RNA – Redes Neutrais Artificiais

SABI – Sistema de Análise de Balanços Ibéricos

SPSS – Statistical Package for the Social Sciences

1. Introdução

O interesse pela temática da falência das empresas tem aumentado ao longo dos anos devido, em grande parte, à crescente globalização da economia o que leva a uma necessidade por parte dos agentes económicos de usar modelos que consigam analisar e antever possíveis situações de insolvência. O estudo alusivo à previsão de falências é relevante uma vez que, apesar da análise neste tema ter sido iniciada em 1930, de acordo com *Gissel (2007)*, é um tema atual dado que nos dias de hoje as empresas ainda têm de ter preocupação no que toca ao comportamento dos seus indicadores. Para as empresas e respetivos *stakeholders*, a falência sempre foi, e continua a ser um fator de extrema importância.

As PME's, para muitos países, são consideradas como a “espinha dorsal” da economia. Desta forma, a previsão de falências neste tipo de empresas tornou-se numa área de estudo com relevo nos últimos tempos devido ao elevado número de PME's existentes e à importância que as mesmas têm no mercado. De acordo com *Neves (2004)* “em Portugal, o peso das PME's é muito importante”. A previsão de falência neste tipo de empresas tem sido igualmente analisada por vários autores, tais como: *Everett e Watson (1998)*, *Altman e Sabato (2007)*, *Cultrera e Brédart (2016)*.

Desta forma, o objetivo primordial do presente trabalho passa pelo desenvolvimento de um modelo de regressão logística que permite uma análise mais objetiva, antecipando uma possível situação de insolvência tendo por base o modelo pioneiro de Edward Altman – Z-score. Posteriormente, como segundo objetivo, irá ser introduzido ao modelo original uma variável adicional, com o propósito de melhorar a capacidade preditiva do mesmo. Vai igualmente verificar-se se a literatura desta temática tem suporte empírico e uma boa capacidade preditiva na realidade das PME's portuguesas.

A literatura em volta deste tema é bastante vasta e, inúmeros foram e são os autores que se debruçam sobre esta temática e que foram melhorando e aprofundando os resultados obtidos ao longo dos anos. O principal contributo apareceu em meados dos anos 60 com o estudo da análise univariada de *William Beaver (1966)*. Este estudo era baseado na análise de cada variável isoladamente com o objetivo de distinguir uma empresa insolvente de uma não insolvente. No final da mesma década, em 1968, surgiu *Edward Altman* com o primeiro estudo, também inovador, a análise discriminante multivariada. Distinguiu-se do anterior devido ao facto de utilizar mais do que uma variável financeira para classificar as empresas como insolventes ou não insolventes. A partir deste estudo muitos outros trabalhos surgiram

de forma a tentar melhorar os resultados obtidos até então por Edward Altman dos quais saliento o trabalho de *Ohlson (1980)* uma vez que foi baseado num novo modelo econométrico, a regressão logística. Os três autores acima mencionados serviram de referência para a elaboração deste trabalho.

Recentemente, dado o avanço tecnológico têm surgido novos modelos para além dos referidos até ao momento como Probit (*Johnston, 2001*), Gombit (*McCulloch, 2007*), redes neurais artificiais, data *envelopment analysis*, entre outros.

A amostra utilizada será então constituída por 80 PME's portuguesas, divididas em 2 grupos – 40 empresas falidas e 40 empresas não falidas - entre 2015 e 2019. Devido aos bons resultados obtidos no modelo Z-Score, as variáveis explicativas deste trabalho serão as mesmas que Altman utilizou no seu modelo, de maneira a verificar se o resultado do modelo é superior quando usadas num modelo econométrico diferente. Numa fase mais adiantada, como já referido acima, irá ser introduzida uma nova variável, *EBIT/ Interest Expenses*. A escolha desta variável não tem sido muito utilizada pelos autores que se tem vindo a debruçar por esta temática e, por isso, pretende-se mostrar a influência que a mesma tem na capacidade preditiva do modelo. Todos os dados para a elaboração do estudo foram retirados da base de dados SABI.

Apesar do tema escolhido já ter sido alvo de vários estudos, este trabalho é inovador por estudar as variáveis analisadas por Altman, num tipo de modelo que foi desenvolvido por Ohlson (Logit), mas numa amostra e período diferentes. Os modelos de previsão de falência em PME's portuguesas desenvolvidos neste trabalho apresentaram uma capacidade preditiva de 78,91% e 78,96%. O primeiro resultado diz respeito ao modelo com as cinco variáveis, e o segundo resultado é referente ao modelo com a variável adicional.

Relativamente à estrutura deste trabalho, o mesmo encontra-se dividido em cinco capítulos. Depois da introdução, no capítulo 2 realizar-se-á uma breve revisão da literatura que fundamentou o presente estudo e que conduziu à escolha da metodologia usada. No capítulo 3 irá proceder-se à descrição da metodologia usada, nomeadamente a seleção e caracterização da amostra e determinação das variáveis a testar. Todo o estudo empírico será desenvolvido neste capítulo. No capítulo 4, serão apresentadas as hipóteses deste estudo. Posteriormente no capítulo 5 irá ser abordada a forma como foi feita a estimação do modelo e respetiva base de dados; as vantagens e limitações do tipo de modelo escolhido. Ao

aproximar-se o fim da dissertação, no capítulo 6 irão ser descritos os resultados obtidos, comparando-os com a literatura empírica (dando destaque ao modelo da análise discriminante multivariada e ao modelo de regressão logística). O capítulo 7 contemplará a conclusão do trabalho.

2. Revisão da Bibliografia

O interesse por esta temática não é de agora e muitos são aqueles que se dedicam à criação e desenvolvimento de modelos aplicáveis à previsão de falência. Desde a primeira abordagem - a análise univariada - às análises de Logit e Probit, muitos são os esforços para tentar prever a falência através da utilização de rácios.

Numa primeira fase irá ser definido o conceito de insolvência e será feita a caracterização das PME's em Portugal. Posteriormente, irei fazer referência aos estudos acerca da temática da previsão de falência que considero como os mais relevantes para o desenvolvimento do meu estudo.

2.1) Definição de Insolvência

Em Portugal, e como já foi referido acima, o conceito de insolvência encontra-se definido no CIRE, no artigo terceiro. De acordo com o mesmo uma empresa encontra-se numa situação de insolvência quando o valor do passivo excede o valor do ativo.

Para além da definição do conceito, é usual neste tipo de estudo, a definição de falência utilizada pois depende do momento em que é definida como é salientado por Watson e Everett (1993). Podemos falar de uma definição de cariz económico que é quando a empresa, como já mencionado, não consegue cumprir com as suas obrigações a nível financeiro ou de uma definição de cariz jurídico que é quando o processo entra em tribunal. Dependendo destes fatores, a capacidade preditiva do modelo pode variar. Por exemplo, *Altman (1968)* e *Deakin (1972)* definiram o termo falência como sendo exclusivamente legal. Por sua vez, *Beaver (1966)* e *Blum (1974)* consideraram a falência como sendo técnica.

Há ainda a necessidade de clarificar dois conceitos distintos, mas confundidos muitas vezes. O conceito de insolvência descrito acima em nada se iguala ao conceito de falência. Este último acontece quando o devedor é responsável por mais dívidas do que a quantidade de bens que possui. Uma pessoa/empresa falida não está necessariamente insolvente. Devido à inexistência do momento da falência na base de dados utilizada, neste estudo utilizou-se o conceito de insolvência de uma forma lata, não o diferenciando de falência. Desta forma, o

momento de falência não é o que se encontra na base de dados ou a data em que foram declarados como é usual em trabalhos desta temática. Destaco *Moro (2008)*, *Rijkers e Thibault (2009)*, *Altman e Sabato (2007)*, *Altman et al (2011)*. Tomada esta opção, estamos a assumir que o momento em que a empresa é considerada em insolvência é igual para todas as presentes na amostra evitando assim as diferenças que poderiam assentar no tempo do processo até a empresa se declarar insolvente ou até mesmo o tempo considerado pela base de dados.

2.2) Caraterização das PME's em Portugal

Em primeira instância, há que clarificar quais os requisitos que são necessários uma empresa ter para ser considerada PME. No caso da União Europeia, este conceito aplica-se às empresas que:

- Tenham entre 50 e 250 trabalhadores;
- Tenham um volume de negócios anual seja inferior ou igual a 50 milhões;
- Ou, que o balanço total anual não ultrapasse os 43 milhões.

(Fonte: IAPMEI¹)

As PME's, para muitos países, são consideradas como a “espinha dorsal” da economia. Assim, a previsão de falências neste tipo de empresas tornou-se numa área de estudo com relevo nos últimos tempos devido ao elevado número de PME's existentes e à importância que as mesmas têm no mercado.

Altman e Sabato (2007) utilizaram, pela primeira vez, a definição de PME's num estudo desta natureza. O objetivo era testar o comportamento dos rácios financeiros neste tipo de empresa. Analisaram um conjunto de rácios e tinham como objetivo identificar aqueles que mais influenciavam a estabilidade financeira de uma empresa. Foi concluído que a capacidade preditiva do modelo é significativamente maior quando usado com base em PME's (comparativamente ao modelo genérico que se baseia nas grandes empresas).

