

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

ROBERTO ZANCANER COSTA

**PREDIÇÃO DE RECUPERAÇÃO JUDICIAL DE EMPRESAS DO SETOR
SUCROALCOOLEIRO NO BRASIL UTILIZANDO ANÁLISE DISCRIMINANTE E
REGRESSÃO LOGÍSTICA**

SÃO PAULO

2020

ROBERTO ZANCANER COSTA

**PREDIÇÃO DE RECUPERAÇÃO JUDICIAL DE EMPRESAS DO SETOR
SUCROALCOOLEIRO NO BRASIL UTILIZANDO ANÁLISE DISCRIMINANTE E
REGRESSÃO LOGÍSTICA**

Dissertação apresentada à Escola de Economia
de São Paulo da Fundação Getulio Vargas,
como requisito para obtenção do título de
Mestre em Economia.

Campo de Conhecimento: Finanças

Orientador: Prof. Dr. Francisco Henrique
Figueiredo de Castro Junior

SÃO PAULO

2020

Costa, Roberto Zancaner.

Predição de recuperação judicial de empresas do setor sucroalcooleiro no Brasil utilizando análise discriminante e regressão logística / Roberto Zancaner Costa. - 2020.

82 f.

Orientador: Francisco Henrique Figueiredo de Castro Junior.

Dissertação (mestrado profissional MPFE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Análise de regressão logística. 2. Função discriminante linear. 3. Sociedades comerciais - Recuperação. 4. Empresas - Finanças. I. Castro Junior, Francisco Henrique Figueiredo de. II. Dissertação (mestrado profissional MPFE) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 336.6

ROBERTO ZANCANER COSTA

**PREDIÇÃO DE RECUPERAÇÃO JUDICIAL DE EMPRESAS DO SETOR
SUCROALCOOLEIRO NO BRASIL UTILIZANDO ANÁLISE DISCRIMINANTE E
REGRESSÃO LOGÍSTICA**

Dissertação apresentada à Escola de Economia
de São Paulo da Fundação Getúlio Vargas,
como requisito para obtenção do título de
Mestre em Economia.

Campo de Conhecimento: Finanças

Banca examinadora:

Orientador: Prof. Dr. Francisco Henrique
Figueiredo de Castro Junior

(Orientador)

Escola de Economia de São Paulo – EESP/FGV

Prof. Dr. Ricardo Ratner Rochman
Escola de Economia de São Paulo – EESP/FGV

Prof. Dr. Rodrigo Takashi Okimura
INSPER

AGRADECIMENTOS

À minha esposa Ana Paula, por compreender a minha ausência em função da dedicação ao curso de mestrado. Pelo incentivo a seguir adiante nos momentos mais difíceis e por me tranquilizar nos períodos de maior ansiedade. Seus exemplos de generosidade e companheirismo me fortaleceram ao longo desta jornada.

Ao meu filho Antônio, que, apesar de ter apenas 3 anos, se comportou muito bem na ausência do papai, proporcionando tranquilidade para o desenvolvimento deste trabalho.

À minha mãe, que sempre esteve ao meu lado nos períodos mais adversos, e por todos os seus ensinamentos, guiando-me sempre na direção da honestidade, honradez e perseverança.

À minha irmã, pelo exemplo de dedicação em todas as atividades que realiza em sua profissão.

Ao meu pai, que infelizmente não está mais presente, porém sempre me orgulhou por sua trajetória de vida.

RESUMO

O principal objetivo desta pesquisa foi desenvolver e comparar modelos estatísticos utilizando as técnicas de Regressão Logística e Análise Discriminante Linear, para compreender quais delas apresenta os melhores resultados na predição de recuperação judicial das empresas do setor sucroalcooleiro brasileiro. Inicialmente foram separados dois grupos de empresas para a coleta de dados, as empresas solventes e as insolventes, sendo este último grupo determinado pelas empresas que realizaram pedido formal de recuperação judicial. Para esses grupos foram levantadas as demonstrações financeiras para um, dois e três anos antes do evento do pedido de recuperação judicial. De posse das demonstrações financeiras, foram calculadas 20 variáveis independentes que, posteriormente, pela técnica de análise de componentes principais, foram reduzidas a 10, tendo sido eliminados principalmente os efeitos de multicolinearidade. A seleção final das variáveis explicativas significativas, tanto dos modelos de Regressão Logística quanto de Análise Discriminante Linear foi realizada pelo método *stepwise*. Os resultados obtidos demonstraram que a acurácia na previsão das empresas insolventes foi superior nos modelos de Regressão Logística em relação à Análise Discriminante Linear, portanto os modelos logit podem ser aplicados para a previsão de falência das empresas do setor com até três anos de antecedência com taxas de acerto superiores a 80%.

Palavras-chave: Regressão Logística, Análise Discriminante Linear, Recuperação Judicial, Sucroalcooleiro.

ABSTRACT

The main purpose of this research was to develop and compare statistical models using the Logistic Regression and Linear Discriminant Analysis techniques, to identify which of these techniques is best suited to predicting the insolvency of Brazilian companies of the sugar and alcohol sector. Initially, two groups of companies were segregated for the data collection: solvent and insolvent, provided that the latter was composed by the companies which had filed a formal request for judicial recovery. The data obtained were the financial statements relating to the period of one, two and three years prior to the formal request for judicial recovery. The calculation of 20 independent variables has been made in possession of these financial statements and, subsequently, using the analysis of main components, the variables have been reduced to 10, eliminating the multicollinearity effects. The final selection of the significant explanatory variables, both in the Logistic Regression model and in the Linear Discriminant Analysis, was made using the *stepwise* method. The obtained results demonstrated that the accuracy in predicting the insolvent companies was higher in the models of Logistic Regression in comparison to the Linear Discriminant Analysis, therefore, the logit models can be used to predict the bankruptcy of companies in the sector up to 3 years in advance with a success rate higher than 80% .

Keywords: Logistic Regression, Linear Discriminant Analysis, Bankruptcy, Sugar and Alcohol Sector.

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Valores da média das variáveis e teste de significância | 33 |
| Tabela 2 – Amostra $t - 1$ de empresas em RJ | 40 |
| Tabela 3 – Teste para diferença de médias | 60 |
| Tabela 4 – Correlação amostra $t - 1$ | 61 |
| Tabela 5 – Componentes principais de liquidez da amostra $t - 1$ | 62 |
| Tabela 6 – Coeficientes de liquidez da amostra $t - 1$ | 62 |
| Tabela 7 – Componentes principais de rentabilidade da amostra $t - 1$ | 63 |
| Tabela 8 – Coeficientes de rentabilidade da amostra $t - 1$ | 63 |
| Tabela 9 – Componentes principais de dívida de curto prazo da amostra $t - 1$ | 63 |
| Tabela 10 – Coeficientes de dívida de curto prazo da amostra $t - 1$ | 64 |
| Tabela 11 – Componentes principais de dívida total da amostra $t - 1$ | 64 |
| Tabela 12 – Coeficientes de dívida total da amostra $t - 1$ | 65 |
| Tabela 13 – Correlação da amostra $t - 2$ | 65 |
| Tabela 14 – Correlação da amostra $t - 3$ | 65 |
| Tabela 15 – Variância explicada pelas componentes principais | 66 |
| Tabela 16 – Lambda de Wilks para a amostra $t - 1$ | 67 |
| Tabela 17 – Função discriminante para $t - 1$ | 67 |
| Tabela 18 – Resultados da classificação para a amostra $t - 1$ | 68 |
| Tabela 19 – Lambda de Wilks para a amostra $t - 2$ | 69 |
| Tabela 20 – Função discriminante para $t - 2$ | 69 |
| Tabela 21 – Resultados da classificação para a amostra $t - 2$ | 70 |
| Tabela 22 – Lambda de Wilks para a amostra $t - 3$ | 70 |
| Tabela 23 – Função discriminante para $t - 3$ | 71 |
| Tabela 24 – Resultados da classificação para a amostra $t - 3$ | 71 |
| Tabela 25 – Resumo dos resultados de classificação | 72 |
| Tabela 26 – Resultados do método <i>stepwise</i> para a amostra $t - 1$ | 73 |
| Tabela 27 – Resultados da classificação para a amostra $t - 1$ (<i>logit</i>) | 73 |
| Tabela 28 – Resultados do método <i>stepwise</i> para a amostra $t - 2$ | 74 |
| Tabela 29 – Resultados da classificação para a amostra $t - 2$ (<i>logit</i>) | 75 |
| Tabela 30 – Resultados do método <i>stepwise</i> para a amostra $t - 3$ | 75 |
| Tabela 31 – Resultados da classificação para a amostra $t - 3$ (<i>logit</i>) | 76 |
| Tabela 32 – Resumo dos resultados da classificação (<i>logit</i>) | 76 |

LISTA DE GRÁFICOS

| | |
|--|----|
| Gráfico 1 – Função Logística | 42 |
| Gráfico 2 – Sobreposição pequena das distribuições | 46 |
| Gráfico 3 – Sobreposição grande das distribuições | 46 |
| Gráfico 4 – Regra de classificação de elementos | 47 |
| Gráfico 5 – <i>Boxplot</i> das variáveis de liquidez | 56 |
| Gráfico 6 – <i>Boxplot</i> das variáveis de rentabilidade | 57 |
| Gráfico 7 – <i>Boxplot</i> das variáveis de endividamento de curto prazo | 58 |
| Gráfico 8 – <i>Boxplot</i> das variáveis de endividamento total | 59 |
| Gráfico 9 – <i>Boxplot</i> das variáveis de Prazos Médios | 60 |

LISTA DE SIGLAS

| | |
|--------------|------------------------------------|
| DFC | Demonstrativo de Fluxo de Caixa |
| EBIT | Earnings before interest and taxes |
| <i>LOGIT</i> | Logistic Regression |
| LDA | Linear Discriminant Analysis |
| MDA | Multiple Discriminant Analysis |
| PCA | Principal Component Analysis |
| PIB | Produto Interno Bruto |
| PL | Patrimônio Líquido |
| PRÓALCOOL | Programa Brasileiro de Álcool |
| ROE | Return on Equity |
| ROA | Return on Assets |
| TWh | Terawatt-hora |
| CPV | Custo do Produto Vendido |

SUMÁRIO

| | |
|---|-----------|
| 1 INTRODUÇÃO | 21 |
| 2 REVISÃO DA LITERATURA | 25 |
| 2.1 Modelos de predição de falência..... | 25 |
| 2.2 Modelo z-score de Altman (1968)..... | 25 |
| 2.3 Modelo ZETA™ Analysis (1977) | 28 |
| 2.4 Modelo de Altman, Baidya e Dias (1979) | 31 |
| 2.5 Modelo de Ohlson (1980) | 34 |
| 2.6 Modelo de Silva (1983) | 35 |
| 2.7 Softwares estatísticos utilizados..... | 37 |
| 3 METODOLOGIA | 39 |
| 3.1 Amostragem | 39 |
| 3.2 Técnicas de análise dos dados | 41 |
| 3.2.1 Regressão logística – <i>logit</i> | 41 |
| 3.2.2 Análise discriminante linear | 44 |
| 3.2.3 Análise de componentes principais | 49 |
| 3.3 Definição das variáveis | 50 |
| 3.3.1 Variáveis de liquidez | 50 |
| 3.3.2 Variáveis de rentabilidade | 51 |
| 3.3.3 Variáveis de endividamento de curto prazo | 52 |
| 3.3.4 Variáveis de endividamento total..... | 52 |
| 3.3.5 Variáveis de giro | 53 |
| 3.3.6 Demais variáveis | 54 |
| 4 RESULTADOS..... | 56 |
| 4.1 Seleção das variáveis | 56 |
| 4.2 Análise discriminante linear | 66 |
| 4.2.1 Modelo discriminante para amostra t – 1..... | 66 |
| 4.2.2 Modelo discriminante para a amostra t – 2 | 68 |
| 4.2.3 Modelo discriminante para a amostra t – 3 | 70 |
| 4.3 Regressão logística | 72 |
| 4.3.1 Regressão logística para a amostra t – 1 | 72 |

| | |
|--|-----------|
| 4.3.2 Regressão logística para a amostra $t - 2$ | 74 |
| 4.3.3 Regressão logística para a amostra $t - 3$ | 75 |
| 5 CONCLUSÃO | 78 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 80 |

1 INTRODUÇÃO

O Brasil atualmente é o maior produtor de cana-de-açúcar do mundo, com uma moagem anual de 620 milhões de toneladas em 2019, e o segundo maior produtor mundial de etanol e açúcar, com uma produção de 33 bilhões de litros e 29 milhões de toneladas, respectivamente. A produção de etanol atende integralmente o mercado interno, que consome 32,8 bilhões de litros por ano, enquanto no mercado de açúcar são exportados 62% da produção, que representam aproximadamente 45% das exportações mundiais da *commodity*.

O setor sucroalcooleiro é responsável por 10% do PIB agrícola, ou seja, cerca de 2% do PIB brasileiro, e 5% do saldo da balança comercial, sendo um importante gerador de divisas para o país, com um valor bruto de produção acima de R\$ 150 bilhões por ano. Ademais, o setor gera cerca de 2,4 milhões de empregos, sendo 800 mil formais diretos e mais 1,6 milhão de empregos indiretos, além de ter como fornecedores de cana-de-açúcar mais de 70 mil produtores rurais independentes.

O Brasil conta com um parque fabril de 402 unidades, sendo a região Centro-Sul, composta pelas regiões Sudeste, Sul e Centro-Oeste, responsável por mais de 90% da produção de cana-de-açúcar. O Estado de São Paulo é o maior produtor, com 167 unidades e uma moagem de 333 milhões de toneladas, seguido pelos Estados de Goiás, com 70 milhões de toneladas, Minas Gerais, com 63 milhões de toneladas, Mato Grosso do Sul, com 49 milhões de toneladas, e Paraná, com 35 milhões de toneladas.

Além dos produtos tradicionalmente conhecidos como açúcar, álcool hidratado e álcool anidro, as usinas também produzem leveduras, etanol celulósico, bioplásticos, biobutanol, e energia elétrica a partir da cogeração de bagaço da cana-de-açúcar. A cana-de-açúcar como fonte de energia é, além de limpa, renovável e complementar à hídrica, pois sua produção concentra-se no período entre abril e novembro, durante a safra canavieira, correspondente ao período menos chuvoso do ano. A energia elétrica produzida pelo setor corresponde a 17,4% da matriz energética nacional e 84% do total de geração por biomassa, com uma produção de 27 TWh por ano, proporcionando uma redução de emissão de CO₂ da ordem de 6,3 milhões de toneladas anuais.

A cana-de-açúcar é uma cultura extremamente eficaz na ajuda ao combate ao efeito estufa, desde o sequestro de CO₂ no próprio desenvolvimento biológico da planta e a redução de emissão de poluentes a partir do etanol (em contrapartida aos combustíveis fósseis) até a geração de energia elétrica por fonte limpa e renovável. Como principal combustível limpo para carros atualmente utilizado em escala global, o etanol ganhou relevante importância com a

tecnologia do carro *flex*, que nasceu em 2003 e atualmente representa 67% da frota de veículos do país, com mais de 30 milhões de carros, e permite a utilização tanto do etanol hidratado quanto da gasolina (que no Brasil possui um percentual obrigatório de 25% de mistura com o etanol anidro).

Desde o desenvolvimento da tecnologia *flex*, o Brasil já reduziu sua emissão de gases de efeito estufa em mais de 520 milhões de toneladas de CO₂ equivalentes, e, para efeito de comparação, essa quantidade representa o total de emissões anuais conjuntas de Argentina, com 209 milhões de toneladas de CO_{2eq}, Chile, com 87 milhões de toneladas de CO_{2eq}, Colômbia, com 85 milhões de toneladas de CO_{2eq} e Equador, com 40 milhões de toneladas de CO_{2eq}.

O setor passou por fases bastante distintas ao longo dos últimos vinte anos, desde a euforia com a nova tecnologia do carro *flex* em 2003 e o forte ritmo de investimentos ao longo dos anos de 2004 a 2008, até as grandes dificuldades enfrentadas de 2008 em diante em função da alta alavancagem financeira das usinas, governança corporativa muito incipiente e ainda os problemas com políticas de congelamentos de preços da gasolina.

O volume de investimentos foi tão expressivo nos anos de 2004 a 2008 que a moagem brasileira cresceu 50% neste período, passando 381 milhões de toneladas para 572 milhões de toneladas, num setor altamente intensivo em capital. As usinas brasileiras, que em sua maioria possuíam controle familiar, com baixos níveis de governança corporativa e uma gestão financeira pouco profissional, acabaram utilizando o farto volume de crédito disponível à época para financiar o crescimento proveniente da expectativa de demanda por etanol em função da tecnologia do carro *flex*.

Apesar da tecnologia do carro *flex* ter se tornado uma realidade, o nível de alavancagem das usinas se tornou insustentável e as empresas do setor começaram a enfrentar grandes dificuldades financeiras a partir de 2008. Com as margens achatadas pela política de congelamento de preços e as despesas financeiras extremamente altas, as usinas começaram a diminuir o nível de investimentos na renovação de seus canaviais, reduzindo significativamente a produtividade agrícola e, portanto, gerando uma elevação substancial dos custos. Diante dessa difícil situação, o setor enfrentou a pior crise desde a criação do Proálcool, e ao longo dos doze anos seguintes, ou seja, de 2008 a 2019, as empresas registraram pedidos de recuperação judicial para 89 unidades de produção. O período de levantamento amostral deste trabalho se concentrará, portanto, nesse período, em que a base de dados se mostra ampla e propícia para análise de empresas em situação de insolvência no setor sucroalcooleiro.

O objetivo deste trabalho é desenvolver modelos estatísticos com elevado nível de acurácia, acima de 70%, na predição de recuperação judicial para as empresas do setor

sucroalcooleiro, utilizando como premissas os principais indicadores financeiros das demonstrações financeiras dessas empresas e como ferramental a regressão logística e a análise discriminante linear.