Com o avançar dos anos, vários foram os autores que continuaram a estudar este tipo de empresas tentando analisar diferentes características e variáveis para alcançar uma

¹ É o Instituto de Apoio às Pequenas e Médias Empresas e à Inovação). Foi fundado em 1975 e tem como objetivo promover o crescimento empresarial das micro, pequenas e médias empresas.

melhor capacidade preditiva: *Gupta et al., 2014* introduziram variáveis *corporate governance* e informação contida em cash-flows; *Ciampi, 2015* realizou um estudo que continha tanto variáveis de *corporate governance* como de rácios financeiros, centrado em empresas italianas.

Um estudo realizado pelo INE em 2018-2019 mostra que, atualmente, as PME's ainda se afirmam face às grandes empresas e, estão presentes em grande número em Portugal.

Tabela 1 – Principais indicadores das sociedades não financeiras, por dimensão (2018-2019)

	Sociedades		Pessoal ao serviço		Volume de negócios		VAB		Gastos com o pessoal		EBE	
	2019	TV. 18-19	2019	TV. 18-19	2019	TV. 18-19	2019	TV. 18-19	2019	TV. 18-19	2019	TV. 18-19
	Nº	%	Nº	%	10 ⁶ Euros	%	10 ⁶ Euros	%	10 ³ Euros	%	10 ⁶ Euros	%
Dimensão												
PME	437 668	6,1	2 340 537	4,4	229 128	4,5	60 058	7,6	39 546	8,0	20 632	6,6
Grandes	1 291	7,7	918 470	5,9	167 694	3,8	36 771	3,9	21 131	10,6	15 488	-3,2

Fonte: INE

Fazendo uma breve análise da tabela acima conclui-se que o desempenho das PME's superou o das grandes empresas.

Desta forma, fica comprovada a importância vital das PME's na economia portuguesa: representam quase a totalidade das empresas e apresentam um volume de negócios que assume ser mais de metade do total a nível nacional. No entanto, continuam também evidentes as diferenças em relação às grandes empresas. Pode-se assumir que as PME's são menos diversificadas e apresentem uma menor produtividade. Talvez por estes motivos exista um menor interesse por parte de investidores externos e, por isso, estejam sujeitas a condições de financiamento mais desvantajosas face às grandes empresas

Perante todos estes fatores, reafirma-se a importância de estudar a previsão da falência nas PME's como já referido no início deste ponto. Este é um tema que já tem vindo a ser desenvolvido por diversos autores, tais como: *Edmister (1972)*, *Phillips e Kirchhoff (1989)*, *Everett e Watson (1998)*, *Lehmann (2003)*, *Altman e Sabato (2007)*, *Altman et al (2010)* *Cultrera e Brédart (2016)*.

2.3) Modelos de previsão de falência e a sua evolução

Com o passar dos anos foram diversos os estudos que surgiram relativamente à temática que se pretende abordar neste trabalho. Contudo, não existe um consenso acerca de quais as variáveis e quais os modelos que apresentam uma maior capacidade preditiva.

Ao longo das últimas duas décadas os métodos de previsão de falência inteligentes têm vindo a obter uma maior popularidade face aos métodos tradicionais estatísticos. Progressivamente, houve a carência de existirem modelos que se adaptassem às necessidades das empresas. Outros fatores para além dos financeiros precisavam de ser considerados de forma a existirem modelos específicos, adequados à tipologia de cada negócio. O início do século XX ficou marcado por efeitos associados à falência empresarial, o que impulsionou os estudos acerca da mesma. Estes desenvolvimentos tornaram-se relevantes a nível da economia mundial uma vez que sempre existiu, e existia a necessidade de antever as dificuldades financeiras por parte dos stakeholders².

Apesar da variedade crescente de métodos inteligentes, existem algumas técnicas que se destacam: as Redes Neutrais Artificiais (RNA), métodos baseados em machine learning, as máquinas de vetores de suporte³, a árvore de decisão entre outras que demonstram alguma superioridade na questão da previsão de falências.

Com o avanço da tecnologia torna-se imperativo falar dos métodos de previsão de falência inspirados no mecanismo do cérebro, aplicando sistemas computacionais. Estes novos modelos eram baseados em técnicas não paramétricas. Aqui, distingo as redes neutrais artificiais. Este é o método mais popular desde os anos 90. O modelo baseia-se em mecanismos computacionais de processamento de informação que, à semelhança do cérebro humano, são capazes de processar várias informações em simultâneo, o que possibilita um maior cruzamento dos dados. Uma outra característica deste modelo é que tem a capacidade de aprender com acontecimentos passados (*Boritz et al., 1995*), possibilitando, assim, detetar tendências e padrões de forma mais rápida. Os modelos que se basearam nas RNA devido à sua capacidade de extração de dados são bastante desejados. No entanto, dado que a capacidade do modelo se prende com a experiência são necessárias atualizações constantes

² Pessoa ou grupo que detêm algum tipo de participação e interessa por uma determinada empresa.

³ É um modelo bastante flexível: tanto pode ser utilizado para a regressão como para a classificação.

de maneira a estarem atuais. O último artigo de *Alaka* (*Alaka et al., 2018*) enumera 38 artigos de autores que utilizaram, de alguma forma, as redes neutrais.

Com base no artigo “*Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets*” é possível perceber a importância da *machine learning*. É uma técnica que faz com que as máquinas aprendam com os dados o que resulta em análises preditivas mais eficientes e, tem sido cada vez mais utilizada para a construção de modelos analíticos. Estas análises com uma maior capacidade preditiva são cada vez mais importantes dado ao crescimento exponencial tecnológico a que se tem assistido e à variedade de dados partilhados na rede. Estamos perante uma nova era e, desta forma, torna-se imprescindível o uso de ferramentas computacionais (*Zoričák, Gnip, Drotár, & Gazda, 2020*).

Um outro modelo que surgiu para tentar corrigir algumas falhas por parte dos modelos tradicionais estatísticos foi o modelo de Regressão Logística de Mínimos Quadrados Parciais (PSL-LR) - *Jabeur (2017)*. Este modelo pode ser obtido através de regressões simples e múltiplas e, caracteriza-se por ser não linear e relacionar uma única variável dependente a um conjunto de variáveis independentes quantitativas ou qualitativas. É considerado um modelo inovador sobretudo, pela inclusão de dois aspetos principais: o uso de todas as variáveis para evitar a falência e, além disso, poderá ser uma alternativa para o problema do algoritmo de convergência para a estimativa da máxima verosimilhança.

No que toca aos métodos inteligentes de previsão de falência é ainda de realçar os processos gaussianos (*Antunes, et al., 2017*). Raramente foram realizados estudos probabilísticos, embora existam inúmeros estudos acerca da previsão de falência. Esta é a grande diferença relativamente a outras abordagens: a previsão de cada classe é sob a forma de probabilidade. Esta classificação probabilística permite decifrar quão certo é ou não o modelo em relação ao estado de falência de uma empresa. De acordo com o autor, o processo gaussiano fornece um quadro muito flexível o que melhora o potencial de modelação de relações complexas não lineares entre rácios contabilísticos e risco de falência em simultâneo. Este modelo é ainda capaz de lidar com a incerteza presente no conjunto de dados de falências estudados e, apresenta como característica, a sua simplicidade – sendo caracterizado pelas suas funções de média e covariância.

No entanto, apesar de toda a inovação tecnológica que se tem vindo a adquirir desde que se começou a debater sobre esta temática, é imprescindível analisar toda a literatura concluída por aqueles que são considerados os pioneiros da temática da previsão de falência. Ao nível dos modelos de previsão, vão ser abordadas três fases de desenvolvimento:

1. O Modelos Univariado – modelo de Beaver (1966)
2. O Modelo Multivariado – modelo de Altman (1968)
3. O Modelos de Regressão Logística – modelo de Ohlson (1980)

William Beaver foi quem apresentou o primeiro grande contributo no que toca ao tema da previsão de falências. O seu estudo era baseado em dois pressupostos: 1) os rácios podem ser utilizados para prever a falência, 2) estes não devem ser escolhidos indiscriminadamente, uma vez que, à partida, uns terão maior capacidade de previsão do que outros. Foi então pioneiro da análise univariada de rácios financeiros onde se focou na análise de cada variável isoladamente. Neste tipo de análise existe a vantagem da facilidade com que pode ser aplicado uma vez que apenas depende da análise isolada da variância das variáveis escolhidas.

A novidade deste estudo face aos realizados até então (estudos baseados na análise de perfis) assenta na análise “multi-ratio” com a abordagem a vários rácios, ainda que tratados separadamente, e pelo facto da informação incluída se tratar de informação de mercado.

Neste estudo foi utilizada uma amostra de 79 empresas que tinham falido entre os anos de 1954 e 1964, e de 79 que estavam na situação contrária. Assim, era possível comparar os dois grupos e determinar em qual deles é que as empresas se inseriam, utilizando valores médios e recorrendo a testes que avaliavam a capacidade do rácio. Foram utilizados 30 rácios, posteriormente subdivididos em cinco categorias: rácios de cash-flow, rácios de rentabilidade, rácios de dívida, rácios de liquidez e rácios de turnover (capacidade de uma empresa utilizar os seus capitais próprios para gerar vendas).

O autor conclui que nas empresas que faliam, à medida que o ano da falência se aproximava, essas apresentavam deteriorações nos seus rácios, algo que não acontecia nas empresas não falidas. Este modelo é considerado um modelo universal uma vez que é independente da indústria e da dimensão, podendo este estudo prever falência com mais de 86,14% de precisão. De acordo com Sheppard (1994), este modelo atingiu um nível moderado de precisão. Contudo, “(...) *it did not provide a measure of the relevant risk*” (Stickney, 1996, p. 507).

Durante a realização deste estudo houve a necessidade de interseção de vários tipos de variáveis financeiras de maneira a se obter uma análise mais pormenorizada. Concluiu-se então que a análise univariada satisfaz as necessidades de uma investigação numa fase inicial.

Contudo, requer mais consistência para estudos com mais exigência dado à metodologia intrínseca do modelo.

Foi então em 1968 que Edward Altman surgiu com o primeiro estudo multivariado, também aplicado à análise de rácios financeiros. Altman, apresentou uma análise discriminante multivariada, denominada de MDA, que se definiu a partir de cinco rácios de rentabilidade e de risco em empresas falidas e não falidas. Inicialmente utilizou um conjunto de 22 rácios, para uma amostra de 66 empresas (33 falidas e as restantes 33 não falidas), num espaço temporal entre 1946 e 1965. Posteriormente, e através do tratamento estatístico e análise de quais os mais significativos, chegou a cinco rácios – de liquidez, de rentabilidade, de alavancagem, de solvência e de atividade. Este autor ficou sobretudo conhecido pelo seu modelo Z-score, o mais popularizado até à data. Este modelo, com recurso aos rácios descritos acima, era capaz de prever com uma enorme percentagem de êxito nos dois anos anteriores à falência. Conclui que os resultados obtidos a partir de uma análise multivariada são mais significativos do que aqueles que são feitos por uma análise univariada.

O propósito deste - tão falado - modelo é atingir um rácio geral que irá servir de base à classificação dos dois grupos de empresas: falidas e não falidas. O modelo Z-score classificou corretamente 94% das empresas falidas e 97% das empresas não falidas (isto um ano antes da falência).

Fórmula do modelo Z-score:

$$\text{Z-Score} = 0.012 (X1) + 0.014 (X2) + 0.033 (X3) + 0.006 (X4) + 0.999 (X5)$$

Em que:

- X1: $\frac{\text{Fundo de maneo}}{\text{Ativo total}}$
- X2: $\frac{\text{Resultados transitados}}{\text{Ativo total}}$
- X3: $\frac{\text{EBIT}}{\text{Ativo total}}$
- X4: $\frac{\text{Valor de mercado}}{\text{Total do passivo}}$
- X5: $\frac{\text{Vendas}}{\text{Ativo total}}$

Apesar deste modelo ainda hoje ser um dos instrumentos mais utilizados, alguns autores apontavam que o uso desta prática violava dois pressupostos base do MDA e, por isso, iam sendo levantadas questões sobre a adequabilidade desta técnica estatística. Um dos pressupostos era a matriz de variância-covariância que deverá ser homogênea nos dois grupos e, o segundo, relacionava-se com o facto de as variáveis explicativas terem que apresentar uma distribuição normal multivariada. Eram ainda apontadas como falhas os erros obtidos na classificação dos resultados e na aplicabilidade dos modelos (*Pereira, Silva, Basto, & Santos, 2016*). Por este motivo houve a necessidade de explorar outras metodologias com o objetivo de contornar as limitações apontadas. Havia ainda o objetivo da maximização da capacidade de previsão. No início da década de 80, surgem os primeiros modelos logísticos para previsão de falência.

Ohlson, como mencionado acima, considera que o resultado obtido na análise discriminante é pouco intuitivo, uma vez que apenas separa as empresas em dois grupos, o que não acontece com o resultado da regressão logística. Assim, desenvolveu um modelo que utiliza na previsão de falência das empresas a análise logística, reconhecendo a existência deste mesmo problema. Foi o pioneiro na utilização desta técnica econométrica e, possibilitou que, com a sua utilização, não se violasse os pressupostos do MDA e o uso de amostras que poderiam ser desproporcionais. Nesta análise utilizou um modelo com nove fatores, em que sete deles eram rácios financeiros e em que os dois restantes eram variáveis binárias. Utilizou ainda uma amostra de 2163 empresas, das quais 105 diziam respeito a empresas falidas, entre os anos de 1970 e 1976.

Diferenciou-se ainda pela base de dados utilizada. Utilizou a *10-K*, ao invés de utilizar a *Moody's Manual* como os anteriores. A base de dados que utilizou exibiu a vantagem de apresentar a data de disponibilização das demonstrações financeiras.

Neste estudo, o autor apresentou três modelos: um ano antes da insolvência, dois anos antes da insolvência e, entre um e dois anos antes da insolvência. De acordo com a tabela abaixo, verifica-se que os modelos obtiveram uma alta capacidade preditiva:

Tabela 2 – Capacidade Preditiva do Modelo de Ohlson

Modelos	Capacidade Preditiva
1 ano antes da falência	96,12%
2 anos antes da falência	95,55%
1 e 2 anos antes da falência	92,84%

Fonte: Adaptado de Ohlson (1980)

A análise logística utiliza a função de probabilidade acumulada (dada por $F(Z)$) para prever a falência. Pode assumir valores entre 0 e 1 podendo ser escrita da seguinte forma:

$$p[v_i = 1] = \frac{1}{1 + e^{-Z_i}} = \frac{1}{1 + e^{-X_i\beta}}$$

Sendo Z representado pela relação linear: $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$

Ohlson, conseguiu definir as relações entre as variáveis pela aplicação dos princípios da probabilidade condicional, através do seu método não linear. Referia que utilizar a informação publicada depois da constatação da insolvência sobrevalorizava a capacidade preditiva dos modelos. Ainda assim, os resultados da estimação mostraram-se menos eficazes do que os de Altman. O autor argumentou que esta discrepância se devia ao facto de o tempo correspondente entre a data de observação e o momento da falência ser maior no seu caso e, seriam necessárias novas e melhores variáveis explicativas para conseguir aumentar a capacidade do modelo.

Em conclusão, os modelos de análise estatística no que toca à temática da previsão de falências continuam a ser os métodos mais utilizados até à data, apesar de terem surgido métodos mais evoluídos devido ao desenvolvimento que houve. A fácil interpretação dos dados e a fácil aplicação, não esquecendo a boa capacidade preditiva, são os principais fatores. O modelo de Altman (Z-Score) é o método que mais adaptado foi para utilização: em diversos países, setores e diferentes intervalos temporais.

3. Metodologia

3.1) Descrição dos dados

A base de dados utilizada será a SABI – que contempla a análise financeira de empresas portuguesas e contém um histórico de contas anuais até 25 anos. Os dados irão ser analisados em painel, dado que são referentes a várias empresas, durante vários anos.

Os dados da amostra retirados são relativos ao período entre 2015-2019. Foi considerado que o período em análise não deverá ter em conta os principais anos de crise económica uma vez que os valores das variáveis podem ser alterados devido a este fator macroeconómico que teve início no ano de 2007.

Para a pesquisa da amostra deste trabalho utilizaram-se alguns critérios, tais como: de localização, sendo que só foram escolhidas empresas portuguesas; com o estado de atividade ativo (40 empresas) e outras com o estado de insolvência (outras 40); e que apresentassem pelo menos três anos consecutivos de contas. À posteriori foi adicionada uma nova variável, mantendo a amostra e restantes variáveis utilizadas, para o desenvolvimento de um segundo modelo, tendo igualmente por base o original de Altman. A variável adicionada – rácio de cobertura EBIT – foi considerada como importante uma vez que é utilizada para avaliar a durabilidade financeira de uma empresa, analisando se a mesma é, pelo menos, suficientemente rentável para cumprir com as suas obrigações com juros utilizando as receitas geradas antes de impostos. Quanto mais elevado o valor do rácio, melhor. Embora o rácio ideal possa variar com o tipo de indústria.

As empresas selecionadas têm em conta o critério da definição europeia de PME's, descrito no capítulo 2.

3.2) Amostra

A amostra que servirá de base à análise econométrica considera 80 empresas, como já mencionado anteriormente, sendo que o período em análise abrange um período de cinco anos para cada empresa, à exceção das empresas em que apenas existem dados de um período inferior. Os dados recolhidos apresentam uma frequência anual e dizem respeito a 40 empresas ativas e 40 empresas insolventes.

Os dados da amostra em estudo são referentes a PME's portuguesas e são obtidos da base de dados SABI. Desta forma, a credibilidade dos dados encontra-se assegurada na medida em que são obtidos através de uma fonte considerada segura.

3.3) Seleção das variáveis

3.3.1) Variáveis dependentes:

Nos modelos de escolha binária, onde está inserido o modelo logístico, a variável dependente refere-se ao estado da empresa, ou seja, permite distinguir as empresas ativas das empresas insolventes. Esta variável vai definir-se sob a forma dummy e apresenta o valor unitário para as empresas insolventes e o valor nulo para as empresas ativas que, neste caso, são aquelas que não se apresentam numa situação de falência. Esta variável é a variável dependente dos modelos a estimar dado que se pretende verificar os fatores que afetam o facto de uma empresa entrar numa situação de insolvência.

- 1 = “falidas”
- 0 = “não falidas”

3.3.2) Variáveis independentes:

As variáveis independentes ou explicativas são as que se espera que permitam explicar as variações existentes na variável dependente. Estas variáveis podem ser quantitativas ou qualitativas, sendo que nestas últimas pode recorrer-se a variáveis dummy. A seleção das variáveis explicativas é efetuada com recurso à bibliografia existente e ao que se espera que influencie a variável dependente, de acordo com a evolução do tecido empresarial e, consequentemente, da situação financeira das mesmas.