A motivação para o trabalho é que os modelos a serem estudados possam contribuir para a predição da recuperação judicial das usinas com tempo hábil para que ainda seja possível a correção dos rumos a fim de se evitar graves problemas econômicos, principalmente nas economias locais, uma vez que as usinas em geral estão localizadas em pequenos municípios e são responsáveis por grande parte da atividade econômica da região. Para tanto, serão utilizados dados com um, dois e três anos antes do pedido de recuperação judicial das empresas para analisar como os modelos capturam esses primeiros sinais de dificuldade financeira. Ao evitar a falência de uma usina são evitadas grandes perdas de postos de trabalhos, fechamentos de estabelecimentos comerciais locais, inadimplência aos fornecedores de cana-de-açúcar e demais insumos, prestadores de serviços e afins, como os que foram observados nos últimos doze anos.

2 REVISÃO DA LITERATURA

2.1 Modelos de predição de falência

A motivação para os estudos de predição de falência está relacionada ao intuito de detectar antecipadamente sinais que indiquem uma possibilidade de falência num futuro próximo a fim de se evitar prejuízos a todos os agentes envolvidos no ambiente da empresa em questão. Apesar de não haver até o momento uma teoria abrangente e definitiva sobre a modelagem de predição de falência, diversos estudos empíricos realizados ao longo das últimas décadas, sempre com a adoção das técnicas estatísticas mais inovadoras na ocasião, foram desenvolvidos e testados com o objetivo de entender o comportamento dos principais indicadores das demonstrações financeiras das empresas na fase de pré-insolvência.

Na presente seção serão visitados: o modelo desenvolvido por Altman (1968), nomeado de modelo *Z-score*, que utilizou a análise discriminante multivariada para uma amostra de empresas do setor industrial; o modelo ZetaTM Analysis (1977) de Altman, Haldeman e Narayanan, que foi construído com o ferramental de análise discriminante linear e quadrática para empresas dos setores manufatureiros e varejista; o modelo de Altman, Baidya e Dias (1979), que abordou os setores têxtil, mobiliário, de papel e celulose, varejista, de plásticos e metalúrgico por meio da análise discriminante linear; o modelo de Ohlson (1980), que utilizou a regressão logística como ferramenta de predição de falência; e o modelo de Silva (1983), que adotou amostras subdivididas para o Estado de São Paulo e outros Estados entre empresas do setor industrial e comercial, com a abordagem por análise discriminante.

Dada a vasta literatura bem-sucedida precedente a este trabalho com a adoção das técnicas estatísticas de regressão logística e análise discriminante linear, foram selecionados estes dois métodos como ferramentas para o desenvolvimento dos modelos de predição de falência do setor sucroalcooleiro brasileiro.

2.2 Modelo *z-score* de Altman (1968)

Segundo Altman (1968), a academia parecia estar caminhando para a eliminação do uso da análise pura por *ratios* como técnica analítica na avaliação do desempenho de empresas. Os teóricos criticavam o uso das regras práticas como comparações de *financial ratios* entre as empresas, técnica amplamente usada pelos profissionais de mercado. Altman (1968) buscava estabelecer um vínculo entre a análise tradicional por meio de *ratios* e as técnicas estatísticas mais rigorosas que se tornaram populares entre os acadêmicos. Para tanto, construiu um modelo

de previsão de falência corporativa com um conjunto de índices financeiros utilizando a análise discriminante multivariada aplicada apenas para empresas do setor industrial.

Os estudos formais relacionados à predição de falências nos negócios começaram a ficar mais evidentes na década de 1930. Os estudos da época apontavam para diferenças significativas entre os índices financeiros de empresas que faliram e empresas que continuavam operando. Também foram elaborados estudos dos índices de grandes corporações que tiveram dificuldades em cumprir suas obrigações de endividamento, e estudos que compararam uma lista de índices para empresas em falência e uma amostra correspondente de empresas normais, sendo que as evidências observadas por cinco anos antes da falência foram citadas como conclusivas de que a análise por *ratios* poderia ser útil na previsão da insolvência, ou seja, os estudos mencionados implicavam os *ratios* como bons preditores de falência. Em geral, os índices que medem rentabilidade, liquidez e solvência prevaleceram como os indicadores mais significativos, sendo a ordem de sua importância não muito clara.

Em quase todos os estudos apresentados, a metodologia era essencialmente univariada, com enfoque nos sinais individuais das variáveis, com isso, a análise por índices apresentada dessa maneira era suscetível a falhas de interpretação. Uma extensão apropriada dos estudos citados anteriormente, portanto, seria combinar várias medidas em um modelo preditivo significativo por meio da análise discriminante múltipla (MDA). Na análise discriminante múltipla, o primeiro passo é estabelecer a classificação de grupos, que podem ser dois ou mais, sendo que o MDA efetua uma combinação linear das características que melhor discriminam os grupos. A técnica MDA tem a vantagem de considerar um perfil completo de características comuns às empresas relevantes, bem como a interação dessas propriedades. Portanto, a função discriminante transforma valores de variáveis individuais em uma única pontuação discriminante ou valor Z , e é descrita da seguinte forma:

$$Z = v_1X_1 + v_2X_2, \dots, v_nX_n \quad (1)$$

v_1, v_2, \dots, v_n , *coeficientes discriminantes*
 X_1, X_2, \dots, X_n , *variáveis independentes*

A amostra inicial do estudo foi composta por 66 empresas, com 33 empresas em cada um dos dois grupos. O grupo insolvente foi formado por empresas do setor industrial que entraram com o pedido de falência no período de 1946 a 1965 e possuíam ativos com valor médio de USD 6,4 milhões. Já o grupo de empresas solventes foi escolhido aleatoriamente entre

empresas com ativos entre USD 1 e USD 25 milhões. Os balanços utilizados foram relativos a um ano antes da falência.

Em função do grande número de variáveis consideradas significantes em estudos anteriores de falência, uma lista de 22 variáveis potencialmente úteis foi levantada, e dividida em cinco categorias: liquidez, lucratividade, alavancagem, solvência e índices da atividade. Dessa lista, cinco variáveis foram selecionadas de forma que conjuntamente pudessem fornecer o melhor modelo para a previsão de falência das empresas. Para tanto, foram utilizadas técnicas de análise de significância estatística, avaliando as contribuições relativas de cada variável independente e a avaliação das correlações entre as variáveis relevantes.

A função discriminante final continha as seguintes variáveis: Capital de giro sobre Total de ativos (X_1), Lucros acumulados sobre Total de ativos (X_2), Lucros antes de juros e impostos sobre Total de ativos (X_3), Valor do *equity* sobre Valor contábil da dívida total (X_4), Vendas sobre Total de ativos (X_5).

$$Z = 0,012X_1 + 0,014X_2 + 0,033X_3 + 0,006X_4 + 0,9999X_5 \quad (2)$$

Para testar a capacidade de discriminação individual das variáveis foi realizado o teste F, relacionando a diferença entre os valores médios das variáveis em cada grupo. As variáveis X_1 a X_4 apresentaram ser significantes, indicando diferenças extremamente importantes nessas variáveis entre os grupos, enquanto a variável X_5 não mostrou diferença significativa entre os grupos.

Após a construção do modelo foi realizado um teste aplicando-o às empresas contidas na base amostral. Com isso, foi realizada a classificação das empresas da amostra de acordo com o modelo discriminante e chegou-se ao número de classificações corretas para cada um dos grupos, solventes e insolventes. O percentual de empresas que fazem parte da base amostral de empresas falidas e foram classificadas como saudáveis pelo modelo caracterizam o erro do Tipo I, enquanto o percentual de empresas que fazem parte da base amostral de empresas saudáveis e foram classificadas erroneamente como insolventes representam o erro do Tipo II. A soma dos acertos dividida pelo número total de empresas classificadas produz a medida de sucesso do MDA na classificação das empresas, e esse percentual é análogo ao R^2 na análise de regressão, que mede o percentual da variação da variável dependente explicada pelas variáveis independentes.

O modelo classificou corretamente 95% da amostra total, sendo que das 33 empresas que pertenciam ao grupo de empresas falidas, 31 foram classificadas corretamente, ou seja, o

erro do Tipo I foi de 6%, enquanto que, das 33 empresas que pertenciam ao grupo de empresas solventes, 32 foram classificadas corretamente, ou seja o erro Tipo II foi de 3%.

O segundo teste realizado tinha como objeto avaliar a capacidade discriminatória do modelo com dados financeiros de 2 anos antes do pedido de falência, portanto, foi testada a capacidade de predição de falência com maior antecedência.

A redução na precisão da classificação do grupo é compreensível porque as indicações de falência são menos claras quanto mais longe as empresas se encontram do evento de insolvência. Todavia, o modelo desenvolvido ainda foi capaz de prever corretamente 72% das falências dois anos antes do evento, apresentando um erro do Tipo I de 28%, contra apenas 6% no modelo com dados do ano anterior à falência. A taxa de acerto para as empresas solventes foi de 94%, o erro do Tipo II foi de 6%, contra 3% no modelo com dados de um ano antes da falência.

Para testar o modelo rigorosamente para empresas solventes e insolventes, Altman (1968) utilizou uma nova amostra com 25 empresas falidas, usando a função discriminante encontrada para um ano antes da falência. Na nova amostra de empresas falidas, os resultados encontrados foram bastante satisfatórios, com uma taxa de acerto de 96% e um erro do Tipo I de apenas 4%.

O autor também procurou entender o comportamento do modelo discriminante encontrado em empresas que sofreram dificuldades temporárias de rentabilidade, mas não chegaram a falir, ou seja, a classificação de falência de uma empresa desse grupo é um exemplo de erro do Tipo II. Um teste da eficácia do modelo discriminante nesse caso seria aplicar a classificação numa amostra de empresas que tiveram prejuízo, mas que não faliram, e analisar os resultados de classificação do MDA. Para tanto, foi selecionada uma amostra de 66 empresas com prejuízo nos anos de 1958 e 1961, com 33 em cada ano. As empresas selecionadas foram classificadas pelo modelo discriminante para determinar seu potencial preditivo de falência e os resultados mostraram que o erro do Tipo II foi de 21%.

2.3 Modelo ZETATM Analysis (1977)

Os objetivos do estudo de Altman, Haldeman e Narayanan (1977) foram construir, analisar e testar um novo modelo de classificação de falências, que considerava desenvolvimentos recentes em relação à predição de falências, com a incorporação de refinamentos na utilização da técnica de estatística discriminante.

Os principais fatores que motivaram os autores a desenvolverem um novo modelo de classificação de falências foram: as mudanças no tamanho e no perfil financeiro das empresas que pediram falência no período do estudo; a questão da atualização temporal dos dados financeiros das empresas falidas; o fato de os modelos anteriores focarem principalmente no setor manufatureiro, ou ainda em setores específicos industriais; a atualização de recentes alterações nos padrões contábeis das demonstrações financeiras; e, ainda, testar e avaliar os recentes avanços e ainda controversos aspectos da análise discriminante.

No que tange à mudança do tamanho das empresas que pediram falência no período do estudo, um ponto importante foi o aumento substancial percebido no tamanho médio das empresas com problemas financeiros e consequente visibilidade e preocupação das instituições financeiras e agências reguladoras. O estudo utilizou uma amostra de empresas falidas em que o tamanho médio do ativo foi de aproximadamente USD 100 milhões, com piso de USD 20 milhões.

O modelo denominado ZETATM também procurou incorporar, na amostra de empresas falidas, empresas do setor varejista, além das empresas do setor manufatureiro tradicionalmente abordadas pelos modelos anteriores.

As amostras foram compostas por 53 empresas insolventes e 58 empresas solventes. Da amostra de 53 empresas que faliram, 29 empresas eram do setor de manufatura e 24 pertenciam ao setor varejista, ou seja, praticamente metade de cada setor, sendo que essa proporção também foi adotada para o grupo de empresas solventes. O trabalho se concentrou no período de 1969 a 1975, em que estavam presentes 50 das 53 empresas que pediram falência, sendo estas empresas de capital aberto e com ativos acima USD 20 milhões. Na amostra também estavam presentes cinco empresas que tiveram apoio substancial do governo, uma empresa cuja fusão foi forçada e uma empresa que os bancos credores assumiram o negócio.

Os autores reuniram diversos índices financeiros úteis na predição de falência utilizados em estudos anteriores e, além disso, incluíram novos índices considerados potencialmente válidos. As variáveis foram classificadas em seis grupos, a saber: rentabilidade (variáveis de 1 a 6), cobertura e medidas de alavancagem (variáveis de 8 a 14), liquidez (variáveis de 15 a 18), índices de capitalização (variáveis de 19 a 23), variabilidade de lucros (variáveis de 24 a 26) e medidas diversas (variáveis 7 e 27).

Os autores efetuaram os ajustes contábeis nas demonstrações financeiras para o modelo ZETATM, conforme relatado anteriormente. Entre esses ajustes estão a capitalização dos arrendamentos, sendo que o valor resultante dessa capitalização foi adicionado aos ativos e passivos das empresas com os respectivos custos financeiros devidos. Para esse cálculo, foram

levantados os cronogramas de pagamentos dos arrendamentos e utilizada uma taxa de juros média para novas emissões mais um prêmio de risco de 10%, sendo as despesas com juros inseridas no demonstrativo de resultados. As adequações em relação às reservas seguiram o seguinte critério: se fossem de natureza contingencial, eram incluídas no patrimônio líquido e o resultado ajustado pela variação líquida da reserva do exercício, se estivessem relacionadas à avaliação de ativos, eram compensadas com esses ativos. As participações minoritárias também foram compensadas com outros ativos e as subsidiárias não consolidadas foram agrupadas com as contas da controladora. Os valores de *goodwill* e os ativos intangíveis foram deduzidos dos ativos e do patrimônio líquido em função da dificuldade de atribuição de valor econômico. Por fim, os gastos acruados com pesquisa e desenvolvimento, juros capitalizados e outros encargos diferidos foram considerados como despesas na data corrente.

Para o modelo de predição de falência, os autores utilizaram a análise discriminante linear e quadrática, sendo que o teste H_1 , Box (1949) foi capaz de orientar qual estrutura seria mais apropriada em função das características de classificação da amostra. Basicamente, se as matrizes de variância-covariância dos grupos fossem estatisticamente idênticas, o formato linear que agrupa todas as observações seria mais apropriado, caso as matrizes não fossem idênticas, a estrutura quadrática forneceria o modelo mais eficiente.

Após um processo iterativo de redução do número de variáveis, foram selecionadas 7 variáveis independentes para o primeiro modelo, cujos testes de validação se mostraram mais confiáveis, ou seja, os resultados não apresentaram melhoras significativas ao se adicionar mais variáveis. As 7 variáveis selecionadas para o modelo ZETATM foram: retorno sobre ativos, ou ROA, (X_1); estabilidade dos lucros, calculada por uma medida normalizada do erro padrão de dez anos em X_1 (X_2); serviço da dívida, medido pelo índice de cobertura de juros, ou seja, lucro antes de juros e impostos sobre as despesas financeiras, incluindo o valor imputado ao passivo de arrendamento em medida logarítmica (X_3); rentabilidade acumulada, medida pelos lucros acumulados sobre o total de ativos (X_4); liquidez, medida pela razão entre o ativo circulante e o passivo circulante (X_5); capitalização, medida pelo patrimônio líquido sobre o capital total (X_6); e tamanho, calculado pelo total de ativos das empresas em medida logarítmica (X_7).

Os resultados de acurácia do modelo ZETATM obtidos a partir da amostra de dados de um ano antes da falência apresentam precisão do modelo linear de 96,2% para o grupo de empresas insolventes e de 89,7% para o grupo de empresas solventes. O modelo quadrático apresentou taxa de acerto de 94,3% na classificação das empresas insolventes e 91,4% na classificação das empresas solventes. Em ambos os modelos foi possível perceber que o grau de acurácia diminui quanto maior o período entre o evento da falência e a data das

demonstrações financeiras, entretanto, a precisão ainda se situa no patamar de 70% para cinco anos antes da falência no modelo linear, resultado superior ao encontrado no estudo de Altman (1968), onde a precisão caiu vertiginosamente depois de dois anos.

O modelo de Altman (1968) recebeu bastante exposição nos principais textos financeiros e tornou-se um padrão de comparação para estudos subsequentes de classificação de falências. Comparando o modelo ZETATM com o modelo de Altman (1968), é possível perceber que o modelo ZETATM é muito mais preciso na classificação de falências nos anos 2 a 5 anteriores à falência, e a acurácia do ano inicial é aproximadamente igual.

2.4 Modelo de Altman, Baidya e Dias (1979)

O trabalho intitulado *Previsão de problemas financeiros em empresas*, publicado por Altman, Baidya e Dias em 1979, tinha como objetivo estudar as recentes falências de empresas brasileiras por meio de um modelo quantitativo que fosse capaz de classificar e fazer predição de empresas com dificuldades financeiras.

O Brasil experimentara na década de 1969 a 1979 um período de forte crescimento econômico com altas taxas de inflação, porém, as empresas apresentavam alto e crescente grau de endividamento. Como consequência desse endividamento, as despesas financeiras cresciam com maior rapidez que o lucro das empresas. Os índices de cobertura de juros caíram de 6,18, em 1971, para 3,79 em 1975, reduzindo fortemente a margem de segurança das empresas, fazendo-se com que houvesse um aumento da preocupação dos principais agentes de mercado, principalmente as instituições financeiras. Somado a essa insegurança no meio corporativo, o governo praticava uma política monetária austera para controlar a inflação, acrescentando mais incerteza sobre a capacidade de pagamento das empresas.

O estudo mostra que o número de falências e concordatas requeridas nos Estados de São Paulo e do Rio de Janeiro subiram 28% e 55%, respectivamente, no ano de 1977 em relação a 1976. Além desse crescimento expressivo, também pode ser observado que o tamanho das empresas que enfrentavam dificuldades financeiras também crescia.

A amostra coletada pelos autores reuniu 23 empresas insolventes (PS), principalmente dos setores têxtil, mobiliário, papel e celulose, varejista, plásticos e metalúrgico. Também foi selecionada uma amostra de 35 empresas saudáveis, grupo de controle (NP), de tamanho e setores equivalentes às empresas em situação de insolvência. Na amostra de empresas insolventes foram levantados os balanços patrimoniais dos três anos anteriores à constatação da

dificuldade financeira de cada empresa, e na amostra de empresas solventes foram adotados os mesmos períodos para análise.

Para a construção do modelo de previsão de falência, os autores utilizaram o método de análise discriminante linear, cujo objetivo é classificar a qual grupo determinada empresa pertence, por exemplo, solventes ou insolventes. Essa classificação é realizada por meio de uma função discriminante Z que combina as variáveis independentes de forma a minimizar a sobreposição de índices de empresas de grupos diferentes.