A escolha das variáveis sempre foi uma questão levantada pelos autores em estudos anteriores. Isto deve-se ao facto de existir um grande número de variáveis consideradas como indicadores significativos no que toca aos problemas empresariais. Desta forma, os autores levantavam a questão de quais é que influenciavam mais no modelo que se queria estimar. Assim, no presente trabalho foram utilizadas como variáveis explicativas as mesmas que Altman utilizou no seu modelo Z-Score, selecionadas de acordo com os cinco grupos de rácios financeiros definidos pelo autor: rácios de liquidez, de atividade, de rentabilidade, de alavancagem e de cobertura.

Tabela 3 – Variáveis Independentes

	X1	X2	X3	X4	X5
Variáveis	Fundo de manejo/ Ativo Total	Resultados transitados/ Ativo Total	Resultados operacionais/ Ativo total	Valor de Mercado dos Capitais Próprios/ Passivo Total	Vendas/ Ativo Total

Fonte: Elaboração própria

X1 = Esta variável mede o peso dos ativos líquidos ou fundo de maneio relativamente à totalidade do balanço da empresa. Assim, permite verificar se a liquidez das empresas afeta o facto das mesmas estarem ativas ou tornarem-se insolventes. A inclusão desta variável vai de encontro ao estudo de *Mervin*, que considerou este como o melhor indicador da descontinuidade.

X2 = Esta variável mede a capacidade de as empresas gerarem lucros tendo em conta o uso que fazem do ativo que detém. Assim, pretende-se verificar o impacto dos resultados obtidos pelas empresas nos diversos períodos e acumulação dos mesmos no facto das mesmas estarem ativas ou se tornarem insolventes. Poderá desfavorecer as empresas mais novas, apresentando um menor valor em X2, o que aumenta a probabilidade de serem classificadas como insolventes face a empresas mais velhas, *ceteris paribus*.

X3 = Esta variável mede a eficiência operacional das empresas em gerar lucros, sem considerar os impostos e alavancagem financeira, tendo em conta o uso que fazem do ativo que detém. Assim, pretende-se verificar o impacto dos resultados antes de impostos, os quais são de curto prazo, no facto das mesmas estarem ativas ou se tornarem insolventes.

X4 = Esta variável mede o peso do valor das empresas em gerarem lucros no total do passivo das mesmas, utilizado com vista ao financiamento das atividades. Assim, pretende-se verificar o impacto deste rácio no facto das mesmas estarem ativas ou se tornarem insolventes.

X5 = Esta variável mede o *turnover*, ou seja, o peso das vendas no total do ativo da empresa. Assim, pretende-se verificar o impacto deste rácio no facto das mesmas estarem ativas ou se tornarem insolventes. É considerada uma medida da capacidade de gestão para lidar com as condições competitivas.

3.3.3) Introdução de uma nova variável (X6):

De maneira a tentar aumentar a capacidade preditiva do modelo de Altman, irá ser introduzida uma nova variável: EBIT/ Interest Expenses.

Esta variável mede o peso do resultado antes de impostos e alavancagem financeira no total de juros suportados com essa alavancagem financeira. Assim, pretende-se verificar o impacto deste rácio no facto das mesmas estarem ativas ou se tornarem insolventes. As empresas que têm, constantemente, um EBIT inferior às despesas de juros, acabam por enfrentar problemas de solvência. Em geral, quanto mais elevado for este rácio, melhor está a saúde financeira da empresa.

O rácio de cobertura de juros informa a curto prazo. Ainda que seja sempre complicado fazer previsões das empresas a longo prazo, esta fórmula pode ajudar a perceber se a dívida se está a tornar um fardo para a empresa, em vez de um passivo que está a utilizar para financiar o seu crescimento.

A introdução desta variável explicativa aparece na medida em que, até ao momento, poucos foram os estudos que tiveram a inclusão da mesma. Em 1977, Altman incluiu-a no seu modelo, desenvolvido em parceria com a empresa Zeta Services. Este modelo foi aplicado às indústrias de produção e retalho. Também no ano de 2007, Altman e Sabato apresentaram um importante contributo neste tema, num período entre 1994-2002, em que tiveram em consideração esta mesma variável. Neste caso foram igualmente seleccionadas cinco categorias de rácios: de liquidez, rendibilidade, financiamento, cobertura e atividade.

Tabela 4 – Variáveis introduzidas no modelo de Altman e Sabato (2007)

Variables Examined	Accounting Ratio Category (in beginning)	Variables Manually Selected	Variables Entered in the Model	Accounting Ratio Category (in the end)
<i>X1=Short Term Debt/Equity (book value)</i>		X1		
<i>X2=Equity (book value)/Total Liabilities</i>			X1	
<i>X3=Liabilities/Total Assets</i>	Leverage	X3		Leverage
<i>X4=Cash/Total Assets</i>		X4		
<i>X5=Working Capital/Total Assets</i>		X5		
<i>X6=Cash/Net Sales</i>			X4	
<i>X7=Intangible/Total Assets</i>	Liquidity			Leverage
<i>X8=EBIT/Sales</i>				
<i>X9=EBITDA/Total Assets</i>		X9		
<i>X10=Net Income/Total Assets</i>				
<i>X11=Retained Earnings/Total Assets</i>	Profitability	X11		Liquidity
<i>X12=Net Income/Sales</i>				
<i>X13=EBITDA/Interest Expenses</i>		X13	X9	
<i>X14=EBIT/Interest Expenses</i>		X14		
<i>X15=Sales/Total Assets</i>	Coverage	X15		Profitability
<i>X16=Account Payable/Sales</i>			X11	
<i>X17=Account Receivable/Liabilities</i>	Activity	X17	X13	Profitability
Total		10	5	5

Fonte: Altman e Sabato (2007, p.341)

Tabela 5 – Sinais esperados das variáveis utilizadas

Falência (Y)	Sinal Esperado
Fundo de manio/ Ativo Total (X1)	-
Resultados transitados/ Ativo Total (X2)	-
Resultados operacionais/ Ativo total (X3)	-
Valor de Mercado dos Capitais Próprios/ Passivo Total (X4)	+
Vendas/ Ativo Total (X5)	+
EBIT/ Interest Expenses (X6)	+

Fonte: Elaboração própria

4. Hipóteses de Investigação

Feita a análise aos artigos estudados pelos autores tidos como referência, e considerando os resultados por eles obtidos, foram delineadas duas hipóteses de investigação.

H1: A regressão logística alcança uma maior capacidade preditiva em relação à análise discriminante.

Esta apresenta-se como a mais completa e surgiu para tentar contornar as limitações apontadas à análise discriminante multivariada. Segundo *Pereira, Silva, Basto, & Santos, 2016*, algumas das falhas da análise discriminante dizem respeito à distribuição das variáveis, à dispersão das matrizes de variância-covariância e ao cálculo individual da significância de cada uma das variáveis.

Assim, apesar de apresentar uma boa capacidade preditiva, foi utilizado o modelo Logit para a estimação dos dois modelos presentes neste trabalho.

H2: A variável introduzida - *EBIT/ Interest Expenses* - está positivamente relacionada com a capacidade preditiva do modelo.

Com o objetivo de diferenciar este estudo dos restantes pretendeu-se adicionar uma variável explicativa que até então não tem vindo a ser muito utilizada pelos autores que se debruçam sobre esta temática. Foi introduzida ao modelo com o objetivo de aumentar a capacidade preditiva do mesmo uma vez que esta avalia a saúde financeira de uma empresa.

É esperado um sinal positivo na relação entre esta nova variável explicativa e a falência.

5. Estimação do modelo

Neste capítulo vai ser descrito o processo de construção da base de dados e o respetivo processo de estimação dos modelos.

5.1) Base de dados

Para o desenvolvimento desta dissertação recorreu-se a um conjunto de dados em painel, de maneira a ser possível analisar informações de várias unidades amostrais, que neste trabalho se traduzem por ser as várias empresas que constituem a amostra.

Ao contrário dos dados seccionais, os dados em painel requerem uma análise mais complexa, Ramalho (2014). Os dados em painel permitem analisar os mesmos indivíduos em diferentes momentos, colocando em causa a independência entre as observações, sendo este um dos fatores necessários para garantir o enviesamento dos estimadores. Ainda assim, a estimação que utiliza dados em painel é mais completa e eficiente – Menezes (2017). No presente trabalho, como já referido, são analisadas diferentes variáveis, em diferentes períodos temporais.

Relativamente aos dados extraídos, todos os dados foram extraídos da base de dados SABI. Nesta extração estão incluídos os dados das 80 empresas da amostra, os dados da variável dependente – Falência – e das seis variáveis explicativas

5.1.1) Vantagens e limitações do modelo

A utilização de dados em painel permite uma análise dinâmica temporal dos comportamentos individuais de cada empresa, fazendo com que os modelos estimados apresentem informações mais credíveis. De acordo com Ramalho (2014) esta é uma das principais vantagens deste tipo de modelo.

Assim, existe um benefício maior nos dados económicos, como os analisados, uma vez que os dados em painel permitem alcançar uma melhor compreensão das dinâmicas de ajustamento dos modelos considerados aos dados em estudo, segundo Marques (2000). Para além da vantagem referida acima, recorrendo a Fernandes (2014), conclui-se que os dados

em painel apresentam ganhos de eficiência, facto que é causado pelo aumento da amostra inerente à consideração dos diferentes países. Pode-se ainda apontar como uma vantagem o recurso a um maior nível de informação, que se traduz no aumento da eficiência dos estimadores e, na diminuição da colinearidade.

Uma das limitações apontadas a este tipo de modelo é o facto de as previsões, por vezes, serem difícil de efetuar. A dificuldade centra-se na correlação entre os regressores e o termo do erro, o que faz com que exista enviesamento nos estimadores e falta de consistência, segundo Marques (2000).

5.2) Estimação

Para a estimação do modelo deste trabalho, de entre as várias técnicas estatísticas existentes, foi utilizado o modelo de regressão logística. Tal como foi falado no capítulo 2, no modelo Logit não é exigido que as variáveis independentes incluídas no modelo sigam uma distribuição normal multivariada e, que as matrizes de variâncias-covariâncias de cada um dos grupos sejam homogêneas. Coisas que, por sua vez, são necessárias num modelo de análise discriminante. Para além deste fator ter sido importante para a escolha de um modelo Logit, também foi decisivo que, na problemática do tema abordado – previsão de falências – a variável dependente seja do tipo dicotómico e, por isso, a aplicação de um modelo logístico é de fácil interpretação uma vez que a classificação resulta entre 0 e 1 e, facilmente se obtém uma probabilidade de falência. Para concluir, foi igualmente tido em consideração a possível interpretação separada dos coeficientes, bem como a significância individual das variáveis explicativas introduzidas. Posto isto, e tal como já referido nos capítulos anteriores, o objetivo primordial deste trabalho passa por desenvolver um modelo logístico de previsão de falência, aplicado às PME's portuguesas, sustentado pelas variáveis do modelo Z-score. Como segundo ponto será adicionada uma sexta variável, a fim de verificar se a capacidade preditiva aumenta ou diminuiu comparativamente ao primeiro modelo estimado.

A função logística é expressa da seguinte forma:

$$\hat{P}(Y_i = 1) = \frac{1}{1 + e^{-YFAL_i}} \quad \text{Em que,} \quad YFAL_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

Nesta fórmula, $\widehat{B_j}$ representa os coeficientes estimados; X_{ji} ($j=2,3,\dots,k$) as variáveis explicativas e $P(\widehat{Y_i} = 1)$ a probabilidade de a empresa ir à falência e YFAL diz respeito à variável dependente (como já referido assume valores entre 0 e 1). A previsão acerca de uma empresa entrar ou não em insolvência é feita no ponto cut-off de 0,5, que quer dizer que para valores da probabilidade estimada abaixo de 0,5 considera-se que a empresa não vai entrar em insolvência. Por outro lado, se o valor estimado for superior a 0,5 é considerado que a empresa vai entrar em insolvência.

Assim sendo, no presente trabalho utilizou-se o software EViews para estimar os dois modelos para os diferentes anos de previsão: um, dois ou três anos antes da insolvência. Os dados estão organizados em painel para um total de 80 empresas (40 não insolventes e 40 insolventes), num período temporal entre 2015 e 2019.

Para as restantes análises, foi utilizado o software SPSS.

6. Resultados

Neste capítulo vão ser descritos os resultados obtidos para os dois modelos analisados. Vão ainda ser apresentadas as correlações entre as variáveis e a análise MANOVA. Por fim, será exposta a capacidade preditiva alcançada neste trabalho e feita a comparação com os resultados obtidos pelos dois autores tidos como referência ao longo deste trabalho, Edward Altman e Ohlson.

Com a amostra recolhida e a base de dados feita, é possível realizar uma análise das estatísticas descritivas para cada uma das variáveis que vão ser utilizadas na estimação dos modelos, como se vê na tabela 6.

Tabela 6 – Estatísticas Descritivas

Empresas não insolventes						
	X1	X2	X3	X4	X5	X6
N	200	200	200	200	200	200
Média	0,177	2,675	0,054	1,055	2,479	9891,293
Desvio- Padrão	0,245	11,293	0,153	5,747	1,910	70171,889
Mínimo	-0,454	-0,458	-1,244	-0,610	0,000	-2860,955
Máximo	0,849	135,383	0,810	72,238	13,237	769016,900

Empresas insolventes						
	X1	X2	X3	X4	X5	X6
N	193	193	193	193	193	193
Média	0,184	0,452	-0,007	0,377	1,631	-1,736
Desvio Padrão	0,278	0,854	0,177	0,601	1,900	28,678
Mínimo	-1,163	-0,572	-1,283	-0,594	0,000	-262,553
Máximo	0,646	6,397	0,508	3,879	13,237	141,657

Variável dependente: Falência
Variáveis explicativas: X1, X2, X3, X4, X5 e X6
Fonte: Elaboração própria

Analisando o quadro geral, existe uma diferença generalizada entre os dois grupos de empresas analisadas (não insolventes e insolventes) relativamente ao valor médio das variáveis estudadas. Nas empresas não insolventes verifica-se uma média positiva em todas as variáveis explicativas, no entanto, tendo em conta o coeficiente de variação, os valores apresentam diferentes comportamentos, atingindo resultados negativos em alguns valores mínimos. No que diz respeito às empresas insolventes, o coeficiente de variação não apresenta resultados tão distintos. No entanto, contrariamente às empresas não insolventes, existem duas variáveis que apresentam uma média negativa (X3 e X6). A variável X6 é a que apresenta os maiores resultados no valor mínimo e no valor máximo, nos dois tipos de empresas analisadas (empresas não insolventes e empresas insolventes).

As variáveis de X2, X3, X4, X5 e X6 contêm resultados superiores no valor médio das empresas não insolventes. Isto diz-nos que, quanto maior for o valor do rácio analisado, menor é a probabilidade de a empresa entrar em falência.

Por sua vez, a variável X1 comporta-se de forma contrária. Apresenta um maior resultado no valor médio nas empresas insolventes o que faz com que os gestores tenham de estar atentos ao possível crescimento desta variável.

Na tabela 7 (abaixo) podemos encontrar as correlações entre as várias variáveis explicativas nos dois modelos testados. É importante recorrer-se à análise de correlação de Pearson entre as variáveis independentes para verificar a colinearidade entre elas. Não será bom se duas variáveis forem muito correlacionadas, uma vez que quer dizer que ao saber de uma, se sabe o que esperar da outra.

Tabela 7 – Correlações de variáveis

		X1	X2	X3	X4	X5	X6
X1	Correlação de Pearson	1					
X2	Correlação de Pearson	-,055	1				
X3	Correlação de Pearson	,271	,246	1			
X4	Correlação de Pearson	-,047	,420	,256	1		
X5	Correlação de Pearson	,035	-,122	,005	-,068	1	
X6	Correlação de Pearson	-,036	,196	,095	,136	-,054	1

Fonte: Elaboração própria

Desta forma, conclui-se que nenhum par de variáveis está fortemente correlacionado. As variáveis X1 e X4 destacam-se por apresentarem o valor mais alto de correlação. Por sua vez, existem 6 pares de correlações que apresentam valores negativos, o que nos indica que a variação de uma é inversa à variação da outra. Estes valores estão entre -0,122 e -0,036.

Dado que os resultados demonstrados na tabela mostram que não existe forte correlação entre as variáveis, conclui-se também que não existe tendência nos modelos apresentados para a multicolinearidade. Isto, por sua vez, significa que os dados não estão enviesados.

Como já referido, foi utilizado o software EViews para a estimação dos coeficientes para os modelos analisados, para os diferentes anos (um, dois e três anos antes da insolvência).

No que diz respeito às variáveis apresentadas, apenas metade apresentaram o sinal esperado (X2, X3 e X4). Segundo o teste do rácio de verosimilhança (teste LR), conclui-se que ambos os modelos analisados são marginalmente significativos a 1%.

Tabela 8 – Coeficientes da Regressão Logística

<i>Modelo 1:</i>						
Variáveis	Ano N-1		Ano N-2		Ano N-3	
	Coeficientes	Sig.	Coeficientes	Sig.	Coeficientes	Sig.
X1	0.994777	0.0643*	1.283935	0.0434**	1.624056	0.0423**
X2	-1.224760	0.0000***	-1.231923	0.0000***	-1.893463	0.0000***
X3	-4.391515	0.0019***	-6.178931	0.0003***	-5.989886	0.0174**
X4	2.241219	0.0000***	2.284935	0.0000***	2.938745	0.0000***
X5	-0.478425	0.0000***	-0.444413	0.0000***	-0.456433	0.0003***
Prob (LR Statistic)	0.000000		0.000000		0.000000	

Modelo 2:						
Variáveis	Ano N-1		Ano N-2		Ano N-3	
	Coefficientes	Sig.	Coefficientes	Sig.	Coefficientes	Sig.
X1	1.601709	0.0086***	2.196533	0.0036***	2.780681	0.0060***
X2	-1.696116	0.0000***	-1.903182	0.0000***	-1.512176	0.0050***
X3	-3.828011	0.0060***	-6.337654	0.0012***	-14.80348	0.0075***
X4	3.159181	0.0000***	3.566119	0.0000***	5.239631	0.0000***
X5	-0.593472	0.0000***	-0.589009	0.0000***	-0.488817	0.0013***
X6	-0.01085	0.2946	-0.000935	0.5584	-0.004155	0.5133
Prob (LR Statistic)	0.000000		0.000000		0.000000	