Para a escolha das variáveis independentes, os autores utilizaram o modelo de previsão de insolvência desenvolvido por Altman em 1968, e apenas efetuaram modificações para adaptar as variáveis ao padrão das demonstrações financeiras utilizadas no Brasil. As variáveis selecionadas, portanto, foram as seguintes:

$$X_1 = \frac{\text{Ativo Corrente} - \text{Passivo Corrente}}{\text{Ativo Total}} \quad (3)$$

$$X_2 = \frac{\text{Lucros Retidos}}{\text{Ativo Total}} \quad (4)$$

$$X_3 = \frac{\text{Lucros antes de Juros e Impostos}}{\text{Ativo Total}} \quad (5)$$

$$X_4 = \frac{\text{Valor de Mercado do Equity}}{\text{Exigível Total}} \quad (6)$$

$$X_5 = \frac{\text{Vendas}}{\text{Ativo Total}} \quad (7)$$

As adaptações para os padrões brasileiros de contabilidade foram realizadas nas variáveis independentes X_2 e X_4 . No que tange à variável X_2 , nos Estados Unidos, os lucros retidos consideram os lucros acumulados menos os dividendos pagos, contudo, nos padrões da contabilidade nacional daquela época, tais lucros eram transferidos sob forma de bonificação em ações para a conta de capital. Para equalizar a variável X_2 do padrão contábil americano para o brasileiro, os autores calcularam os lucros retidos como o valor do capital menos os recursos efetivamente aportados pelos acionistas. Com relação à variável X_4 , muitas empresas de capital fechado não possuíam cotação de mercado para o valor do *equity*, portanto, os autores utilizaram como *proxy* o valor do patrimônio líquido.

A tabela a seguir demonstra o valor do teste F para as médias das variáveis de acordo com o grupo a qual pertencem, solventes (NP) ou insolventes (PS).

Tabela 1 – Valores da média das variáveis e teste de significância

| Variável | Média | | Teste F |
|----------------|------------|------------|---------|
| | Amostra PS | Amostra NP | |
| X ₁ | 0,12 | 0,23 | 4,8 |
| X ₂ | 0,01 | 0,24 | 54,9 |
| X ₃ | 0,05 | 0,16 | 24,8 |
| X ₄ | 0,35 | 1,14 | 20,8 |
| X ₅ | 0,88 | 1,12 | 7,3 |

Fonte: Altman, Baidya e Dias (1979)

Como é possível observar na tabela, as cinco variáveis foram significantes ao nível de 1%, ou seja, existe diferença estatística entre as médias das variáveis, fato que indica que a análise discriminante pode ser uma ferramenta útil.

Os resultados empíricos obtidos demonstraram que a variável X₁ não contribuía para o poder explicativo do modelo, portanto, o modelo Z₁ possui apenas as variáveis X₂, X₃, X₄ e X₅. No modelo Z₂, foi retirada a variável X₂, pois X₂ e X₄ apresentaram-se semelhantes, ou seja, o modelo Z₂ permaneceu apenas com X₁, X₃, X₄ e X₅.

$$Z_1 = -1,44 + 4,03 \cdot X_2 + 2,25 \cdot X_3 + 0,14 \cdot X_4 + 0,42 \cdot X_5 \quad (8)$$

$$Z_2 = -1,84 - 0,51 \cdot X_1 + 6,32 \cdot X_3 + 0,71 \cdot X_4 + 0,52 \cdot X_5 \quad (9)$$

Para os dois modelos, Z₁ e Z₂, o ponto de separação dos grupos solventes e insolventes é o valor zero, ou seja, empresas com valores de Z negativos são classificadas como insolventes pelo modelo e empresas com Z positivos são classificadas como solventes. O modelo Z₁ apresentou desempenho de classificação melhor que o modelo Z₂ para os dados do penúltimo e antepenúltimo anos antes do problema de insolvência, portanto, as análises das classificações foram voltadas para o modelo Z₁. Da amostra de 23 empresas que apresentaram problemas de solvência, o modelo Z₁ classificou 20 empresas de maneira correta, detectando valores negativos para Z, com isso o erro do Tipo I nesse caso foi de 13% (classificar uma empresa com problemas financeiros no grupo das empresas saudáveis). Em relação à classificação das

empresas solventes, da amostra de 35 empresas, o modelo classificou corretamente 31, apresentando uma taxa de acerto de 89% e um erro do Tipo II de 11% (classificar uma empresa saudável no grupo das empresas com problemas).

Como era de se esperar, quanto mais distante da data da confirmação da insolvência, menor a precisão do modelo, ou seja, o modelo Z_1 apresentou grau de precisão de 87% no ano anterior ao problema de insolvência, 84,2% dois anos antes e 77,8% três anos antes.

Os autores aplicaram posteriormente testes para a validação do modelo. O primeiro deles foi o teste de Lachenbruch (1967), no qual se retira uma empresa da amostra e se calcula novamente os coeficientes do modelo para as firmas remanescentes e classifica-se a firma retirada com o novo modelo encontrado. Esse processo foi realizado 58 vezes com a exclusão de cada uma das empresas da amostra, e nesse caso apresentou o mesmo resultado obtido pelo teste-padrão.

2.5 Modelo de Ohlson (1980)

Em seu artigo de 1980, James Ohlson apresenta os resultados empíricos do estudo de predição de falências corporativas com a utilização da metodologia de regressão logística. O autor optou pela metodologia *logit* para evitar alguns problemas bastante conhecidos associados à Análise Discriminante Multivariada (MDA), entre eles, a existência de certos requisitos estatísticos das distribuições das variáveis explicativas, como a necessidade de as matrizes de variância-covariância serem as mesmas para os dois grupos, solventes e insolventes. Ademais, os resultados da aplicação de um modelo MDA são apresentados em *scores* com pouca interpretação intuitiva, uma vez que é basicamente um dispositivo de classificação discriminatório. O uso do modelo *logit*, por outro lado, evita essencialmente todos os problemas discutidos em relação ao MDA, além de o modelo proporcionar como resultado a probabilidade de uma empresa falir dentro de um período de tempo predeterminado.

A amostra selecionada por Ohlson foi composta de empresas que pediram concordata no período de 1970 a 1976, que tivessem suas ações negociadas nos últimos três anos e que não fossem ligadas aos setores financeiros, de serviços ou transportes. Dadas essas características foram levantadas informações para 105 empresas insolventes e 2.058 empresas solventes.

O estudo identificou quatro fatores básicos estatisticamente significantes que afetam a probabilidade de falência das empresas no intervalo de um ano, a saber: (i) o tamanho da empresa; (ii) a estrutura de capital; (iii) medidas de *performance*; e (iv) medidas de liquidez

atual. Portanto, os modelos desenvolvidos foram compostos pelas seguintes variáveis explicativas:

- SIZE: $\log(\text{Ativo total}/\text{índice de nível de preços})$;
- TLTA: $\text{Passivo total}/\text{Ativo total}$;
- WCTA: $\text{Capital de Giro}/\text{Ativo total}$;
- CLCA: $\text{Passivo Circulante}/\text{Ativo Circulante}$;
- OENEG: 1 se o Passivo total for maior que o Ativo total, 0 em caso contrário;
- NITA: $\text{Lucro Líquido}/\text{Ativo total}$;
- FUTL: $\text{Resultado operacional}/\text{Passivo Total}$;
- INTWO: 1 se o Lucro líquido for negativo pelos dois últimos anos, 0 em caso contrário;
- CHIN: $(NI_t - NI_{t-1}) / (|NI_t| + |NI_{t-1}|)$, onde NI é o lucro líquido.

O modelo com a amostra de dados de um ano antes da falência apresentou taxa de acerto para as empresas insolventes de 96,1%, enquanto o modelo com dois anos reportou percentual de acerto ligeiramente abaixo, 95,6%. Por fim, o último modelo que previa a falência para um ou dois anos antes do pedido obteve acurácia de 92,8%.

Como o número de empresas da amostra de empresas insolventes difere das empresas solventes, não é possível simplesmente adotar o critério de corte de 50%. Para tanto, o autor utilizou como corte o valor que minimiza os erros de classificação, e, portanto, o ponto de corte do primeiro modelo deveria ser 3,8% e, para o segundo, 8%. Adotando esse critério, o modelo $t - 1$ classificou erroneamente 12,4% das empresas insolventes, enquanto o modelo $t - 2$, apenas 8,6%.

2.6 Modelo de Silva (1983)

O autor José Pereira da Silva, em seu artigo “Avaliação da saúde financeira das empresas”, publicado na *Revista de Administração de Empresas* em 1983, tinha como objetivo a utilização da análise discriminante para a construção de um modelo capaz de avaliar a saúde financeira das empresas. Para tanto, utilizou uma amostra de 419 empresas, das quais 160 eram insolventes e 259 solventes, e direcionou os testes levando em consideração o setor de atividade dessas empresas. Os dados amostrais contemplaram o último ano antes do evento da insolvência e o penúltimo ano.

A ferramenta estatística utilizada foi a análise discriminante, na qual se busca classificar uma determinada empresa num dos grupos existentes, neste caso o grupo das empresas solventes e das empresas insolventes. Para tanto, é necessário que o elemento a ser classificado pertença a um dos grupos existentes, de forma a permitir que sejam comparadas suas características com as características dos grupos. As características são observadas por meio de um conjunto de variáveis independentes que, de acordo com seus correspondentes pesos, formarão a função discriminante Z .

De posse dos dados amostrais das 419 empresas, o autor classificou como ótimas as 10 melhores empresas de cada setor de atividade, de acordo com o *ranking* da revista *Exame*, publicada em setembro de 1980, excluindo-se da base ainda as empresas que tiveram prejuízos. Não foram consideradas as empresas dos setores de prestação de serviços e agrícolas, pelo fato de o autor considerá-las com características peculiares. As empresas classificadas como insolventes foram as que tiveram processos de concordata ou falência requeridos ou deferidos. O autor também realizou uma terceira classificação nomeada de regular para as empresas que não estavam no *ranking* de melhores empresas da revista *Exame*, nem envolvidas em processos de concordata ou falência. Após a realização de testes, o autor preferiu reunir num grupo só as empresas ótimas e regulares, classificando-as como boas.

Ademais, com objetivo de obtenção de resultados mais precisos, as amostras foram subdivididas entre empresas do setor industrial e comercial, bem como empresas do Estado de São Paulo e de outros Estados. Os resultados obtidos com o modelo definido pela amostra das empresas industriais de São Paulo e aplicado às 58 empresas comerciais mostraram taxa de acerto para as empresas saudáveis de 97,5%, e apenas 55,56% para as empresas insolventes.

Também foi analisado o modelo desenvolvido com a amostra das empresas comerciais de São Paulo e testado nas empresas industriais, com taxa de acerto de 69,59% para as empresas solventes e 81,97% para as empresas insolventes. Por fim, foi realizado o teste II, com o modelo desenvolvido com a amostra de empresas industriais e aplicado em outro grupo de empresas também industriais, sendo que neste caso a previsão correta para empresas saudáveis foi de 90,91% e, para empresas insolventes, de 86,36%.

O autor também aplicou um teste com o modelo desenvolvido com amostras de empresas industriais de São Paulo em empresas industriais de outros Estados e o resultado foi uma taxa de acerto total de 80,49%, sendo 75% para as empresas boas e 82,26% para as empresas insolventes.

O modelo de seis variáveis independentes para as empresas industriais de São Paulo Z_{1i} foi construído com dados do último ano antes do evento de insolvência e o modelo Z_{2i} utilizou os dados do penúltimo ano antes do evento de insolvência. A seguir o modelo para $t - 1$.

$$Z_{1i} = 0,72 - 5,124 \cdot E_{23} + 11,016 \cdot L_{19} - 0,342 \cdot L_{21} - 0,048 \cdot L_{26} + 8,605 \cdot R_{13} - 0,004 \cdot R_{29} \quad (10)$$

$$E_{23} = \frac{\text{Duplicatas descontadas}}{\text{Duplicatas a receber}} \quad (11)$$

$$L_{19} = \frac{\text{Estoques}}{\text{Vendas} - \text{Lucro Bruto}} \quad (12)$$

$$L_{21} = \frac{\text{Fornecedores}}{\text{Vendas}} \quad (13)$$

$$L_{26} = \frac{\text{Estoque Médio}}{\text{Vendas} - \text{Lucro Bruto}} \cdot 360 \quad (14)$$

$$R_{13} = \frac{\text{Lucro Operacional} + \text{Despesa Financeira}}{\text{Ativo Total Médio} - \text{Investimento Médio}} \quad (15)$$

$$R_{29} = \frac{\text{Passivo Circulante} + \text{Exigível a Longo Prazo}}{\text{Lucro Líquido} + 0,1 \cdot \text{Imobilizado Médio} \pm \text{Resultado da Correção Monetária}} \quad (16)$$

Segundo o autor, os resultados dos testes efetuados confirmam a capacidade preditiva dos modelos quantitativos de análise discriminante e demonstram que a seleção das variáveis e a determinação de seus pesos por processo estatístico eliminam o entendimento subjetivo da análise.

2.7 Softwares estatísticos utilizados

Para a construção dos modelos de regressão logística e análise discriminante linear, bem como a realização dos testes de acurácia dos modelos foram utilizados os *softwares* Stata, R e SPSS (*Statistical Package for Social Sciences*). O *software* Excel foi utilizado para a preparação e a adequação da base de dados levantada a partir das demonstrações financeiras apuradas nos diários oficiais estaduais e no Serasa.

3 METODOLOGIA

O presente trabalho tem como objetivo utilizar duas ferramentas estatísticas para a construção dos modelos de predição de falências das empresas do setor sucroalcooleiro. A primeira ferramenta a ser visitada será a regressão logística, onde são avaliadas as demonstrações financeiras das empresas insolventes nos anos $t - 1$, $t - 2$ e $t - 3$, onde n é o ano do evento que corresponde ao pedido formal de recuperação judicial pela empresa, em contraste com as demonstrações financeiras das empresas solventes nos mesmos períodos. A segunda ferramenta a ser utilizada será a análise discriminante linear, levando-se em conta também os anos $t - 1$, $t - 2$ e $t - 3$. Feito isso, o propósito é avaliar qual dos dois métodos possui melhor acurácia em relação à classificação das empresas insolventes, bem como avaliar, para cada modelo, qual o efeito temporal da taxa de acerto à medida que nos distanciamos do ano em que a recuperação judicial foi requerida.

As empresas cujas demonstrações financeiras foram pesquisadas são empresas brasileiras, tanto de capital aberto quanto de capital fechado, sociedades anônimas e limitadas, todas pertencentes ao setor sucroalcooleiro, com seu devido registro no Ministério da Agricultura, Pecuária e Abastecimento – MAPA. Foram utilizadas na amostra de empresas insolventes apenas as empresas com pedido formal de recuperação judicial, portanto, não constam nesta amostra as empresas que se encontram falidas, nem empresas em situação de recuperação extrajudicial. A amostra de empresas solventes foi composta por empresas que não fossem falidas, nem com pedido formal de recuperação judicial, nem em situação formal de recuperação extrajudicial.

Os dados foram coletados no período de 2008 a 2019, conforme mencionado no início deste trabalho, uma vez que nesse período pós-crise financeira americana a oferta de crédito foi substancialmente reduzida e as empresas do setor começaram a enfrentar grandes dificuldades em função do alto grau de alavancagem, fato esse que culminou em grande número de pedidos formais de recuperação judicial, sendo, portanto, um período rico de informações para o objetivo pretendido neste trabalho.

3.1 Amostragem

Um dos grandes desafios deste trabalho está relacionado à coleta de dados, uma vez que a maioria das empresas do setor é de capital fechado, ou seja, não possuem suas demonstrações financeiras disponíveis na CVM por obrigação. Portanto, houve um grande esforço na parte de

levantamento de dados junto aos diários oficiais de cada Estado onde as empresas têm sede, além do levantamento adicional junto ao Serasa.

A coleta de dados junto ao Serasa foi muito importante no sentido de aumentar a base de empresas presentes nas amostras. Contudo, as informações disponíveis em sua base são apenas relativas ao Balanço Patrimonial e Demonstrativo de Resultados, não abrangendo, portanto, a Demonstração de Fluxo de Caixa pelo método indireto, fato esse que limitou a análise de mais variáveis valiosas, especificamente as relacionadas à combinação entre DFC e Balanço Patrimonial.

Foram levantadas as demonstrações financeiras de 58 empresas, sendo 29 empresas insolventes e 29 solventes, no ano imediatamente anterior ao pedido de recuperação judicial das empresas insolventes. A base de dados acima foi utilizada para a construção tanto do modelo *logit* quanto da análise discriminante com dados em $t - 1$.

Tabela 2 – Amostra $t - 1$ de empresas em RJ

| Grupo | UF | Situação | Status jurídico | Região | Ano RJ | Ano-1 |
|---------------------------------------|----|----------|----------------------|--------------|--------|-------|
| Santa Terezinha | PR | OPERANDO | Recuperação judicial | Sul | 2019 | 2018 |
| Usina Rio Parana S.A. | MS | OPERANDO | Recuperação judicial | Centro-Oeste | 2019 | 2018 |
| Grupo Clealco | SP | OPERANDO | Recuperação judicial | Sudeste | 2018 | 2017 |
| Destilaria Porto Alegre | AL | OPERANDO | Recuperação judicial | Nordeste | 2017 | 2016 |
| Olival Tenório - Ind. Porto Rico | AL | OPERANDO | Recuperação judicial | Nordeste | 2017 | 2016 |
| Copertrading | AL | OPERANDO | Recuperação judicial | Nordeste | 2017 | 2016 |
| Agro Industrial Vista Alegre | SP | OPERANDO | Recuperação judicial | Sudeste | 2017 | 2016 |
| Grupo Farias - Vale Verde Emp. | AC | OPERANDO | Recuperação judicial | Norte | 2016 | 2015 |
| Energética São Simão | GO | OPERANDO | Recuperação judicial | Centro-Oeste | 2016 | 2015 |
| Usina Goianésia | GO | OPERANDO | Recuperação judicial | Centro-Oeste | 2016 | 2015 |
| Usina Santa Elisa S.A. | SP | OPERANDO | Recuperação judicial | Sudeste | 2016 | 2015 |
| Unialco S.A. | MS | OPERANDO | Recuperação judicial | Centro-Oeste | 2015 | 2014 |
| Alcool Vale S.A | MS | OPERANDO | Recuperação judicial | Centro-Oeste | 2015 | 2014 |
| Tonon | MS | OPERANDO | Recuperação judicial | Centro-Oeste | 2015 | 2014 |
| Energética Santa Helena | MS | OPERANDO | Recuperação judicial | Centro-Oeste | 2015 | 2014 |
| Usivale | PE | OPERANDO | Recuperação judicial | Nordeste | 2015 | 2014 |
| Renuka | PR | OPERANDO | Recuperação judicial | Sul | 2015 | 2014 |
| Abengoa Bioenergia | SP | OPERANDO | Recuperação judicial | Sudeste | 2015 | 2014 |
| Centroálcool | GO | OPERANDO | Recuperação judicial | Centro-Oeste | 2014 | 2013 |
| Usina Carolo | SP | OPERANDO | Recuperação judicial | Sudeste | 2014 | 2013 |
| Dail / Clarion | PR | OPERANDO | Recuperação judicial | Sul | 2013 | 2012 |
| Cafeálcool | SP | PARADA | Recuperação judicial | Sudeste | 2013 | 2012 |
| Destilaria Nova Era Ltda | SP | PARADA | Recuperação judicial | Sudeste | 2012 | 2011 |
| Destilaria Americana | PR | OPERANDO | Recuperação judicial | Sul | 2011 | 2010 |
| Alvorada do Bebedouro | MG | PARADA | Recuperação judicial | Sudeste | 2010 | 2009 |
| Alcana - Infinity | MG | PARADA | Recuperação judicial | Sudeste | 2009 | 2008 |
| Central Energetica Paraíso - Infinity | MG | PARADA | Recuperação judicial | Sudeste | 2009 | 2008 |
| Usina Naviraí - Infinity | MG | PARADA | Recuperação judicial | Sudeste | 2009 | 2008 |
| Destilaria de Alcool Libra | MT | OPERANDO | Recuperação judicial | Centro-Oeste | 2009 | 2008 |

Fonte: Resultados da Pesquisa.

Com o objetivo de analisar o comportamento dos modelos em relação à sua eficácia com o distanciamento temporal em relação ao evento do pedido de recuperação judicial, também foram construídos modelos com amostra total de 64 empresas, sendo 32 empresas insolventes e 32 solventes, para $t - 2$; e para 60 empresas, sendo 30 empresas insolventes e 30 solventes, para $t - 3$.

3.2 Técnicas de análise dos dados

3.2.1 Regressão logística – *logit*

Existe uma gama de casos em que o modelo de regressão linear não se aplica, como nos casos em que a variável dependente é uma variável categórica, ou seja, variável que pode ser mensurada com apenas um número limitado de valores ou categorias. Os modelos de regressão para variáveis categóricas são muito úteis para avaliação de crédito de empresas e para as ciências sociais.

A regressão logística é uma ferramenta que permite estudar variáveis categóricas dicotômicas, ou seja, variáveis que apresentam apenas duas opções de eventos, sim ou não. Dessa forma, a regressão logística traz a possibilidade de estimação da probabilidade associada à ocorrência de determinado evento em função das variáveis explicativas que podem ser contínuas ou categóricas. Como os resultados esperados são a probabilidade de ocorrência de determinado evento, os valores estarão sempre no intervalo de zero a um.

No caso específico deste trabalho, a variável dependente Y utilizada é uma variável binária, que apresenta valor igual a um, caso a empresa tenha efetuado o pedido de recuperação judicial, ou zero, em caso contrário, ou seja, apresente-se solvente. Portanto, dada a característica binária da variável dependente, foi adotado o modelo de regressão logística, ou simplesmente *logit*, como um dos modelos adequados para a elaboração do modelo de predição de falência pretendido.

O modelo de resposta binária a ser tratado neste trabalho será o modelo abaixo, em que a função G assume valores estritamente entre zero e um.

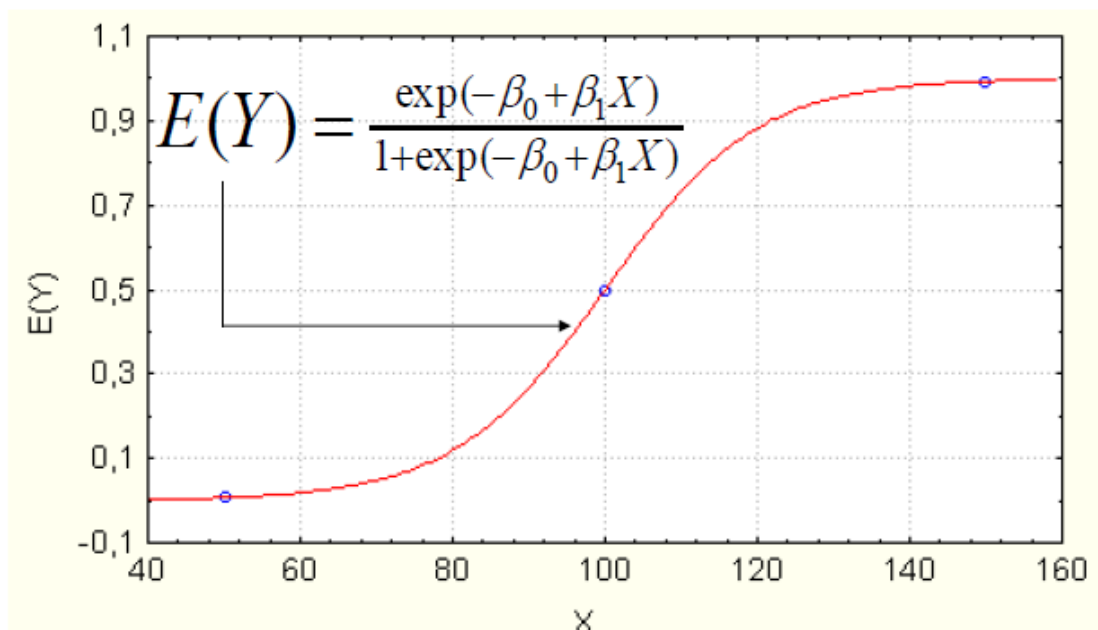
$$P(y = 1|x) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) = G(\beta_0 + \mathbf{x}\boldsymbol{\beta}) \quad (17)$$

No caso específico da regressão logística ou modelo *logit*, a função G será representada pela função de distribuição cumulativa de uma variável aleatória logística padrão, expressa por:

$$G(z) = \frac{e^z}{1+e^z} = \Lambda(z) \quad (18)$$

A função $G(z)$ descrita acima é uma função crescente com taxa máxima de crescimento com $z = 0$, e $G(z) \rightarrow 0$ quando $z \rightarrow -\infty$ e $G(z) \rightarrow 1$ quando $z \rightarrow +\infty$, fatos que podem ser observados no gráfico a seguir, que descreve o comportamento da função logística.

Gráfico 1 – Função Logística



Fonte: DCP UFMG (2012).

No modelo *logit*, as magnitudes de cada β_j não são de grande valia, diferentemente do que ocorre no modelo de probabilidade linear. O efeito parcial das variáveis aproximadamente contínuas será obtido pela derivada parcial da função G em relação a z .

$$\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + \mathbf{x}\boldsymbol{\beta}), \text{ onde } g(z) \equiv \frac{dG}{dz}(z) \quad (19)$$

Como G é uma função de distribuição acumulada de uma variável aleatória contínua, logo g é uma fdp, e como G é uma fdc estritamente crescente, $g(z)$ será positiva para todo z . Assim, conclui-se que o efeito parcial de x_j sobre $p(\mathbf{x})$ terá sempre o mesmo sinal de β_j .

A estimação dos parâmetros do modelo *logit* não pode ser realizada pelo método dos mínimos quadrados ordinários, dado que a função de regressão logística não é linear. Dessa forma, para a estimação dos parâmetros do modelo, faz-se necessária a utilização da estimação

por máxima verossimilhança, sendo que, no caso do modelo *logit*, a função de log-verossimilhança a ser maximizada é a seguinte:

$$f(y|\mathbf{x}_i; \boldsymbol{\beta}) = [G(\mathbf{x}_i\boldsymbol{\beta})]^y [1 - G(\mathbf{x}_i\boldsymbol{\beta})]^{1-y}, y = 0,1 \quad (20)$$

$$\ell_i(\boldsymbol{\beta}) = y_i \log [G(\mathbf{x}_i\boldsymbol{\beta})] + (1 - y_i) \log [1 - G(\mathbf{x}_i\boldsymbol{\beta})] \quad (21)$$

$$\text{máx } L(\boldsymbol{\beta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \log [G(\mathbf{x}_i\boldsymbol{\beta})] + (1 - y_i) \log [1 - G(\mathbf{x}_i\boldsymbol{\beta})] \quad (22)$$

Devido à natureza não linear do problema de maximização, não é possível obter equações para as estimativas de máxima verossimilhança do modelo *logit*, porém, as soluções podem ser rapidamente encontradas por meio computacional através de métodos iterativos de maximização de uma função.

No modelo *logit* pode-se utilizar o teste t de forma análoga à utilizada no método dos mínimos quadrados ordinários, uma vez que os *softwares* fornecem informações sobre os erros-padrão dos coeficientes. Sendo assim, o teste t deve ser realizado de maneira habitual com a estatística t dada pelo coeficiente sobre o erro-padrão para o teste da hipótese nula de $\beta = 0$.

Também é possível testar no modelo *logit* restrições múltiplas, sendo na maioria dos casos efetuados testes de múltiplas restrições de exclusão. Para tanto, podem ser aplicados os testes do multiplicador de Lagrange, no qual apenas se estima o modelo sob a hipótese nula; o teste de Wald, que exige a estimação somente do modelo irrestrito, cuja estatística possui uma distribuição qui-quadrada assintótica, com os graus de liberdade iguais ao número de restrições presentes no teste; e o teste da razão de verossimilhança baseado no mesmo conceito do teste F, cuja estatística é calculada pelo dobro da diferença nas log-verossimilhanças dos modelos restritos e irrestritos.

Conforme mencionado anteriormente, não é possível analisar o efeito parcial das variáveis por seus betas, entretanto, por meio dos betas é possível analisar o sinal dos efeitos parciais de cada variável, bem como avaliar a significância estatística de cada variável pelo teste da hipótese nula de beta igual a zero dado um determinado nível de significância aceitável. No modelo *logit*, diferentemente do modelo de probabilidade linear, o cálculo do indicador de qualidade de ajuste passa por um cálculo inicial que define se o valor encontrado pelo *logit* para cada observação pertence à classificação binária zero ou um. Portanto, para cada observação é necessária a aplicação da seguinte fórmula de classificação:

$$se \begin{cases} G(\beta^{\wedge}_0 + \mathbf{x}_i \beta^{\wedge}) < 0,5, y_i = 0 \\ G(\beta^{\wedge}_0 + \mathbf{x}_i \beta^{\wedge}) \geq 0,5, y_i = 1 \end{cases} \quad (23)$$

Com esses resultados torna-se possível avaliar a taxa de acerto do modelo *logit* ao confrontar os resultados obtidos pela fórmula acima com os valores efetivos para cada observação. Sendo assim, uma observação cujo valor efetivo é zero pode apresentar pelo modelo *logit* valores de zero ou um, demonstrando predição correta no primeiro caso e errônea no segundo. Esse raciocínio também vale quando o valor efetivo da observação é igual a um. Portanto, a percentagem corretamente prevista pelo modelo *logit* será o total de acertos do modelo dividido pelo total de observações.

Apesar de a percentagem corretamente prevista pelo modelo ser muito útil como indicador de qualidade de ajuste, é preciso avaliá-la com parcimônia, uma vez que é possível obter percentagens de acerto relativamente altas mesmo em situações em que a capacidade preditiva do modelo seja extremamente baixa. Para elucidar essa questão, basta analisar uma situação em que uma amostra com 100 observações apresente valores efetivos de 80 observações com valor igual a zero e 20 observações com valor igual a um. Nesse caso, o modelo *logit* que apresente um resultado de 70 observações iguais a zero dentre as 80 aparentemente possui uma taxa de acerto de 87,5%. Não obstante, esse mesmo modelo pode ter apresentado nenhuma predição correta em relação às observações iguais a um. Portanto, uma análise mais abrangente em relação aos resultados obtidos pelo modelo *logit* seria realizar também a análise de percentagem corretamente prevista para cada um dos resultados.

3.2.2 Análise discriminante linear

A técnica de análise discriminante linear tem como objetivo comparar as diferenças entre os grupos de interesse e classificar um novo elemento como pertencente ao grupo com perfil mais semelhante. Para tanto, o modelo utiliza uma variável categórica como variável dependente e múltiplas variáveis métricas como independentes.

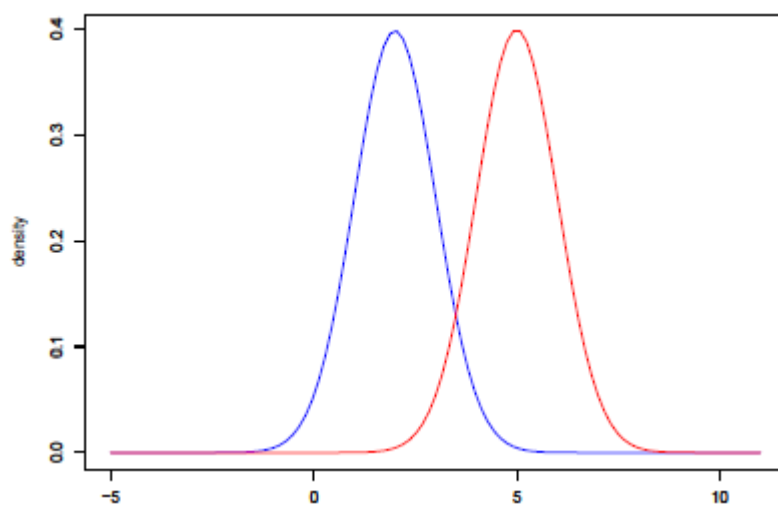
Com o conhecimento prévio das características dos grupos, a análise discriminante é a ferramenta adequada para testar a hipótese de que as médias de um conjunto de variáveis independentes para dois ou mais grupos é igual, e com isso checar se as variáveis independentes são capazes de discriminar os grupos, além de avaliar a importância relativa de cada uma delas para o modelo.

A função discriminante resulta de uma variável que é a combinação linear de duas ou mais variáveis independentes com maior capacidade de discriminação entre os grupos definidos inicialmente. Essa função é definida com a determinação dos pesos para as variáveis explicativas de forma a maximizar a variância entre os grupos em relação à variância dentro dos grupos. A multiplicação dos pesos pelas variáveis independentes trará um *score* Z de discriminação para cada elemento do grupo, sendo que a média desses *scores* para um determinado grupo é chamado de centroide. O centroide sinaliza a localização mais provável de um elemento num determinado grupo, e, ao se comparar a distância entre os centroides, é possível entender quão próximos ou afastados estão os grupos em relação ao discriminante Z.

O teste de significância estatística da função discriminante é determinado pela distância entre os centroides dos grupos. Esse cálculo é feito por meio da comparação entre a distribuição dos *scores* discriminantes para cada um dos grupos. Caso a sobreposição das distribuições seja pequena, a função discriminante possui boa capacidade de discriminação, e, caso a sobreposição seja grande, a função discriminante não é uma boa alternativa a ser usada para a separação dos grupos.

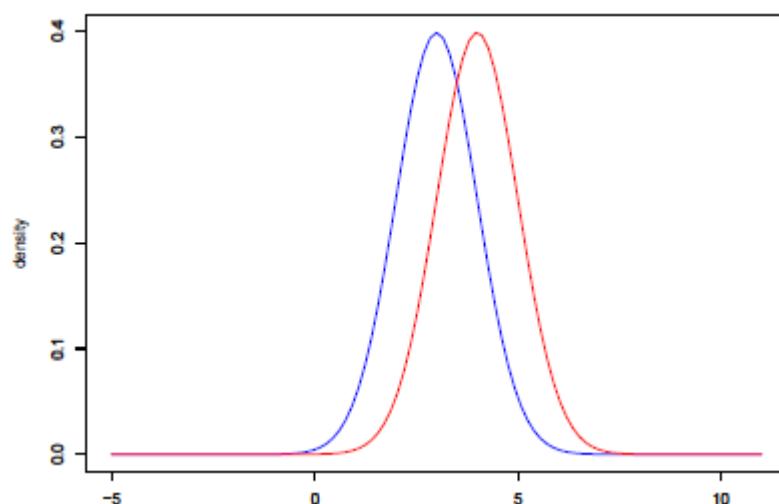
Para melhor entendimento, a figura a seguir demonstra no primeiro gráfico a situação em que a sobreposição das distribuições é pequena e, portanto, a função discriminante é eficiente, e o segundo gráfico apresenta a situação em que a sobreposição é grande, e a função discriminante não é capaz de gerar uma discriminação satisfatória.

Gráfico 2 – Sobreposição pequena das distribuições



Fonte: ICMC USP (2019).