Variável dependente: Falência
Variáveis explicativas: X1, X2, X3, X4, X5 e X6
Fonte: Elaboração própria

Legenda:

*Variável significativa para uma significância de 10%

** Variável significativa para uma significância de 5%

*** Variável significativa para uma significância de 1%

Daqui resultam os modelos abaixo representados:

Modelo 1:

$$Z_1 = 0.994777 - 1.224760X_1 - 4.391515X_2 + 2.241219X_3 - 0.478425X_4$$

$$Z_2 = 1.283935 - 1.231923X_1 - 6.178931X_2 + 2.284935X_3 - 0.444413X_4$$

$$Z_3 = 1.624056 - 1.893463X_1 - 5.989886X_2 + 2.9387455X_3 - 0.456433X_4$$

Modelo 2:

$$Z_1 = 1.601709 - 1.696116X_1 - 3.828011X_2 + 3.159181X_3 - 0.593472X_4 - 0.01085X_5$$

$$Z_2 = 2.196533 - 1.903182X_1 - 6.337654X_2 + 3.566119X_3 - 0.589009X_4 - 0.000935X_5$$

$$Z_3 = 2.780681 - 1.512176X_1 - 14.80348X_2 + 5.239631X_3 - 0.488817X_4 - 0.004155X_5$$

Conforme a tabela 8 nos indica, verifica-se que as variáveis são estatisticamente significativas. As variáveis X2, X4 e X5 são estatisticamente significativas para o nível de significância de 1%. Existem apenas duas variáveis que são estatisticamente significativas a 10% (no modelo 1 a variável X1, em N-1; e no modelo 2 a variável X6, para todos os anos analisados). Para o nível de confiança de 95% destaca-se no modelo 1, a variável X1 para N-1 e N-2.

A variável adicionada no segundo modelo – X6 – não tem significância no modelo (em nenhum dos três anos analisados).

Tabela 9 – MANOVA

<i>Origem</i>	<i>Variável dependente</i>	<i>F</i>	<i>Sig</i>
<i>Falência</i>	X1	0,265	0,607
	X2	7,449	0,007
	X3	13,144	0,000
	X4	2,586	0,109
	X5	18,540	0,000
	X6	0,952	0,330

Variável dependente: Falência
Fonte: Elaboração própria

A análise MANOVA também é conhecida como sendo a análise multivariada da variância. É utilizada quando existem duas ou mais variáveis dependentes e serve para comparar médias amostrais multivariadas.

A partir da tabela acima (calculada pelo software SPSS), conclui-se que as variáveis X1, X4 e X6 **não são** estatisticamente significativas. Isto diz-nos que nestas variáveis não há diferenças significativas entre as empresas em falência e em não falência.

De acordo com a análise MANOVA percebe-se que existem diferenças significativas entre as empresas em falência e em não falência nas variáveis X2, X3 e X5.

Foi utilizado o nível de significância de 5%.

Na avaliação de uma empresa podem surgir dois tipos de erros: erros do tipo I e tipo II. Os erros do tipo I dizem respeito às empresas insolventes que são consideradas como ativas e, os erros do tipo II é a percentagem de empresas ativas que são classificadas como insolventes. Daqui concluímos que o modelo Logit, comparado com a análise discriminante multivariada, é mais capaz de prever a insolvência.

Tabela 10 – Capacidade Preditiva Erros tipo I e tipo II

<i>Modelo</i>	Ano N-1		Ano N-2		Ano N-3	
1:	Não insolvente	Insolvente	Não insolvente	Insolvente	Não insolvente	Insolvente
%	73.75%	84.31%	75.00%	84.07%	70.00%	80.82%
% correta						
%	26.25%	15.69%	25.00%	15.93%	30.00%	19.18%
% incorreta						
<i>Modelo</i>	Ano N-1		Ano N-2		Ano N-3	
2:	Não insolvente	Insolvente	Não insolvente	Insolvente	Não insolvente	Insolvente
% correta	74.38%	83.89%	70.05%	67.03%	78.75%	80.28%
%						
% incorreta	25.62%	16.11%	29.95%	32.97%	21.25%	19.72%

Fonte: Elaboração própria

A tabela 10 foi então elaborada com o propósito de simplificar a análise relativamente à capacidade preditiva dos dois modelos desenvolvidos, no que diz respeito à classificação das empresas como não insolventes ou insolventes.

Encontra-se a percentagem de casos que satisfazem a condição no cruzamento de cada linha com uma coluna. Por exemplo: para o modelo 1, com a previsão a um ano, verifica-se que 73.75% das empresas não insolventes foram classificadas corretamente como tal. Por sua vez, para o mesmo modelo, e para as mesmas empresas, 26.25% foram incorretamente classificadas como não insolventes.

Tabela 11 – Capacidade Preditiva - Final

	Ano N-1	Ano N-2	Ano N-3
<i>Modelo 1</i>	78.91%	79.40%	75.16%
<i>Modelo 2</i>	78.96%	68.62%	79.47%

Fonte: Elaboração própria

Os dois modelos comportam-se de maneira diferente no que toca aos anos de previsão analisados. No modelo 1 a capacidade preditiva para N-1 é de 78.91%, aumentando para 79.40% em N-2 e para N-3 diminui para 75.16%. Por sua vez, o segundo modelo para N-1 apresenta uma capacidade preditiva de 78.96%, diminuindo em N-2 para 68.62% e, por fim, apresenta um aumento para 79.47% em N-3.

Comparando os resultados obtidos em ambas as estimações, verifica-se que o segundo modelo (com a variável adicional) três anos antes da falência é o que apresenta uma melhor capacidade preditiva, apresentando um valor de 79.47%.

Por fim, foi elaborada uma última tabela que tem como único objetivo comparar a capacidade preditiva dos dois modelos estimados no presente trabalho com os modelos de dois dos grandes contribuidores para este tema, e que estiveram na base deste trabalho – Altman (Z-score) e Ohlson.

Tabela 12 – Capacidade Preditiva – Análise comparativa

Modelos	N-1	N-2	N-3
<i>Modelo Z-Score</i>	76.90%	72.61%	69.31%
<i>Modelo Ohlson</i>	76.82%	70.82%	71.67%
<i>Modelo 1 (5 variáveis)</i>	78.91%	79.40%	75.16%
<i>Modelo 2 (6 variáveis)</i>	78.96%	68.62%	79.47%

Fonte: Elaboração própria

Na tabela acima apresentada são exibidos os resultados da capacidade preditiva dos modelos para os três anos de análise. No que diz respeito aos dois modelos estimados neste trabalho, o resultado que apresenta uma capacidade preditiva mais baixa tem o valor de

68.62%, tendo o melhor resultado atingido os 79.47% - ambos no modelo 2 (com a variável adicional) nos anos N-2 e N-3, respetivamente. A nível global, e a título de comparação com os trabalhos anteriores, no ano anterior ao da insolvência o modelo 2 é o que apresenta uma maior capacidade preditiva, apenas com uma diferença de 0,05% face ao modelo 1. No ano N-2 é o modelo 1 que apresenta melhores resultados. Por fim, no ano N-3, volta a ser o modelo 2 que apresenta uma melhor capacidade preditiva, desta vez com uma margem de 4.31% face ao segundo modelo com melhores resultados.

Concluída a análise, é finalmente possível estabelecer uma relação com as hipóteses de investigação levantadas no início do trabalho e concluir se as mesmas se verificam. Assim, verifica-se H1. Através dos resultados da tabela 12 verificamos que, a utilização da regressão logística face à análise discriminante multivariada aumentou a capacidade preditiva do modelo. Também se verifica H2: a introdução da sexta variável - *EBIT/ Interest Expenses* – apesar de ter apresentado um sinal contrário ao esperado, exibiu um sinal positivo na sua relação com a falência. Para o ano anterior e para os três anos anteriores à insolvência, o modelo aumentou a sua capacidade de previsão em 0.05% e 4.31%, respetivamente.

7. Conclusão

O presente trabalho teve como propósito estudar a previsão de falência em PME's portuguesas através de uma regressão logística, tendo sido utilizadas as mesmas variáveis que Edward Altman utilizou no modelo Z-Score e foi composto por uma amostra de 80 empresas seleccionada da base de dados SABI que contém dados contabilísticos e financeiros de PME's portuguesas, no período entre 2015 a 2019.

No desenvolvimento do mesmo, foram tidos como referência sobretudo dois estudos: a análise multivariada de Altman (modelo Z-score) e o modelo logístico de Ohlson. A mistura de ferramentas destes dois estudos (variáveis de Altman e tipo de regressão utilizada por Ohlson) permitiu melhorar a capacidade preditiva, para o período analisado até aos três anos anteriores à insolvência.

Apesar dos inúmeros estudos acerca da Previsão de Falência e apesar dos estudos constantes dado a necessidade nesta temática, é um tema de difícil análise uma vez que o mundo empresarial está em constante evolução. Assim, é necessário compreender quais os modelos que alcançam uma maior capacidade preditiva face ao contexto do modelo que se pretende desenvolver.

Assim, o principal contributo da presente dissertação centra-se no facto da mesma ter sido desenvolvida com base em dois autores tomados como referência nesta temática. Foi escolhido o modelo econométrico desenvolvido por Ohlson e as variáveis seleccionadas por Altman.

Relativamente aos resultados obtidos, os mesmos suportaram as hipóteses de investigação. A regressão logística revela-se como o método econométrico mais adequado para as PME's portuguesas. Ao introduzir neste tipo de modelo as variáveis utilizadas por Altman os modelos apresentam capacidades preditivas superiores. No que toca à segunda hipótese, conclui-se primeiramente que a adição da sexta variável ($\frac{EBIT}{Interest\ Expenses}$) melhora a capacidade preditiva do modelo.

Uma vez que as demonstrações financeiras são a maior fonte acerca da atividade da empresa, é importante salientar que toda a metodologia utilizada no presente trabalho é baseada em rácios financeiros. Desta forma, poderá ser apontada aqui uma limitação do modelo, para além da exposta no capítulo 5, que é o facto de serem desvalorizados os componentes qualitativos que podem ter incidência nas empresas. Uma outra limitação é a definição do momento de falência. Dado que na base de dados utilizada não existem dados para o momento exato, de forma a evitar enviesamento da amostra, será importante encontrar dados mais exatos.

Para estudos futuros, é importante realçar alguns aspetos que não foram tidos em conta neste trabalho e podem ter influência. A amostra utilizada é pequena e, por isso, pode não identificar todos os dados possíveis – seria mais fiável ter uma amostra mais diversificada. Os modelos foram analisados tendo em conta os dados económicos das empresas que constituem a amostra. Assim, seria enriquecedor os trabalhos futuros terem em consideração variáveis macroeconómicas, que afetem, por exemplo, a conjuntura nacional como uma crise económica/social/política, uma pandemia, etc.

Por fim, conclui-se que os modelos desenvolvidos alcançaram resultados bastante significativos, contribuindo assim para o estudo da Previsão de Falências em PME's portuguesas.

8. Bibliografia

- Altman, E. (1968). FINANCIAL RATIOS, DISCRIMINANT ANALYSIS AND THE PREDICTION OF CORPORATE BANKRUPTCY. *The Journal Of Finance*, 23(4), 589-609.
- Altman, E. I., Sabato, G., & Wilson, N. (2010). The Value of Non-Financial Information in Small and Medium-Sized Enterprise Risk Management. *The Journal of Credit Risk*, 6, 1-33.
- Altman, E., & Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus*, 43(3), 332-357.
- Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing*, 60, 831-843.
- Beaver, W. (1966). Financial Ratios As Predictors of Failure. *Journal Of Accounting Research*, 4, 71.
- Blum, M. (1974). Failing Company Discriminant Analysis. *Journal Of Accounting Research*, 12(1), 1.
- Chen, K., & Shimerda, T. (1981). An Empirical Analysis of Useful Financial Ratios. *Financial Management*, 10(1), 51.
- Chen, M. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems With Applications*, 38(9), 11261-11272.
- Chen, Z., Chen, W., & Shi, Y. (2020). Ensemble learning with label proportions for bankruptcy prediction. *Expert Systems With Applications*, 146, 113155.
- Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. *Journal Of Business Research*, 68(5), 1012-1025.
- Cultrera, L., & Brédart, X. (2016). Bankruptcy prediction: the case of Belgian SMEs. *Review Of Accounting And Finance*, 15(1), 101-119.
- Deakin, E. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal Of Accounting Research*, 10(1), 167.
- Dislivro. (2009). *Código da insolvência e recuperação de empresas (CIRE)*. Lisboa.

Drotár, P., Gnip, P., Zoričák, M., & Gazda, V. (2019). Small- and medium-enterprises bankruptcy dataset. *Data In Brief*, 25, 104360.

Edmister, R. (1972). An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *The Journal Of Financial And Quantitative Analysis*, 7(2), 1477.

Fama, E., & French, K. (2002). Testing Trade-Off and Pecking Order Predictions About Dividends and Debt. *Review Of Financial Studies*, 15(1), 1-33.

Gissel Bellovary, J., & Johnstone, K. (2007). Descriptive Evidence from Audit Practice on SAS No. 99 Brainstorming Activities. *SSRN Electronic Journal*.

Grünberg, M., & Lukason, O. (2014). Predicting Bankruptcy of Manufacturing Firms. *International Journal Of Trade, Economics And Finance*, 93-97.

Gupta, J., Wilson, N., Gregoriou, A., & Healy, J. (2014). The value of operating cash flow in modelling credit risk for SMEs. *Applied Financial Economics*, 24(9), 649-660.

Keasey, K., & Watson, R. (1986). The Prediction of Small Company Failure: Some Behavioural Evidence for the UK. *Accounting And Business Research*, 17(65), 49-57.

Kumar Khatri, D. (2016). Financial Ratios, Econometrics and Prediction of Corporate Bankruptcy-an Empirical Study. *International Journal Of Accounting Research*, 4(1).

Lana, H. (2019). Alteração da lei de recuperação de empresas e falência: uma avaliação a partir de dados empíricos depois de mais de uma década de vigência. *Revista Do Curso De Direito Do UNIFOR*, 10(1), 15-45.

Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal Of Accounting Research*, 18(1), 109.

TUAN, K., & BORAK, M. (2020). Earnings Management and Financial Performance: An Empirical Investigation of the Istanbul Stock Exchange. *Sosyoekonomi*, 381-393.

Wu, Y., Gaunt, C., & Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal Of Contemporary Accounting & Economics*, 6(1), 34-45.

Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems With Applications*, 38(7), 8336-8342.

Zoričák, M., Gnip, P., Drotár, P., & Gazda, V. (2020). Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets. *Economic Modelling*, 84, 165-176.

9. Anexos

Anexo 1) Estimação modelo Logit – cinco variáveis N-1

Dependent Variable: FALENCIA

Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)

Sample (adjusted): 2004 2019

Included observations: 313 after adjustments

Convergence achieved after 8 iterations

Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.(Sig)
X1(-1)	0.994777	0.537596	1.850418	0.0643
X2(-1)	-1.224760	0.226020	-5.418807	0.0000
X3(-1)	-4.391515	1.416097	-3.101140	0.0019
X4(-1)	2.241219	0.423156	5.296440	0.0000
X5(-1)	-0.478425	0.089376	-5.352970	0.0000
C	0.913435	0.233202	3.916935	0.0001
McFadden R-squared	0.200547	Mean dependent var		0.488818
S.D. dependent var	0.500675	S.E. of regression		0.420238
Akaike info criterion	1.146216	Sum squared resid		54.21624
Schwarz criterion	1.218028	Log likelihood		-173.3828
Hannan-Quinn criter.	1.174914	Deviance		346.7655
Restr. deviance	433.7536	Restr. log likelihood		-216.8768
LR statistic	86.98807	Avg. log likelihood		-0.553939
Prob(LR statistic)	0.000000			
Obs with Dep=0	160	Total obs		313
Obs with Dep=1	153			

Expectation-Prediction Evaluation for Binary Specification

Equation:

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	118	24	142	160	153	313
P(Dep=1)>C	42	129	171	0	0	0
Total	160	153	313	160	153	313
Correct	118	129	247	160	0	160
% Correct	73.75	84.31	78.91	100.00	0.00	51.12
% Incorrect	26.25	15.69	21.09	0.00	100.00	48.88
Total Gain*	-26.25	84.31	27.80			
Percent Gain**	NA	84.31	56.86			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	103.07	56.93	160.00	81.79	78.21	160.00
E(# of Dep=1)	56.93	96.07	153.00	78.21	74.79	153.00
Total	160.00	153.00	313.00	160.00	153.00	313.00
Correct	103.07	96.07	199.14	81.79	74.79	156.58
% Correct	64.42	62.79	63.62	51.12	48.88	50.03
% Incorrect	35.58	37.21	36.38	48.88	51.12	49.97
Total Gain*	13.30	13.91	13.60			
Percent Gain**	27.21	27.21	27.21			

N-2

Dependent Variable: FALENCIA
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Sample (adjusted): 2005 2019
Included observations: 233 after adjustments
Convergence achieved after 8 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1(-2)	1.283935	0.635675	2.019796	0.0434
X2(-2)	-1.231923	0.255926	-4.813587	0.0000
X3(-2)	-6.178931	1.696742	-3.641643	0.0003
X4(-2)	2.284935	0.480272	4.757585	0.0000
X5(-2)	-0.444413	0.101713	-4.369298	0.0000
C	0.830633	0.264987	3.134618	0.0017
McFadden R-squared	0.209884	Mean dependent var		0.484979
S.D. dependent var	0.500850	S.E. of regression		0.421184
Akaike info criterion	1.146122	Sum squared resid		40.26897
Schwarz criterion	1.234990	Log likelihood		-127.5232
Hannan-Quinn criter.	1.181957	Deviance		255.0464
Restr. deviance	322.7963	Restr. log likelihood		-161.3981
LR statistic	67.74986	Avg. log likelihood		-0.547310
Prob(LR statistic)	0.000000			
Obs with Dep=0	120	Total obs		233
Obs with Dep=1	113			

Expectation-Prediction Evaluation for Binary Specification

Equation:

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	90	18	108	120	113	233
P(Dep=1)>C	30	95	125	0	0	0
Total	120	113	233	120	113	233
Correct	90	95	185	120	0	120
% Correct	75.00	84.07	79.40	100.00	0.00	51.50
% Incorrect	25.00	15.93	20.60	0.00	100.00	48.50
Total Gain*	-25.00	84.07	27.90			
Percent Gain**	NA	84.07	57.52			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	77.98	42.02	120.00	61.80	58.20	120.00
E(# of Dep=1)	42.02	70.98	113.00	58.20	54.80	113.00
Total	120.00	113.00	233.00	120.00	113.00	233.00
Correct	77.98	70.98	148.97	61.80	54.80	116.61
% Correct	64.99	62.82	63.93	51.50	48.50	50.05
% Incorrect	35.01	37.18	36.07	48.50	51.50	49.95
Total Gain*	13.48	14.32	13.89			
Percent Gain**	27.80	27.80	27.80			