Gráfico 3 – Sobreposição grande das distribuições



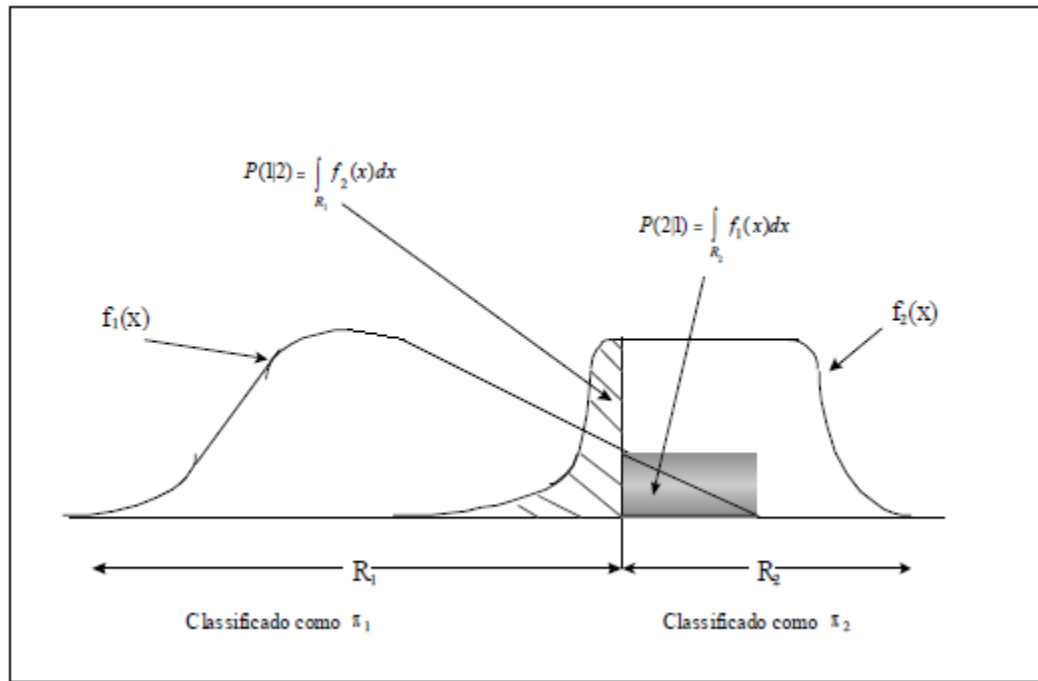
Fonte: ICMC USP (2019).

Para a construção matemática da explanação teórica acima acerca da análise discriminante, são selecionados dois grupos, π_1 e π_2 , cujos elementos serão separados ordinariamente ou classificados com base em p variáveis aleatórias, em que os valores observados de X diferem em alguma quantidade de um grupo para o outro.

$$\mathbf{X}' = [X_1, X_2, X_3, \dots, X_p] \quad (24)$$

Assumindo que as populações desses dois grupos podem ser descritas por suas funções de densidade $f_1(\mathbf{x})$ e $f_2(\mathbf{x})$, e que o conjunto de todos os resultados amostrais possíveis são divididos em duas regiões, R_1 e R_2 , se um novo elemento pertencer à região R_1 , este será classificado como pertencente ao grupo π_1 , e caso o novo elemento pertença à região R_2 , este será classificado como pertencente ao grupo π_2 .

Gráfico 4 – Regra de classificação de elementos



Fonte: ICMC USP (2019).

Como pode ser observado na figura, o critério para uma boa classificação do novo elemento em determinado grupo seria criar uma regra que minimize a chance de classificar incorretamente um elemento da amostra, para tanto será adotada a técnica de máxima verossimilhança.

No caso em que \mathbf{X} possui uma distribuição normal, a população 1 tem média μ_1 e a população 2 tem média μ_2 e ambas as populações tenham o mesmo desvio-padrão, a razão de verossimilhança entre as duas populações pode ser descrita pela equação a seguir:

$$\lambda(x) = \frac{f(x_1)}{f(x_2)} = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2}} = e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_2}{\sigma}\right)^2\right]} \quad (25)$$

Portanto, nesse caso a regra de classificação se daria de acordo com a seguinte formulação:

$$se \begin{cases} \lambda(x) > 1, o \text{ elemento} \in \pi_1 \\ \lambda(x) < 1, o \text{ elemento} \in \pi_2 \\ \lambda(x) = 1, classif. indefinida \end{cases} \quad (26)$$

Considerando ainda a razão entre as funções de densidade, é possível chegar na função discriminante por meio de simples manipulação algébrica com a aplicação do logaritmo neperiano em ambos os lados da equação de $\lambda(x)$, e, por fim, entender que a função discriminante está diretamente relacionada com a diferença das distâncias euclidianas ponderadas ao quadrado.

$$\begin{aligned} \lambda(x) &= e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_2}{\sigma}\right)^2\right]} \\ -2 \ln(\lambda(x)) &= \left[\left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_2}{\sigma}\right)^2\right] \\ -2 \ln(\lambda(x)) &= \frac{1}{\sigma^2}[(x-\mu_1)^2 - (x-\mu_2)^2] \end{aligned} \quad (27)$$

Portanto, nesse caso, a regra de classificação se daria de acordo a seguinte formulação:

$$se \begin{cases} -2 \ln(\lambda(x)) < 0, o \text{ elemento} \in \pi_1 \\ -2 \ln(\lambda(x)) > 0, o \text{ elemento} \in \pi_2 \\ -2 \ln(\lambda(x)) = 0, classif. indefinida \end{cases} \quad (28)$$

No caso em que as variâncias das populações são diferentes, a função discriminante terá a forma descrita a seguir e o critério de classificação dos elementos segue a mesma regra já apresentada anteriormente.

$$-2 \ln(\lambda(x)) = -2 \ln\left(\frac{\sigma_2}{\sigma_1}\right) + \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] \quad (29)$$

Quando o modelo de análise discriminante for multivariado e as matrizes de covariâncias forem iguais, tem-se o caso da função discriminante de Fisher, cuja forma e os critérios de classificação podem ser observados a seguir:

$$f_d(\mathbf{x}) = \ln(\lambda(\mathbf{x})) = \left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right]$$

$$se \begin{cases} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{x} > \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \mathbf{x} \in \pi_1 \\ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{x} > \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \mathbf{x} \in \pi_2 \end{cases} \quad (30)$$

3.2.3 Análise de componentes principais

A análise de componentes principais é uma técnica multivariada que tem como objetivo a análise de dados buscando sua redução, eliminação e sobreposição por meio da combinação linear das variáveis originais. Essa técnica multivariada de modelagem da estrutura de covariância também é conhecida como Transformada Discreta de Karhunen-Loève, ou também Transformada de Hotelling. O conceito de redução da massa de dados, com a menor perda possível das informações, é materializado pela transformação de um conjunto de variáveis iniciais num outro conjunto de variáveis com a mesma dimensão.

A técnica de análise de componentes apresenta como principal vantagem a eliminação da multicolinearidade das variáveis, dado que transforma o conjunto de variáveis iniciais intercorrelacionadas em um novo conjunto de variáveis descorrelacionadas. As desvantagens da técnica são a sensibilidade a *outliers*, a não recomendação de utilização de amostra com dados ausentes e quando o número de variáveis é maior que unidades amostrais.

Considerando as (p) variáveis $X_1, X_2, X_3, \dots, X_p$, cada uma com n elementos, formando a matriz original \mathbf{X} ($n \times p$), para obtenção dos componentes principais são definidos os vetores das médias, $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3, \dots, \mu_p]$ e a matriz de covariância $\boldsymbol{\Sigma}$. Com isso, se encontram os pares de autovalores e autovetores $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), (\lambda_3, \mathbf{e}_3), \dots, (\lambda_p, \mathbf{e}_p)$, com $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p$, e, portanto, a variável latente $\mathbf{Z}_i = \mathbf{e}_i' \mathbf{X}$, projetando os pontos coordenados originais em um plano, maximizando a distância euclidiana entre eles, ou seja, maximizando a variabilidade de \mathbf{Z}_i .

A contribuição de cada componente principal \mathbf{Z}_i é expressa em porcentagem, sendo que a explicação individual de cada uma das componentes é calculada pela proporção entre a variância individual de cada componente principal \mathbf{Z}_i pela variância total.

$$C = \frac{\text{Var}(\mathbf{Z}_i)}{\sum_{i=1}^p \text{Var}(\mathbf{Z}_i)} \cdot 100 \quad (31)$$

Como os componentes principais são obtidos por meio da matriz de transformação cujas colunas são os autovetores da matriz de covariância estimada dos dados, e a matriz de covariância é simétrica positiva semidefinida, a realização desse tipo de cálculo por diversos *softwares* de cálculos matriciais se torna aplicável.

Em geral define-se o componente principal como o de maior variância, ou seja, que explique o máximo da variabilidade das informações. Recomenda-se a adoção da componente principal com proporção de explicação acima de 70%.

3.3 Definição das variáveis

As variáveis selecionadas para o estudo foram levantadas de acordo com o conjunto de variáveis utilizadas nos trabalhos precedentes de predição de falência, tais como variáveis de liquidez, endividamento, estrutura de capital e rentabilidade, sempre adotando a utilização de *ratios* financeiros em vez dos dados financeiros puros, a fim de se evitar problemas inflacionários ao longo dos 12 anos de dados coletados. Infelizmente, todos os *ratios* que envolvem o patrimônio líquido não puderam ser utilizados, uma vez que várias empresas do setor possuem patrimônio líquido a descoberto, principalmente as empresas insolventes. Como exemplo para a inviabilização da utilização do PL, é possível analisar o efeito no ROE de uma empresa com patrimônio líquido a descoberto e prejuízo anual, neste caso, haveria um ROE positivo, fato que não faz sentido para a análise. Portanto, como será apresentado adiante, os indicadores que tradicionalmente utilizam o patrimônio líquido no denominador tiveram esse termo substituído pelo ativo total.

As variáveis selecionadas para a construção dos modelos estão listadas a seguir.

3.3.1 Variáveis de liquidez

As variáveis utilizadas para a avaliação de liquidez das empresas do setor sucroalcooleiro serão tanto melhores quanto maiores seus indicadores, portanto, espera-se que nos modelos finais o sinal de seus coeficientes seja negativo, ou seja, quanto maior a liquidez da empresa, menor sua chance de insolvência.

- **LiqCorr:** A liquidez corrente é a razão entre o ativo circulante e o passivo circulante, e mede a capacidade de curto prazo de a empresa saldar suas obrigações financeiras. O indicador de liquidez corrente acima de 1 demonstra que a empresa possui ativos a curto prazo suficientes para cobrir suas obrigações de curto prazo, lembrando que existem certos ativos que possuem menor liquidez que outros.

- **LiqSeca:** O índice de liquidez corrente também avalia a capacidade de curto prazo de a empresa pagar suas obrigações financeiras, contudo, é um indicador mais conservador, uma vez que não considera os estoques no seu cálculo, ou seja, o ativo menos líquido do ativo circulante é expurgado do cálculo para melhor avaliação da saúde financeira de curto prazo da empresa.
- **LiqImed:** A liquidez imediata considera apenas o caixa e equivalentes de caixa no ativo circulante, ou seja, apenas o ativo mais líquido de todos, e confronta essa posição com o passivo circulante.

3.3.2 Variáveis de rentabilidade

As variáveis selecionadas para a avaliação da rentabilidade das empresas tendem a apresentar sinais negativos para seus coeficientes, uma vez que quanto maiores as margens operacionais e o retorno do ativo, menor a probabilidade de insolvência da empresa.

- **MargBru:** A margem bruta é a razão entre o lucro bruto e a receita líquida e expressa basicamente a eficiência da empresa em relação à gestão de seus custos, uma vez que o lucro bruto é expresso pela receita líquida menos o custo dos produtos vendidos. Em outras palavras, o quanto a empresa é capaz de extrair de lucro bruto percentualmente em relação às suas vendas.
- **MargEBIT:** A margem EBIT é calculada pela razão entre o EBIT e a receita líquida e demonstra, em termos percentuais, qual o lucro operacional da empresa em relação às suas vendas, ou seja, é uma medida de eficiência operacional da empresa. A margem EBIT desperta o interesse tanto dos acionistas quanto dos credores, uma vez que será responsável pela remuneração de ambos.
- **MargLiq:** A margem líquida é a razão entre o lucro líquido e a receita líquida e traduz em termos percentuais quanto a empresa é capaz de extrair de lucros em suas operações em relação às suas vendas. A margem líquida é uma medida de interesse para os acionistas, dado que os credores já foram remunerados pelos juros.
- **ROA_oper:** O *Return on Assets* é medido pela razão entre o EBIT e o ativo total e tem por objetivo demonstrar a rentabilidade das operações. Esse é um indicador importante, pois além de ser de interesse tanto dos credores quanto dos acionistas, demonstra a eficiência com que foram alocados os capitais de ambos.

3.3.3 Variáveis de endividamento de curto prazo

Para as variáveis de endividamento de curto prazo, espera-se que os coeficientes presentes nas equações dos modelos de análise discriminante linear e regressão logística sejam positivos, dado que quanto maior o nível de endividamento de curto prazo, maiores as chances de falência das empresas.

- **EndivCP:** O endividamento de curto prazo é a razão entre as dívidas onerosas pertencentes ao passivo circulante em relação ao total de dívidas onerosas da empresa, e reflete basicamente a qualidade do perfilamento da dívida e, conseqüentemente, a pressão sobre o caixa no curto prazo.
- **EndivCP_Rec:** A relação entre as dívidas onerosas do passivo circulante e a receita líquida foi utilizada em substituição ao tradicional índice de dívida de curto prazo sobre EBIT, uma vez que esse indicador perde o sentido quando temos empresas que possuem EBIT negativo. Esse indicador foi elaborado com o intuito de compreender a dimensão do endividamento de curto prazo em relação ao porte da empresa, e acredita-se ser útil para discriminar empresas solventes das insolventes.
- **EndivLiqCP_Rec:** A razão entre dívida líquida de curto prazo e receita líquida tem o mesmo objetivo do indicador de endividamento de curto prazo sobre receita líquida, porém, nesse caso, o caixa e os equivalentes de caixa são descontados da dívida de curto prazo, a fim de diferenciar as empresas pelo seu nível de liquidez imediata.

3.3.4 Variáveis de endividamento total

Analogamente às variáveis de endividamento de curto prazo, as variáveis de endividamento total também devem apresentar sinal positivo para seus coeficientes nas regressões, uma vez que, quanto mais endividada está a empresa, mais provável será a insolvência.

- **EndivTot:** O endividamento total foi calculado pela razão entre a dívida onerosa total, de curto prazo e longo prazo, e o ativo total. Esse indicador demonstra quão alavancada a empresa está em relação à sua base de ativos que irá gerar o caixa necessário para quitá-la. Quanto maior o nível de endividamento total da empresa, maiores suas chances de problemas de inadimplência.
- **EndivLiq:** O endividamento total líquido tem o mesmo objetivo do indicador de endividamento total, no entanto, neste caso o caixa e os equivalentes de caixa são

descontados da dívida total, discriminando empresas que possuem boa liquidez das demais.

- **EndivTot_Rec:** A relação entre o endividamento total e a receita líquida também foi utilizada em substituição ao tradicional índice de dívida total sobre EBIT, pelos mesmos motivos de EBIT negativo citados anteriormente. Esse indicador mostra quão alavancada está a empresa em relação ao seu porte, ou ainda quantos anos de receita líquida são necessários para quitar o endividamento total.
- **EndivLiq_Rec:** A razão entre dívida líquida total e receita líquida tem o mesmo objetivo do indicador de endividamento total sobre receita líquida, contudo, neste caso, o caixa e os equivalentes de caixa são descontados da dívida total. Esse indicador premia as empresas com boa liquidez imediata e penaliza as empresas em baixo nível de caixa.

3.3.5 Variáveis de giro

- **GiroAtivo:** O giro do ativo é uma medida de capacidade de extração de receita com base nos ativos da empresa. Esse indicador é medido pela razão entre a receita líquida e os ativos totais, e quanto maior, melhor a utilização da base de ativos pela empresa. Dado o exposto acima, espera-se que o coeficiente da variável giro do ativo nas equações de regressão seja negativo, pois quanto mais eficiente for a empresa na utilização de sua base de ativos, menor serão suas chances de falência.
- **PMRV:** O prazo médio de recebimento das vendas reflete em média quantos dias a empresa concede de prazo para seus clientes. Empresas com prazos altos de recebimento de suas vendas são penalizadas com alta demanda por capital de giro. Este é um caso especial para se analisar quando se trata de empresas próximas da falência, pois se espera que empresas com má gestão de capital de giro tenham altos prazos de recebimentos. Todavia, pode acontecer de as empresas em dificuldades efetuarem descontos de recebíveis, para aliviar a pressão de caixa, baixando o nível de suas duplicatas. Portanto, neste caso não há evidências de qual sinal se esperar nas equações de regressão.
- **PMPF:** O prazo médio de pagamento de fornecedores reflete o prazo médio, em dias de CPV, que a empresa consegue com seus fornecedores para o pagamento de suas obrigações operacionais. Em tese, quanto maior esse prazo, melhor para a empresa, pois reduz sua necessidade de capital de giro. No entanto, no caso específico de empresas em dificuldade financeira, é comum encontrar elevados prazos de pagamento de

fornecedores em função de atrasos nas obrigações. Portanto, neste caso também não há evidências de qual sinal se esperar nas equações de regressão.

- **PMRE:** O prazo médio de renovação dos estoques é uma medida que demonstra o número de dias que a empresa leva em média para girar seu estoque e é apresentada em dias de CPV. Quanto maior o prazo para girar o estoque, maior será a demanda por capital de giro da empresa e menos eficiente será a gestão do seu capital de giro. Todavia, empresas em pré-insolvência tendem a carregar o mínimo de estoque possível, podendo apresentar baixos níveis de PMRE mesmo não sendo eficientes na gestão do capital de giro. Da mesma forma, também não é possível concluir com antecedência qual sinal dos coeficientes esperar.

3.3.6 Demais variáveis

- **PLneg:** A variável *dummy* de patrimônio líquido negativo tem valor igual a um se a empresa tiver patrimônio líquido a descoberto e zero em caso contrário. Essa é uma variável importante, uma vez que empresas em fase de pré-insolvência podem apresentar esse tipo de comportamento. Portanto, espera-se o sinal positivo para essa variável nas regressões.
- **PL_Rec:** A variável de patrimônio líquido sobre receita líquida tem por objetivo entender o montante de capital dos acionistas em relação ao porte da empresa. Essa variável não é tradicionalmente utilizada na literatura, porém pode ser de grande valia à medida que empresas com pouco capital próprio em relação ao seu tamanho tendem a estar mais expostas a riscos de insolvência. Portanto, conclui-se que essa variável tende a apresentar valores negativos para seus coeficientes.

4 RESULTADOS

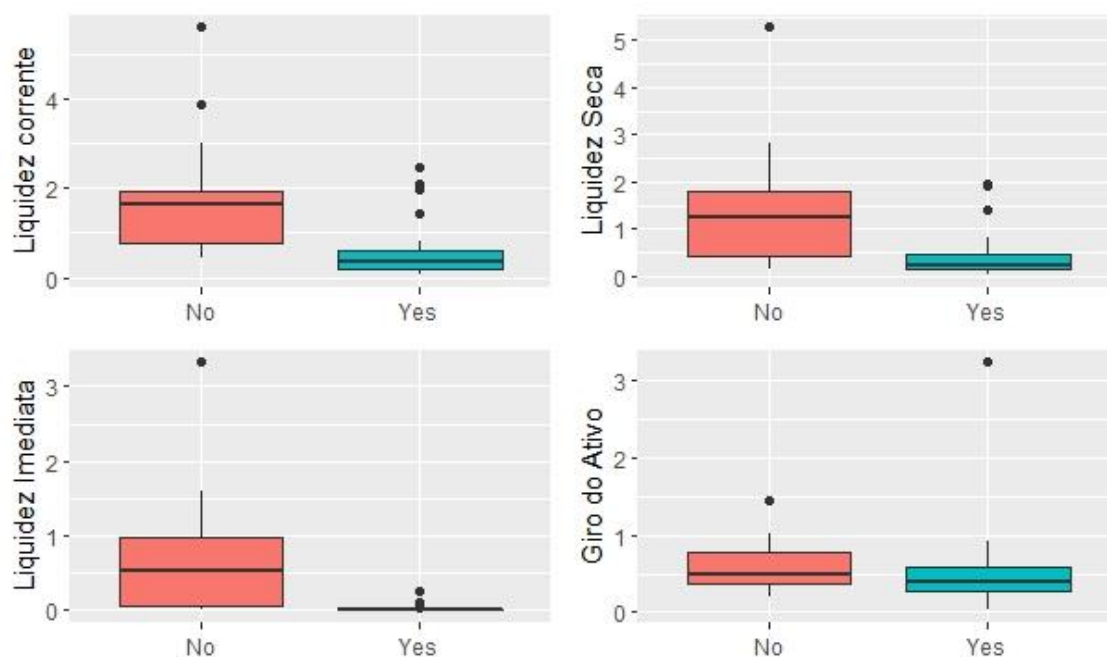
4.1 Seleção das variáveis

No caso do modelo de análise discriminante linear, quanto mais distantes os centroides das amostras entre as empresas solventes e insolventes, maior será o poder discriminante da variável independente para o modelo. Para tanto, com o intuito de compreender melhor o comportamento de cada uma das variáveis independentes pré-selecionadas, foram elaborados gráficos *boxplot* que mostram com clareza visual quais variáveis terão boa capacidade de discriminação entre os grupos.

Os gráficos apresentados a seguir refletem os dados coletados das amostras para o ano imediatamente anterior ao pedido de recuperação judicial das empresas insolventes, e serão nomeados daqui em diante por $t - 1$.

No caso das variáveis de liquidez corrente, seca e imediata, todas possuem diferenças importantes com relação às médias dos grupos de empresas solventes (*No*) e insolventes (*Yes*), sendo que os grupos das empresas solventes possuem maior liquidez com mais variância que as empresas insolventes. Pode-se observar que a variável de liquidez imediata para as empresas insolventes possui mediana próxima a zero, o que demonstra que as empresas em recuperação judicial administram suas operações sob forte pressão de caixa. A variável giro do ativo, que expressa o quão eficiente é a empresa em extrair receita da sua base de ativos, não apresentou visualmente diferenças importantes para a segregação dos grupos.

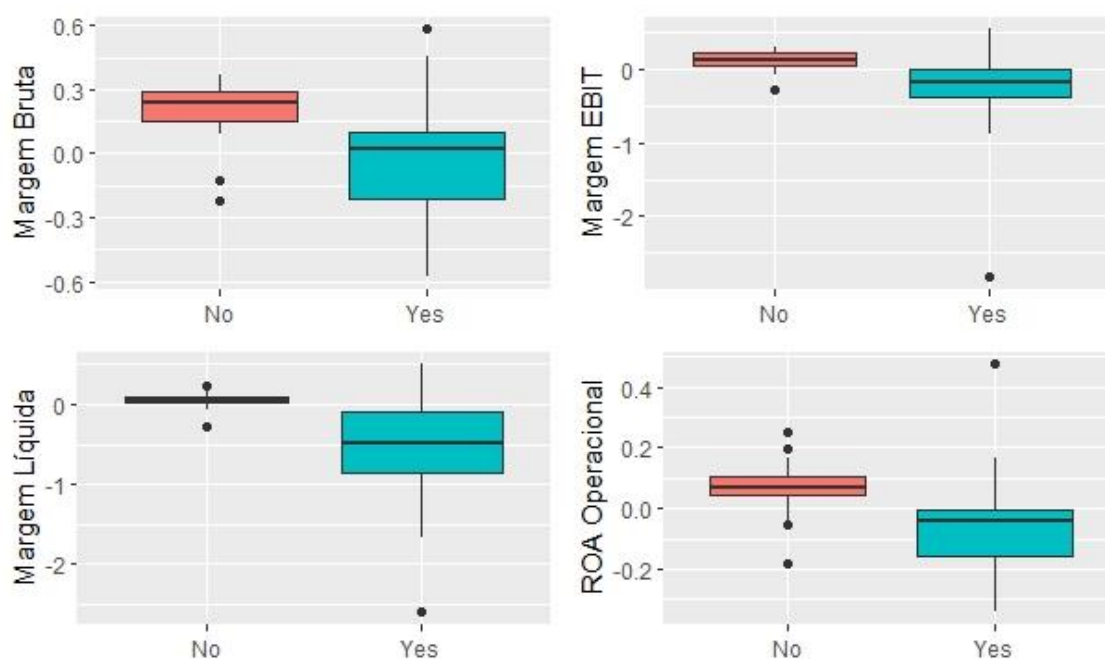
Gráfico 5 – *Boxplot* das variáveis de liquidez



Fonte: Resultados da Pesquisa.

A margem bruta, que leva em conta apenas a dedução dos impostos sobre produtos e custo dos produtos vendidos, para as empresas solventes, apresenta mediana maior e variância menor do que o grupo de empresas insolventes, como era de se esperar. As margens EBIT, para o grupo das empresas solventes, são ligeiramente maiores do que as das empresas insolventes e também possuem menor variância. A margem líquida das empresas solventes é próxima de zero, evidenciando que o setor é intensivo em capital e opera alavancado, enquanto as empresas insolventes apresentam mediana negativa, fruto da combinação entre baixas margens operacionais e alto nível de endividamento e consequentemente elevado gasto com juros. As empresas solventes também apresentam ROA mais elevado com menor variância que as empresas insolventes, demonstrando que as empresas insolventes apresentam métricas de rentabilidade operacional menores que as empresas solventes.

Gráfico 6 – *Boxplot* das variáveis de rentabilidade

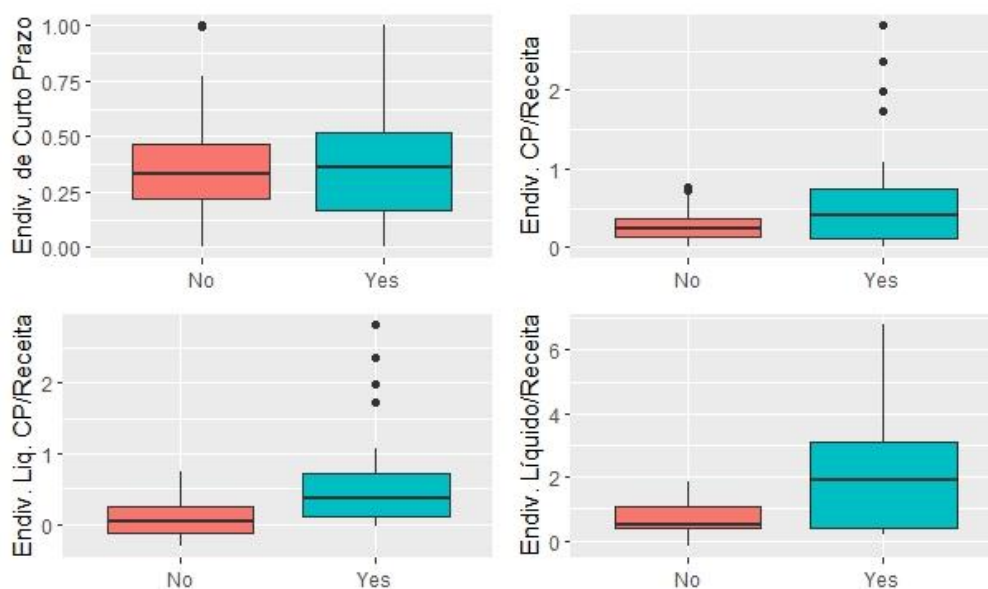


Fonte: Resultados da Pesquisa.

O endividamento de curto prazo das empresas insolventes é ligeiramente maior que das empresas solventes. Ademais, quando se analisa o endividamento de curto prazo em relação à receita, é possível notar que as empresas insolventes apresentam maior índice, demonstrando que são mais alavancadas em relação ao seu porte, ou faturamento, apresentando, portanto, menor capacidade para honrar suas dívidas. Quando se observa o endividamento líquido de curto prazo sobre a receita, que leva em conta a posição de caixa e equivalentes de caixa das

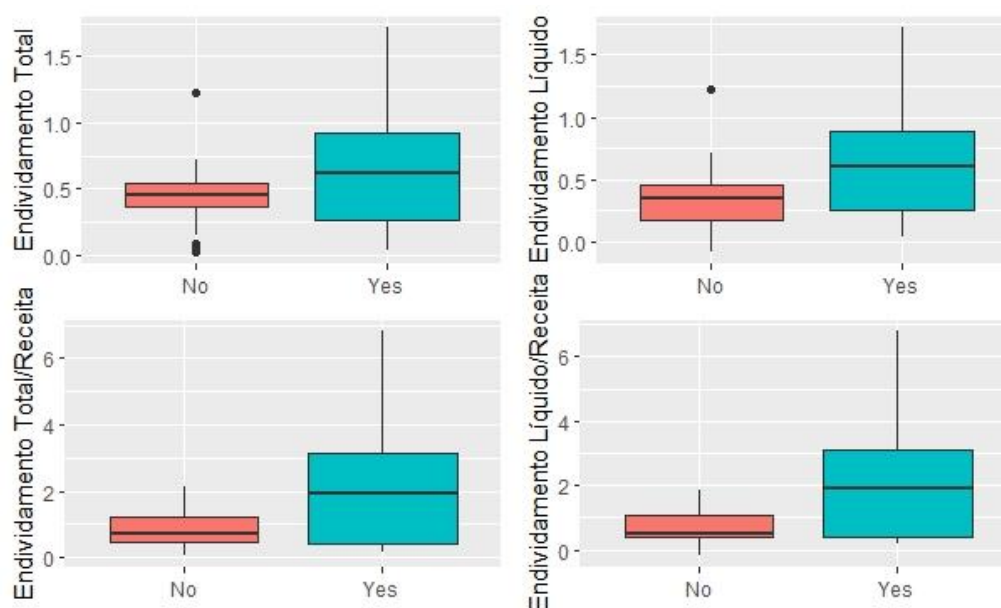
empresas, é possível observar maior diferença entre os grupos solventes e insolventes, fato esse que demonstra o baixo nível de liquidez imediata das empresas insolventes.

Gráfico 7 – *Boxplot* das variáveis de endividamento de curto prazo



Fonte: Resultados da Pesquisa.

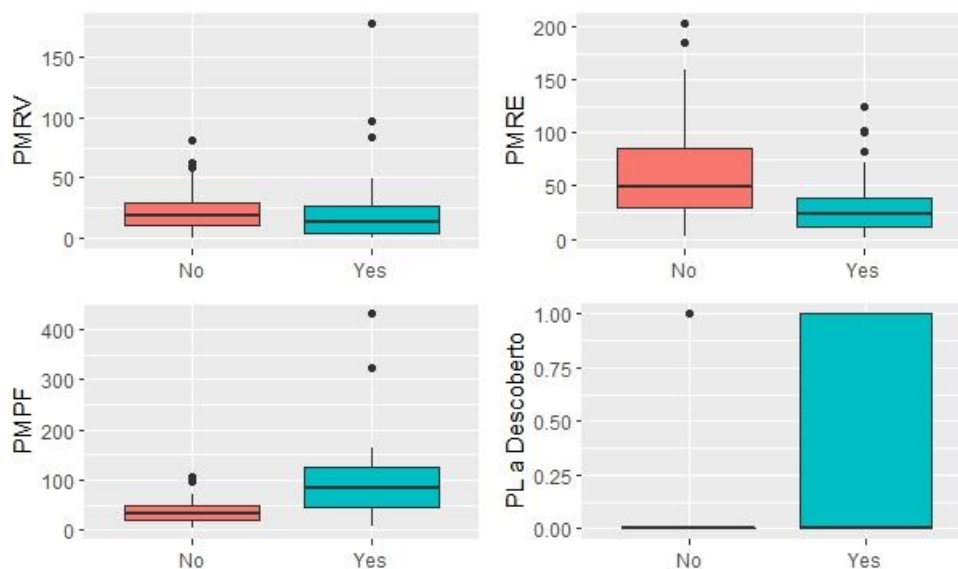
O endividamento total, que relaciona o total de dívidas onerosas em relação ao ativo total, é maior para o grupo das empresas insolventes, e quando se analisa o endividamento líquido, pode-se notar que a falta de caixa afeta diretamente as empresas insolventes. No tocante ao índice de endividamento total sobre receita líquida, as empresas insolventes também são mais alavancadas em relação ao seu porte, e essa diferença acentua-se no cálculo do endividamento líquido sobre a receita.

Gráfico 8 – *Boxplot* das variáveis de endividamento total

Fonte: Resultados da Pesquisa.

No caso do ciclo operacional, uma análise sob a óptica da gestão de recursos de curto prazo em empresas com dificuldades financeiras se faz necessária para melhor compreensão dos indicadores. As empresas em dificuldades financeiras em geral, devido à falta de liquidez, podem acabar antecipando seus recebíveis para fazer caixa, e, portanto, diminuir seus recebíveis. Pelo gráfico *Bloxplot*, observam-se os PMRV muito próximos para ambos os grupos, com o grupo de insolventes ligeiramente menor e com maior variância.

O prazo médio de renovação dos estoques também é menor em função das empresas insolventes não conseguirem montar posições importantes de estoques à espera de melhores preços no final da entressafra e, portanto, serem obrigadas a vender rapidamente seus produtos acabados para fazer caixa. O indicador de prazo médio de pagamento de fornecedores também exige a compreensão da situação das empresas insolventes para análise adequada. Quando as empresas se encontram na situação de pré-insolvência, acabam atrasando os pagamentos a fornecedores, o que faz com que o saldo da conta dos fornecedores a pagar se eleve, acarretando mais prazos de pagamentos de fornecedores. No entanto, esse prazo estendido não condiz com a análise tradicional de melhor gestão de capital de giro com negociações de prazos dilatados com fornecedores. Por fim, pode-se observar que grande parte das empresas insolventes possui patrimônio líquido a descoberto, ao passo que as empresas solventes tendem a ter patrimônio líquido positivo.

Gráfico 9 – *Boxplot* das variáveis de Prazos Médios

Fonte: Resultados da Pesquisa.

Com o intuito de analisar analiticamente quais variáveis possuem maior capacidade de discriminação entre os grupos de empresas solventes e insolventes, foram aplicados testes t para diferença de médias para cada uma das variáveis com a hipótese nula de que a média entre os grupos é igual. As variáveis EndivCP, GiroAtivo, PMRV, PMRE e PL_Rec apresentaram p-valores acima de 5%, não rejeitando a hipótese nula, portanto, são pouco eficientes para a segregação necessária para o modelo.

Tabela 3 – Teste para diferença de médias

| Variável | t | p-value | |
|----------------|---------|---------|---|
| LiqCorr | 4,3605 | 0,0001 | |
| LiqSeca | 3,8195 | 0,0002 | |
| LiqImed | 4,4214 | 0,0000 | |
| MargBru | 3,7360 | 0,0004 | |
| MargEBIT | 3,6741 | 0,0005 | |
| MargLiq | 5,2452 | 0,0000 | |
| ROA_oper | 4,0283 | 0,0002 | |
| EndivCP | 0,3691 | 0,7135 | * |
| EndivCP_Rec | -2,5345 | 0,0141 | |
| EndivLiqCP_Rec | -3,5718 | 0,0007 | |
| EndivTot | -2,1054 | 0,0398 | |
| EndivLiq | -3,0433 | 0,0036 | |
| EndivTot_Rec | -3,6411 | 0,0006 | |
| EndivLiq_Rec | -4,1485 | 0,0001 | |
| GiroAtivo | 0,5626 | 0,5759 | * |
| PMRV | -0,0266 | 0,9788 | * |
| PMRE | -0,6690 | 0,5063 | * |
| PMPF | -1,9587 | 0,0551 | |
| PL_Rec | 0,6028 | 0,5491 | * |
| PLneg | -4,4590 | 0,0000 | |

Fonte: Resultados da Pesquisa.

Além da questão de capacidade de discriminação entre as variáveis, também é importante analisar a possibilidade de ocorrência de multicolinearidade, dado que algumas variáveis tendem a demonstrar informações semelhantes a outras. Nesse caso, foi elaborada uma matriz de correlação entre as variáveis para analisar quais seriam redundantes para o modelo.