N-3

Dependent Variable: FALENCIA
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Sample (adjusted): 2006 2019
Included observations: 153 after adjustments
Convergence achieved after 8 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1(-3)	1.624056	0.799900	2.030324	0.0423
X2(-3)	-1.893463	0.466443	-4.059365	0.0000
X3(-3)	-5.989886	2.517804	-2.379013	0.0174
X4(-3)	2.938745	0.676339	4.345076	0.0000
X5(-3)	-0.456433	0.127529	-3.579052	0.0003
C	0.872350	0.351699	2.480388	0.0131
McFadden R-squared	0.237866	Mean dependent var		0.477124
S.D. dependent var	0.501117	S.E. of regression		0.423914
Akaike info criterion	1.133378	Sum squared resid		26.41631
Schwarz criterion	1.252219	Log likelihood		-80.70342
Hannan-Quinn criter.	1.181653	Deviance		161.4068
Restr. deviance	211.7827	Restr. log likelihood		-105.8913
LR statistic	50.37583	Avg. log likelihood		-0.527473
Prob(LR statistic)	0.000000			
Obs with Dep=0	80	Total obs		153
Obs with Dep=1	73			

Expectation-Prediction Evaluation for Binary Specification

Equation:

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	56	14	70	80	73	153
P(Dep=1)>C	24	59	83	0	0	0
Total	80	73	153	80	73	153
Correct	56	59	115	80	0	80
% Correct	70.00	80.82	75.16	100.00	0.00	52.29
% Incorrect	30.00	19.18	24.84	0.00	100.00	47.71
Total Gain*	-30.00	80.82	22.88			
Percent Gain**	NA	80.82	47.95			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	53.02	26.98	80.00	41.83	38.17	80.00
E(# of Dep=1)	26.98	46.02	73.00	38.17	34.83	73.00
Total	80.00	73.00	153.00	80.00	73.00	153.00
Correct	53.02	46.02	99.04	41.83	34.83	76.66
% Correct	66.27	63.04	64.73	52.29	47.71	50.10
% Incorrect	33.73	36.96	35.27	47.71	52.29	49.90
Total Gain*	13.99	15.33	14.63			
Percent Gain**	29.32	29.32	29.32			

Anexo 2) Estimação modelo Logit – seis variáveis

N-1

Dependent Variable: FALENCIA
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Sample (adjusted): 2004 2019
Included observations: 309 after adjustments
Convergence achieved after 28 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1(-1)	1.601709	0.609998	2.625762	0.0086
X2(-1)	-1.696116	0.289130	-5.866279	0.0000
X3(-1)	-3.828011	1.393970	-2.746121	0.0060
X4(-1)	3.159181	0.543185	5.816034	0.0000
X5(-1)	-0.593472	0.103854	-5.714491	0.0000
X6(-1)	-0.001085	0.001035	-1.048001	0.2946
C	0.987648	0.251861	3.921402	0.0001
McFadden R-squared	0.272671	Mean dependent var		0.482201
S.D. dependent var	0.500494	S.E. of regression		0.400248
Akaike info criterion	1.052678	Sum squared resid		48.37995
Schwarz criterion	1.137252	Log likelihood		-155.6387
Hannan-Quinn criter.	1.086491	Deviance		311.2774
Restr. deviance	427.9733	Restr. log likelihood		-213.9866
LR statistic	116.6959	Avg. log likelihood		-0.503685
Prob(LR statistic)	0.000000			
Obs with Dep=0	160	Total obs		309
Obs with Dep=1	149			

Expectation-Prediction Evaluation for Binary Specification

Equation:

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	119	24	143	160	149	309
P(Dep=1)>C	41	125	166	0	0	0
Total	160	149	309	160	149	309
Correct	119	125	244	160	0	160
% Correct	74.38	83.89	78.96	100.00	0.00	51.78
% Incorrect	25.62	16.11	21.04	0.00	100.00	48.22
Total Gain*	-25.62	83.89	27.18			
Percent Gain**	NA	83.89	56.38			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	109.33	50.67	160.00	82.85	77.15	160.00
E(# of Dep=1)	50.67	98.33	149.00	77.15	71.85	149.00
Total	160.00	149.00	309.00	160.00	149.00	309.00
Correct	109.33	98.33	207.66	82.85	71.85	154.70
% Correct	68.33	65.99	67.20	51.78	48.22	50.06
% Incorrect	31.67	34.01	32.80	48.22	51.78	49.94
Total Gain*	16.55	17.77	17.14			
Percent Gain**	34.32	34.32	34.32			

N-2

Dependent Variable: FALENCIA

Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)

Sample (adjusted): 2005 2019

Included observations: 229 after adjustments

Convergence achieved after 27 iterations

Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1(-2)	2.196533	0.755314	2.908106	0.0036
X2(-2)	-1.903182	0.367069	-5.184806	0.0000
X3(-2)	-6.337654	1.962136	-3.229976	0.0012
X4(-2)	3.566119	0.692536	5.149359	0.0000
X5(-2)	-0.589009	0.123044	-4.786964	0.0000
X6(-2)	-0.000935	0.001598	-0.585185	0.5584
C	0.898879	0.296016	3.036587	0.0024
McFadden R-squared	0.303852	Mean dependent var		0.475983
S.D. dependent var	0.500517	S.E. of regression		0.395025
Akaike info criterion	1.024595	Sum squared resid		34.64193
Schwarz criterion	1.129556	Log likelihood		-110.3161
Hannan-Quinn criter.	1.066939	Deviance		220.6323
Restr. deviance	316.9328	Restr. log likelihood		-158.4664
LR statistic	96.30052	Avg. log likelihood		-0.481730
Prob(LR statistic)	0.000000			
Obs with Dep=0	120	Total obs		229
Obs with Dep=1	109			

Expectation-Prediction Evaluation for Binary Specification

Equation:

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	94	20	114	120	109	229
P(Dep=1)>C	26	89	115	0	0	0
Total	120	109	229	120	109	229
Correct	94	89	183	120	0	120
% Correct	78.33	81.65	79.91	100.00	0.00	52.40
% Incorrect	21.67	18.35	20.09	0.00	100.00	47.60
Total Gain*	-21.67	81.65	27.51			
Percent Gain**	NA	81.65	57.80			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	84.07	35.93	120.00	62.88	57.12	120.00
E(# of Dep=1)	35.93	73.07	109.00	57.12	51.88	109.00
Total	120.00	109.00	229.00	120.00	109.00	229.00
Correct	84.07	73.07	157.13	62.88	51.88	114.76
% Correct	70.05	67.03	68.62	52.40	47.60	50.12
% Incorrect	29.95	32.97	31.38	47.60	52.40	49.88
Total Gain*	17.65	19.43	18.50			
Percent Gain**	37.09	37.09	37.09			

N-3

Dependent Variable: FALENCIA

Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)

Sample (adjusted): 2006 2019

Included observations: 151 after adjustments

Convergence achieved after 19 iterations

Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1(-3)	2.780681	1.011236	2.749783	0.0060
X2(-3)	-1.512176	0.538517	-2.808035	0.0050
X3(-3)	-14.80348	5.539105	-2.672540	0.0075
X4(-3)	5.239631	1.268809	4.129568	0.0000
X5(-3)	-0.488817	0.152151	-3.212715	0.0013
X6(-3)	-0.004155	0.006356	-0.653692	0.5133
C	0.297288	0.389416	0.763419	0.4452
McFadden R-squared	0.353757	Mean dependent var		0.470199
S.D. dependent var	0.500772	S.E. of regression		0.377019
Akaike info criterion	0.986302	Sum squared resid		20.46859
Schwarz criterion	1.126176	Log likelihood		-67.46578
Hannan-Quinn criter.	1.043126	Deviance		134.9316
Restr. deviance	208.7937	Restr. log likelihood		-104.3969
LR statistic	73.86216	Avg. log likelihood		-0.446793
Prob(LR statistic)	0.000000			
Obs with Dep=0	80	Total obs		151
Obs with Dep=1	71			

Expectation-Prediction Evaluation for Binary Specification

Equation:

Success cutoff: C = 0.5

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	63	14	77	80	71	151
P(Dep=1)>C	17	57	74	0	0	0
Total	80	71	151	80	71	151
Correct	63	57	120	80	0	80
% Correct	78.75	80.28	79.47	100.00	0.00	52.98
% Incorrect	21.25	19.72	20.53	0.00	100.00	47.02
Total Gain*	-21.25	80.28	26.49			
Percent Gain**	NA	80.28	56.34			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	58.41	21.59	80.00	42.38	37.62	80.00
E(# of Dep=1)	21.59	49.41	71.00	37.62	33.38	71.00
Total	80.00	71.00	151.00	80.00	71.00	151.00
Correct	58.41	49.41	107.83	42.38	33.38	75.77
% Correct	73.02	69.60	71.41	52.98	47.02	50.18
% Incorrect	26.98	30.40	28.59	47.02	52.98	49.82
Total Gain*	20.04	22.58	21.23			
Percent Gain**	42.62	42.62	42.62			