Tabela 4 – Correlação amostra t – 1

| | LiqCorr | LiqSeca | LiqImed | MargBru | MargEBIT | MargLiq | ROA_oper | EndivTot | EndivLiq | EndivTot_Rec | EndivLiq_Rec | EndivCP_Rec | EndivLiqCP_Rec | PLneg |
|----------------|---------|---------|---------|---------|----------|---------|----------|----------|----------|--------------|--------------|-------------|----------------|-------|
| LiqCorr | 1,00 | 0,89 | 0,71 | 0,12 | 0,10 | 0,19 | 0,18 | 0,14 | 0,26 | 0,09 | 0,13 | 0,05 | 0,13 | 0,18 |
| LiqSeca | 0,89 | 1,00 | 0,81 | 0,10 | 0,09 | 0,16 | 0,16 | 0,08 | 0,19 | 0,11 | 0,15 | 0,12 | 0,23 | 0,12 |
| LiqImed | 0,71 | 0,81 | 1,00 | 0,10 | 0,07 | 0,11 | 0,09 | 0,07 | 0,19 | 0,08 | 0,13 | 0,07 | 0,19 | 0,10 |
| MargBru | 0,12 | 0,10 | 0,10 | 1,00 | 0,47 | 0,51 | 0,65 | 0,07 | 0,10 | 0,18 | 0,20 | 0,02 | 0,04 | 0,17 |
| MargEBIT | 0,10 | 0,09 | 0,07 | 0,47 | 1,00 | 0,75 | 0,54 | 0,03 | 0,05 | 0,41 | 0,42 | 0,00 | 0,01 | 0,15 |
| MargLiq | 0,19 | 0,16 | 0,11 | 0,51 | 0,75 | 1,00 | 0,52 | 0,15 | 0,19 | 0,50 | 0,53 | 0,04 | 0,08 | 0,31 |
| ROA_oper | 0,18 | 0,16 | 0,09 | 0,65 | 0,54 | 0,52 | 1,00 | 0,05 | 0,08 | 0,14 | 0,15 | 0,00 | 0,01 | 0,23 |
| EndivTot | 0,14 | 0,08 | 0,07 | 0,07 | 0,03 | 0,15 | 0,05 | 1,00 | 0,95 | 0,31 | 0,32 | 0,07 | 0,09 | 0,37 |
| EndivLiq | 0,26 | 0,19 | 0,19 | 0,10 | 0,05 | 0,19 | 0,08 | 0,95 | 1,00 | 0,35 | 0,38 | 0,10 | 0,15 | 0,42 |
| EndivTot_Rec | 0,09 | 0,11 | 0,08 | 0,18 | 0,41 | 0,50 | 0,14 | 0,31 | 0,35 | 1,00 | 0,99 | 0,22 | 0,23 | 0,18 |
| EndivLiq_Rec | 0,13 | 0,15 | 0,13 | 0,20 | 0,42 | 0,53 | 0,15 | 0,32 | 0,38 | 0,99 | 1,00 | 0,23 | 0,27 | 0,20 |
| EndivCP_Rec | 0,05 | 0,12 | 0,07 | 0,02 | 0,00 | 0,04 | 0,00 | 0,07 | 0,10 | 0,22 | 0,23 | 1,00 | 0,95 | 0,00 |
| EndivLiqCP_Rec | 0,13 | 0,23 | 0,19 | 0,04 | 0,01 | 0,08 | 0,01 | 0,09 | 0,15 | 0,23 | 0,27 | 0,95 | 1,00 | 0,02 |
| PLneg | 0,18 | 0,12 | 0,10 | 0,17 | 0,15 | 0,31 | 0,23 | 0,37 | 0,42 | 0,18 | 0,20 | 0,00 | 0,02 | 1,00 |

Fonte: Resultados da Pesquisa.

Após a análise da matriz de correlação das variáveis, foram efetuadas tentativas de construção do modelo de análise discriminante retirando-se as variáveis com correlação acima de 0,8. Contudo, em função ainda do grande número de variáveis disponíveis, os modelos apresentaram variáveis estatisticamente não significantes. Também foram testados modelos utilizando-se o método de seleção de variáveis *stepwise*, porém os resultados também não foram satisfatórios.

Diante desse desafio, foi utilizada a técnica de análise de componentes principais para reduzir o número de variáveis e então obter modelos com variáveis com nível de significância aceitável. O critério para a utilização da técnica de análise de componentes principais foi dividir as variáveis em grupos de acordo com suas características. Para tanto, foram criados os seguintes grupos: LIQUIDEZ, que contém as variáveis LiqCorr, LiqSeca e LiqImed; Rentabilidade (RENT), com as variáveis MargBru, MargEBIT, MargLiq e ROA_oper; Endividamento de Curto Prazo (ENDIVCURTO), composto por EndivCP, EndivCP_Rec e EndivLiqCP_Rec; e Endividamento Total (ENDIVTOTAL); com as variáveis EndivTot, EndivLiq, EndivTot_Rec e EndivLiq_Rec.

A análise de componentes principais para as variáveis de liquidez apresentou a primeira componente com capacidade explicativa de aproximadamente 93%, acima dos 70% recomendados por Regazzi (2000).

Tabela 5 – Componentes principais de liquidez da amostra t – 1

| Variância total explicada | | | | | | |
|----------------------------------|----------------------|----------------|--------------|--|----------------|--------------|
| Componente | Autovalores iniciais | | | Somadas de extração de carregamentos ao quadrado | | |
| | Total | % de variância | % cumulativa | Total | % de variância | % cumulativa |
| 1 | 2,790 | 92,996 | 92,996 | 2,790 | 92,996 | 92,996 |
| 2 | 0,161 | 5,364 | 98,360 | | | |
| 3 | 0,049 | 1,640 | 100,000 | | | |

Fonte: Resultados da Pesquisa.

Ao utilizar-se a nova variável LIQUIDEZ, busca-se resolver os problemas de multicolinearidade encontrados anteriormente, por meio da transformação de um conjunto de variáveis intercorrelacionadas em um novo conjunto de variáveis não correlacionadas. Pelos coeficientes da equação da primeira componente principal, é possível notar que os três indicadores de liquidez possuem relevância semelhante.

Tabela 6 – Coeficientes de liquidez da amostra t – 1

| Matriz de componente | |
|-----------------------------|--------------|
| Variável | Componente 1 |
| LiqCorr | 0,963 |
| LiqSeca | 0,982 |
| LiqImed | 0,947 |

Fonte: Resultados da Pesquisa.

No caso das variáveis de rentabilidade, foram agrupadas as variáveis de margens do demonstrativo de resultados (Margem Bruta, EBIT e Líquida) juntamente com o ROA. A primeira componente foi capaz de explicar aproximadamente 82% da variância total e, portanto, será testada nos modelos de análise discriminante linear e regressão logística.

Tabela 7 – Componentes principais de rentabilidade da amostra t – 1

| Variância total explicada | | | | | | |
|----------------------------------|----------------------|----------------|--------------|--|----------------|--------------|
| Componente | Autovalores iniciais | | | Somadas de extração de carregamentos ao quadrado | | |
| | Total | % de variância | % cumulativa | Total | % de variância | % cumulativa |
| 1 | 3,268 | 81,695 | 81,695 | 3,268 | 81,695 | 81,695 |
| 2 | 0,407 | 10,165 | 91,860 | | | |
| 3 | 0,202 | 5,059 | 96,919 | | | |
| 4 | 0,123 | 3,081 | 100,000 | | | |

Fonte: Resultados da Pesquisa.

Ao analisar os coeficientes da equação da primeira componente principal, ou seja, a nova variável RENT, pode-se concluir que as margens EBIT e Líquida são as variáveis mais relevantes, seguidas pelo ROA e Margem Bruta.

Tabela 8 – Coeficientes de rentabilidade da amostra t – 1

| Matriz de componente | |
|-----------------------------|--------------|
| Variável | Componente 1 |
| MargBru | 0,886 |
| MargEBIT | 0,913 |
| MargLiq | 0,915 |
| ROA_oper | 0,901 |

Fonte: Resultados da Pesquisa.

As variáveis EndivCP, EndivCP_Rec e EndivLiqCP_Rec foram reduzidas à nova variável Endividamento de Curto Prazo (ENDIVCURTO), cujo poder explicativo atingiu aproximadamente 75%, considerado um nível aceitável para ser utilizada.

Tabela 9 – Componentes principais de dívida de curto prazo da amostra t – 1

| Variância total explicada | | | | | | |
|----------------------------------|----------------------|----------------|--------------|--|----------------|--------------|
| Componente | Autovalores iniciais | | | Somadas de extração de carregamentos ao quadrado | | |
| | Total | % de variância | % cumulativa | Total | % de variância | % cumulativa |
| 1 | 2,258 | 75,265 | 75,265 | 2,258 | 75,265 | 75,265 |
| 2 | 0,714 | 23,815 | 99,080 | | | |
| 3 | 0,028 | 0,920 | 100,000 | | | |

Fonte: Resultados da Pesquisa.

As variáveis EndivCP_Rec e EndivLiqCP_Rec possuem maior peso e consequentemente maior relevância na equação da primeira componente principal da nova variável Endividamento de Curto Prazo (ENDIVCURTO), como pode ser observado na tabela a seguir.

Tabela 10 – Coeficientes de dívida de curto prazo da amostra t – 1

| Matriz de componente | |
|-----------------------------|--------------|
| Variável | Componente 1 |
| EndivCP | 0,646 |
| EndivCP_Rec | 0,960 |
| EndivLiqCP_Rec | 0,958 |

Fonte: Resultados da Pesquisa.

A análise de componentes principais no caso das variáveis de endividamento total reduziu quatro variáveis (EndivTot, EndivLiq, EndivTot_Rec e EndivLiq_Rec), sendo que a primeira componente principal, nomeada ENDIVTOTAL, tem capacidade explicativa de aproximadamente 79% da variância total.

Tabela 11 – Componentes principais de dívida total da amostra t – 1

| Variância total explicada | | | | | | |
|----------------------------------|----------------------|----------------|--------------|--|----------------|--------------|
| Componente | Autovalores iniciais | | | Somadas de extração de carregamentos ao quadrado | | |
| | Total | % de variância | % cumulativa | Total | % de variância | % cumulativa |
| 1 | 3,152 | 78,803 | 78,803 | 3,152 | 78,803 | 78,803 |
| 2 | 0,821 | 20,521 | 99,323 | | | |
| 3 | 0,026 | 0,650 | 99,973 | | | |
| 4 | 0,001 | 0,027 | 100,000 | | | |

Fonte: Resultados da Pesquisa.

No caso dos coeficientes da equação da primeira componente principal, observa-se que as variáveis reduzidas, EndivTot, EndivLiq, EndivTot_Rec e EndivLiq_Rec, possuem praticamente o mesmo peso, ou seja, não apresentam distinção importante de relevância entre si na composição da primeira componente principal.

Tabela 12 – Coeficientes de dívida total da amostra t – 1

| Matriz de componente | |
|----------------------|-----------------|
| Variável | Componente 1 |
| EndivTot | 0,873 |
| EndivLiq | 0,895 |
| EndivTot_Rec | 0,886 |
| EndivLiq_Rec | 0,896 |

Fonte: Resultados da Pesquisa.

Também foram analisadas as correlações entre os dados dos períodos t – 2 e t – 3, ou seja, as variáveis das demonstrações financeiras de dois e três anos anteriores ao pedido de recuperação judicial pelas empresas insolventes. Algumas variáveis para esses dois períodos também apresentaram problemas de alta correlação, conforme pode-se observar nas matrizes a seguir.

Tabela 13 – Correlação da amostra t – 2

| | LiqCorr | LiqSeca | LiqImed | MargBru | MargEBIT | MargLiq | ROA_oper | PMRE | PMPF | EndivTot | EndivLiq | EndivTot_Rec | EndivLiq_Rec | EndivLiqCP_Rec | PLneg |
|----------------|---------|---------|---------|---------|----------|---------|----------|------|------|----------|----------|--------------|--------------|----------------|-------|
| LiqCorr | 1,00 | 0,74 | 0,44 | 0,04 | 0,05 | 0,12 | 0,18 | 0,12 | 0,20 | 0,04 | 0,10 | 0,03 | 0,04 | 0,14 | 0,04 |
| LiqSeca | 0,74 | 1,00 | 0,63 | 0,01 | 0,02 | 0,06 | 0,07 | 0,00 | 0,16 | 0,02 | 0,06 | 0,01 | 0,02 | 0,15 | 0,02 |
| LiqImed | 0,44 | 0,63 | 1,00 | 0,05 | 0,05 | 0,08 | 0,10 | 0,00 | 0,06 | 0,06 | 0,16 | 0,04 | 0,07 | 0,09 | 0,04 |
| MargBru | 0,04 | 0,01 | 0,05 | 1,00 | 0,72 | 0,47 | 0,49 | 0,08 | 0,04 | 0,20 | 0,20 | 0,08 | 0,08 | 0,01 | 0,13 |
| MargEBIT | 0,05 | 0,02 | 0,05 | 0,72 | 1,00 | 0,68 | 0,72 | 0,12 | 0,05 | 0,15 | 0,17 | 0,05 | 0,05 | 0,00 | 0,27 |
| MargLiq | 0,12 | 0,06 | 0,08 | 0,47 | 0,68 | 1,00 | 0,57 | 0,07 | 0,10 | 0,28 | 0,32 | 0,30 | 0,32 | 0,18 | 0,25 |
| ROA_oper | 0,18 | 0,07 | 0,10 | 0,49 | 0,72 | 0,57 | 1,00 | 0,11 | 0,08 | 0,14 | 0,19 | 0,10 | 0,11 | 0,02 | 0,31 |
| PMRE | 0,12 | 0,00 | 0,00 | 0,08 | 0,12 | 0,07 | 0,11 | 1,00 | 0,02 | 0,04 | 0,04 | 0,00 | 0,00 | 0,01 | 0,02 |
| PMPF | 0,20 | 0,16 | 0,06 | 0,04 | 0,05 | 0,10 | 0,08 | 0,02 | 1,00 | 0,00 | 0,01 | 0,03 | 0,03 | 0,02 | 0,05 |
| EndivTot | 0,04 | 0,02 | 0,06 | 0,20 | 0,15 | 0,28 | 0,14 | 0,04 | 0,00 | 1,00 | 0,92 | 0,44 | 0,45 | 0,09 | 0,31 |
| EndivLiq | 0,10 | 0,06 | 0,16 | 0,20 | 0,17 | 0,32 | 0,19 | 0,04 | 0,01 | 0,92 | 1,00 | 0,48 | 0,51 | 0,13 | 0,33 |
| EndivTot_Rec | 0,03 | 0,01 | 0,04 | 0,08 | 0,05 | 0,30 | 0,10 | 0,00 | 0,03 | 0,44 | 0,48 | 1,00 | 0,99 | 0,30 | 0,03 |
| EndivLiq_Rec | 0,04 | 0,02 | 0,07 | 0,08 | 0,05 | 0,32 | 0,11 | 0,00 | 0,03 | 0,45 | 0,51 | 0,99 | 1,00 | 0,33 | 0,04 |
| EndivLiqCP_Rec | 0,14 | 0,15 | 0,09 | 0,01 | 0,00 | 0,18 | 0,02 | 0,01 | 0,02 | 0,09 | 0,13 | 0,30 | 0,33 | 1,00 | 0,00 |
| PLneg | 0,04 | 0,02 | 0,04 | 0,13 | 0,27 | 0,25 | 0,31 | 0,02 | 0,05 | 0,31 | 0,33 | 0,03 | 0,04 | 0,00 | 1,00 |

Fonte: Resultados da Pesquisa.

Tabela 14 – Correlação da amostra t – 3

| | acpc | caixapc | mbruta | mebit | mliq | ebitat | PMRE | PMPF | dtat | dlat | dtrl | dlrl | pldescob |
|--------------|------|---------|--------|-------|------|--------|------|------|------|------|------|------|----------|
| LiqCorr | 1,00 | 0,41 | 0,00 | 0,00 | 0,03 | 0,05 | 0,13 | 0,13 | 0,04 | 0,09 | 0,00 | 0,01 | 0,03 |
| LiqImed | 0,41 | 1,00 | 0,02 | 0,02 | 0,05 | 0,04 | 0,00 | 0,07 | 0,01 | 0,11 | 0,02 | 0,04 | 0,02 |
| MargBru | 0,00 | 0,02 | 1,00 | 0,71 | 0,61 | 0,21 | 0,10 | 0,02 | 0,06 | 0,08 | 0,28 | 0,29 | 0,05 |
| MargEBIT | 0,00 | 0,02 | 0,71 | 1,00 | 0,58 | 0,31 | 0,12 | 0,06 | 0,02 | 0,03 | 0,06 | 0,07 | 0,08 |
| MargLiq | 0,03 | 0,05 | 0,61 | 0,58 | 1,00 | 0,23 | 0,14 | 0,07 | 0,21 | 0,27 | 0,30 | 0,32 | 0,24 |
| ROA_oper | 0,05 | 0,04 | 0,21 | 0,31 | 0,23 | 1,00 | 0,05 | 0,07 | 0,02 | 0,04 | 0,03 | 0,04 | 0,17 |
| PMRE | 0,13 | 0,00 | 0,10 | 0,12 | 0,14 | 0,05 | 1,00 | 0,01 | 0,05 | 0,04 | 0,00 | 0,00 | 0,03 |
| PMPF | 0,13 | 0,07 | 0,02 | 0,06 | 0,07 | 0,07 | 0,01 | 1,00 | 0,01 | 0,00 | 0,00 | 0,01 | 0,08 |
| EndivTot | 0,04 | 0,01 | 0,06 | 0,02 | 0,21 | 0,02 | 0,05 | 0,01 | 1,00 | 0,93 | 0,37 | 0,36 | 0,20 |
| EndivLiq | 0,09 | 0,11 | 0,08 | 0,03 | 0,27 | 0,04 | 0,04 | 0,00 | 0,93 | 1,00 | 0,41 | 0,43 | 0,22 |
| EndivTot_Rec | 0,00 | 0,02 | 0,28 | 0,06 | 0,30 | 0,03 | 0,00 | 0,00 | 0,37 | 0,41 | 1,00 | 0,99 | 0,03 |
| EndivLiq_Rec | 0,01 | 0,04 | 0,29 | 0,07 | 0,32 | 0,04 | 0,00 | 0,01 | 0,36 | 0,43 | 0,99 | 1,00 | 0,04 |
| PLneg | 0,03 | 0,02 | 0,05 | 0,08 | 0,24 | 0,17 | 0,03 | 0,08 | 0,20 | 0,22 | 0,03 | 0,04 | 1,00 |

Fonte: Resultados da Pesquisa.

A mesma solução aplicada para o período $t - 1$ foi utilizada para os períodos $t - 2$ e $t - 3$, ou seja, a redução do número total de variáveis pela técnica de análise de componentes principais. Os resultados obtidos para os períodos $t - 2$ e $t - 3$ estão descritos a seguir.

A nova variável LIQUIDEZ, obtida pela análise de componentes principais, tem poder de explicação de 85% da variância total para $t - 2$ e 84% para $t - 3$. Das três variáveis reduzidas, a variável LiqSeca é a que possui maior peso na equação tanto para $t - 2$ quanto para $t - 3$. A primeira componente principal RENT captura aproximadamente 83% da variância total das quatro variáveis de rentabilidade no período $t - 2$ e 74% em $t - 3$, sendo que a MargEBIT é a principal variável da equação.

Para a nova variável ENDIVCURTO, o poder de explicação das três variáveis reduzidas é de 74% para $t - 2$ e 78% para $t - 3$, e em ambos os casos as variáveis EndivCP_Rec e EndivLiqCP_Rec são as mais relevantes. Por fim, a variável obtida pela primeira componente principal ENDIVTOTAL é capaz de explicar 84% da variância total das quatro variáveis de endividamento total em $t - 2$ e 81% em $t - 3$, sendo todas as variáveis com praticamente igual importância.

Tabela 15 – Variância explicada pelas componentes principais

| Variância total explicada | | | |
|---------------------------|--------|--------|--------|
| Variável | n-1 | n-2 | n-3 |
| LIQUIDEZ | 92,996 | 84,851 | 83,886 |
| RENT | 81,695 | 83,465 | 74,163 |
| ENDIV CURTO | 75,265 | 73,504 | 77,671 |
| ENDIVTOTAL | 78,803 | 83,804 | 80,802 |

Fonte: Resultados da Pesquisa.

4.2 Análise discriminante linear

4.2.1 Modelo discriminante para amostra $t - 1$

De acordo com as análises das variáveis realizadas na seção anterior, o modelo de análise discriminante foi desenvolvido inicialmente a partir das seguintes variáveis originais GiroAtivo, PMRV, PMRE, PMPF, PL_Rec e PLneg e pelas variáveis desenvolvidas pela análise de componentes principais, a saber: LIQUIDEZ, RENT, ENDIVTOTAL, ENDIVCURTO.

Para a construção do melhor modelo discriminante linear e seleção final das variáveis a partir da pré-seleção realizada foi utilizado o método *stepwise*. Os resultados abaixo demonstram que a cada etapa realizada pelo método *stepwise* foi possível obter um Lambda de

Wilks menor, até chegar-se no modelo final cuja função discriminante possui quatro variáveis e é altamente significativa com p-valor próximo a zero.

Tabela 16 – Lambda de Wilks para a amostra t – 1

| Lambda de Wilks | | | | | | | | | |
|-----------------|---------------------|--------|-----|-----|-----|-------------|-----|-----|-------|
| Etapa | Número de variáveis | Lambda | df1 | df2 | df3 | F exato | | | |
| | | | | | | Estatística | df1 | df2 | Sig. |
| 1 | 1 | 0,713 | 1 | 1 | 56 | 23 | 1 | 56 | 0,000 |
| 2 | 2 | 0,611 | 2 | 1 | 56 | 18 | 2 | 55 | 0,000 |
| 3 | 3 | 0,566 | 3 | 1 | 56 | 14 | 3 | 54 | 0,000 |
| 4 | 4 | 0,523 | 4 | 1 | 56 | 12 | 4 | 53 | 0,000 |

Fonte: Resultados da Pesquisa.

A tabela a seguir mostra as variáveis selecionadas pelo método *stepwise*, bem como os coeficientes da função discriminante utilizados para o cálculo do *score* discriminante empregado para a classificação das empresas em solventes ou insolventes. O sinal de cada uma das variáveis da função discriminante está coerente com o esperado. A variável PLneg possui sinal positivo, uma vez que, caso a empresa possua patrimônio líquido a descoberto, maiores suas chances de insolvência, e também é a variável com maior poder de discriminação entre as quatro selecionadas. As variáveis LIQUIDEZ e RENT possuem sinal negativo, confirmando que, quanto maiores a liquidez e a rentabilidade de uma empresa, menores suas chances de falência, porém, a variável LIQUIDEZ tem o menor poder discriminante entre todas, e RENT, o segundo maior poder discriminante. Por fim, a variável ENDIVCURTO possui sinal positivo, corroborando com o fato de que empresas mais alavancadas no curto prazo possuem maiores chances de falência e é a penúltima variável em termos de poder discriminante.

Tabela 17 – Função discriminante para t – 1

| Coeficientes de função discriminante canônica | |
|---|----------|
| Variáveis | Função 1 |
| PLneg | 1,274 |
| LIQUIDEZ | -0,394 |
| RENT | -0,603 |
| ENDIVCURTO | 0,444 |
| (Constante) | -0,330 |

Fonte: Resultados da Pesquisa.

A acurácia total do modelo discriminante desenvolvido a partir da amostra de dados $t - 1$ foi de 84,5%. Do total de 29 empresas insolventes da base de dados, 22 foram classificadas corretamente, ou seja, 75,9%. No caso das empresas solventes, o modelo apresentou acurácia de 93,1%. Os testes de acurácia do modelo foram realizados pelo método *leave-one-out*. Após essa análise, é possível constatar que o erro de classificar uma empresa insolvente no grupo das solventes, erro mais crítico para o modelo, foi de 24,1%, enquanto o erro de classificar uma empresa solvente no grupo das insolventes foi de apenas 6,9%.

Tabela 18 – Resultados da classificação para a amostra $t - 1$

| Resultados da classificação | | | | | |
|-----------------------------|----------|--------|------------------------------|------|-------|
| | | RecJud | Associação ao grupo prevista | | Total |
| | | | 0 | 1 | |
| Original | Contagem | 0 | 27 | 2 | 29 |
| | | 1 | 7 | 22 | 29 |
| | % | 0 | 93,1 | 6,9 | 100,0 |
| | | 1 | 24,1 | 75,9 | 100,0 |
| Com validação cruzada | Contagem | 0 | 27 | 2 | 29 |
| | | 1 | 8 | 21 | 29 |
| | % | 0 | 93,1 | 6,9 | 100,0 |
| | | 1 | 27,6 | 72,4 | 100,0 |

Fonte: Resultados da Pesquisa.

4.2.2 Modelo discriminante para a amostra $t - 2$

Os resultados da tabela a seguir mostram que foram realizadas duas etapas pelo método *stepwise*, com a última apresentando Lambda de Wilks de 0,625. O modelo discriminante encontrado para a amostra de dados $t - 2$ conta com apenas duas variáveis, ou seja, um número inferior de variáveis em relação ao modelo $t - 1$; contudo, também se apresenta altamente significativa com p-valor próximo a zero. O fato de o modelo ter um número menor de variáveis explicativas pode estar ligado ao distanciamento temporal entre os dados amostrais e o evento do pedido de recuperação judicial, uma vez que, quanto mais distante do evento, mais solvente a empresa pode aparentar estar.

Tabela 19 – Lambda de Wilks para a amostra t – 2

| Lambda de Wilks | | | | | | | | | |
|-----------------|---------------------|--------|-----|-----|-----|-------------|-----|-----|-------|
| Etapa | Número de variáveis | Lambda | df1 | df2 | df3 | F exato | | | |
| | | | | | | Estatística | df1 | df2 | Sig. |
| 1 | 1 | 0,738 | 1 | 1 | 62 | 22 | 1 | 62 | 0,000 |
| 2 | 2 | 0,625 | 2 | 1 | 62 | 18 | 2 | 61 | 0,000 |

Fonte: Resultados da Pesquisa.

O sinal de cada uma das variáveis da função discriminante está condizente com o esperado, uma vez que a variável PMPF pode assumir coeficiente positivo quando se trata da análise de empresas com atrasos de pagamentos, conforme mencionado na sessão descritiva das variáveis, e a variável RENT possui coeficiente negativo, corroborando com o fato de que, quanto maior a rentabilidade, menores as chances de falência. Por fim, a variável RENT possui maior poder discriminante entre as duas variáveis selecionadas pelo modelo.

Tabela 20 – Função discriminante para t – 2

| Coeficientes de função discriminante canônica | |
|---|----------|
| Variáveis | Função 1 |
| PMPF | 0,019 |
| RENTABILIDADE | -0,718 |
| (Constante) | -1,179 |

Fonte: Resultados da Pesquisa.

O modelo discriminante elaborado a partir da amostra de dados t – 2 obteve taxa de acerto total de 79,7%. Considerando-se as 32 empresas insolventes da base de dados t – 2, foram classificadas corretamente 75,0% das empresas. Em relação às empresas solventes, a taxa de acerto foi de 84,4%, portanto, acima da taxa obtida para a amostra das empresas insolventes.

Tabela 21 – Resultados da classificação para a amostra t – 2

| Resultados da classificação | | | | | |
|------------------------------------|----------|--------|------------------------------|------|-------|
| | | RecJud | Associação ao grupo prevista | | Total |
| | | | 0 | 1 | |
| Original | Contagem | 0 | 27 | 5 | 32 |
| | | 1 | 8 | 24 | 32 |
| | % | 0 | 84,4 | 15,6 | 100,0 |
| | | 1 | 25,0 | 75,0 | 100,0 |
| Com validação cruzada ^b | Contagem | 0 | 27 | 5 | 32 |
| | | 1 | 9 | 23 | 32 |
| | % | 0 | 84,4 | 15,6 | 100,0 |
| | | 1 | 28,1 | 71,9 | 100,0 |

Fonte: Resultados da Pesquisa.

4.2.3 Modelo discriminante para a amostra t – 3

O modelo discriminante para a amostra de dados t – 3 construído a partir do método de seleção final de variáveis *stepwise* foi definido após três etapas, com a presença de três variáveis, Lambda de Wilks de 0,68 e a função discriminante significativa com p-valor próximo a zero.

Tabela 22 – Lambda de Wilks para a amostra t – 3

| Lambda de Wilks | | | | | | | | | |
|-----------------|---------------------|----------|-----|-----|-----|-------------|-----|-----|-------|
| Etapa | Número de variáveis | Lambda | df1 | df2 | df3 | F exato | | | |
| | | | | | | Estatística | df1 | df2 | Sig. |
| 1 | 1 | 0,866267 | 1 | 1 | 58 | 9 | 1 | 58 | 0,004 |
| 2 | 2 | 0,762486 | 2 | 1 | 58 | 9 | 2 | 57 | 0,000 |
| 3 | 3 | 0,676715 | 3 | 1 | 58 | 9 | 3 | 56 | 0,000 |

Fonte: Resultados da Pesquisa.

Permaneceram no modelo final as variáveis PMRE, PMPF e ENDIVTOTAL. A variável ENDIVTOTAL é a variável com maior poder de discriminação do modelo e apresenta sinal positivo conforme o esperado. A variável PMPF, assim como encontrado no modelo t – 2, apresentou sinal positivo, corroborando com a ideia de que empresas pré-insolventes tendem a ter atrasos de pagamento de fornecedores. Por fim, a variável PMRE apresentou o menor poder de discriminação entre as três variáveis do modelo, e o sinal de seu coeficiente foi negativo,

demonstrando que as empresas em estado pré-falimentar tendem a liquidar seus estoques para fazer caixa, mantendo o volume mínimo possível de recursos nessa rubrica.

Tabela 23 – Função discriminante para t – 3

| Coeficientes de função discriminante canônica | |
|---|----------|
| Variáveis | Função 1 |
| PMRE | -0,006 |
| PMPF | 0,016 |
| ENDIVTOTAL | 0,716 |
| (Constante) | -0,327 |

Fonte: Resultados da Pesquisa.

O resultado geral do modelo discriminante t – 3 realizado pelo método *leave-one-out* foi de 76,7% de acerto. Para esse modelo, o teste de acurácia apresentou o mesmo valor tanto no caso das empresas insolventes quanto das empresas solventes, ou seja, 76,7%.

Tabela 24 – Resultados da classificação para a amostra t – 3

| Resultados da classificação | | | | | |
|------------------------------------|----------|---|------------------------------|------|-------|
| RecJud | | | Associação ao grupo prevista | | Total |
| | | | 0 | 1 | |
| Original | Contagem | 0 | 23 | 7 | 30 |
| | | 1 | 7 | 23 | 30 |
| | % | 0 | 76,7 | 23,3 | 100,0 |
| | | 1 | 23,3 | 76,7 | 100,0 |
| Com validação cruzada ^b | Contagem | 0 | 23 | 7 | 30 |
| | | 1 | 7 | 23 | 30 |
| | % | 0 | 76,7 | 23,3 | 100,0 |
| | | 1 | 23,3 | 76,7 | 100,0 |

Fonte: Resultados da Pesquisa.

Quando se analisam os resultados por ordem cronológica em relação ao distanciamento temporal do evento pedido de recuperação judicial, observa-se que os modelos de predição de falência pelo método de análise discriminante linear apresentam uma redução do poder de previsibilidade, sendo, nesse caso, uma redução de 84,5% em t – 1 para 76,7% em t – 3. Ao avaliar-se isoladamente os resultados, verifica-se que a taxa de acerto para as empresas

solventes é reduzida de 93,1% para 76,7%, e, no caso das empresas insolventes, alvo de maior interesse do modelo, a capacidade preditiva permanece praticamente constante.

Tabela 25 – Resumo dos resultados de classificação

| Resultados da classificação | | | |
|-----------------------------|------|------|-------|
| Período | 0 | 1 | Total |
| n-1 | 93,1 | 75,9 | 84,5 |
| n-2 | 84,4 | 75,0 | 79,7 |
| n-3 | 76,7 | 76,7 | 76,7 |

Fonte: Resultados da Pesquisa.

4.3 Regressão logística

4.3.1 Regressão logística para a amostra t – 1

Assim como no modelo de análise discriminante descrito na seção anterior, o modelo de regressão logística considerou as variáveis originais selecionadas previamente, a saber: GiroAtivo, PMRV, PMRE, PMPF, PL_Rec e PLneg. Além das variáveis originais, também foram inseridas na construção do modelo de regressão logística as variáveis LIQUIDEZ, RENT, ENDIVTOTAL, ENDIVCURTO, desenvolvidas pela análise de componentes principais.

Para o desenvolvimento da regressão logística da amostra t – 1, foram utilizados os métodos *Forward stepwise* (Condicional), *Forward stepwise* (Razão de Verossimilhança), *Forward stepwise* (Wald), *Backward stepwise* (Condicional), *Backward stepwise* (Razão de Verossimilhança) e *Backward stepwise* (Wald). Todos os métodos resultaram na seleção das mesmas variáveis finais presentes no modelo, bem como nos mesmos coeficientes.

A tabela a seguir mostra a primeira e a sétima etapas dos resultados obtidos pelo método *Backward stepwise* (Wald). As variáveis finais significantes na regressão logística foram: PMRV, PMPF, LIQUIDEZ e RENT. As variáveis PMRV, LIQUIDEZ e RENT apresentaram sinal de seus coeficientes conforme esperado pela literatura tradicional de finanças, enquanto a variável PMPF apresentou sinal positivo, corroborando com os resultados encontrados no modelo discriminante, e que reforçam a ideia que o aumento do PMPF nas empresas em fase de pré-insolvência está relacionado ao atraso de pagamentos de fornecedores, e não de gestão eficiente de capital de giro.

Tabela 26 – Resultados do método *stepwise* para a amostra t – 1

| Variáveis na equação | | | | | | | |
|----------------------|-------------|--------|-------|-------|-------|---------|--------|
| | | B | S.E. | Wald | df | p-valor | Exp(B) |
| Etapa 1 ^a | GiroAtivo | 1,863 | 1,634 | 1,301 | 1,000 | 0,254 | 6,446 |
| | PMRV | 0,057 | 0,028 | 4,270 | 1,000 | 0,039 | 1,059 |
| | PMRE | -0,010 | 0,007 | 2,279 | 1,000 | 0,131 | 0,990 |
| | PMPF | 0,034 | 0,017 | 3,778 | 1,000 | 0,052 | 1,034 |
| | PL_Rec | 0,925 | 0,661 | 1,955 | 1,000 | 0,162 | 2,521 |
| | PLneg(1) | -0,240 | 2,055 | 0,014 | 1,000 | 0,907 | 0,787 |
| | LIQUIDEZ | -3,033 | 2,311 | 1,722 | 1,000 | 0,189 | 0,048 |
| | RENT | -2,880 | 1,307 | 4,854 | 1,000 | 0,028 | 0,056 |
| | ENDIVTOTAL | 0,494 | 0,882 | 0,314 | 1,000 | 0,575 | 1,639 |
| | ENDIV CURTO | 0,556 | 1,044 | 0,284 | 1,000 | 0,594 | 1,744 |
| | Constante | -4,669 | 3,452 | 1,830 | 1,000 | 0,176 | 0,009 |
| Etapa 7 ^a | PMRV | 0,049 | 0,019 | 7,044 | 1,000 | 0,008 | 1,050 |
| | PMPF | 0,028 | 0,013 | 4,547 | 1,000 | 0,033 | 1,028 |
| | LIQUIDEZ | -1,908 | 1,138 | 2,811 | 1,000 | 0,094 | 0,148 |
| | RENT | -2,634 | 0,867 | 9,235 | 1,000 | 0,002 | 0,072 |
| | Constante | -2,857 | 1,038 | 7,582 | 1,000 | 0,006 | 0,057 |

Fonte: Resultados da Pesquisa.

A acurácia total da regressão logística desenvolvida a partir da amostra de dados t – 1 foi de 87,9%. Ao se analisarem individualmente os grupos de empresas insolventes e solventes, para o primeiro grupo foram classificadas corretamente 25 empresas do total de 29, ou seja, 86,2%; e para o segundo grupo foram classificadas corretamente 26 empresas, ou 89,7% da amostra. O erro de classificar uma empresa insolvente no grupo das solventes, erro mais crítico para o modelo, foi de 13,8%, e o erro de classificar uma empresa solvente no grupo das insolventes foi ligeiramente menor, 10,3%.

Tabela 27 – Resultados da classificação para a amostra t – 1 (*logit*)

| Resultados da classificação | | | | | |
|-----------------------------|----------|--------|------------------------------|------|-------|
| | | RecJud | Associação ao grupo prevista | | Total |
| | | | 0 | 1 | |
| Original | Contagem | 0 | 26 | 3 | 29 |
| | | 1 | 4 | 25 | 29 |
| | % | 0 | 89,7 | 10,3 | 100,0 |
| | | 1 | 13,8 | 86,2 | 100,0 |

Fonte: Resultados da Pesquisa.

4.3.2 Regressão logística para a amostra t – 2

A regressão logística com a amostra de dados t – 2 também foi desenvolvida pelos métodos *Forward stepwise* e *Backward stepwise*, e assim como no modelo t – 1, os resultados obtidos foram idênticos. O método *Backward stepwise* para a amostra de dados t – 2 também levou sete etapas para chegar ao modelo final, que contou com as variáveis PMRE, PMPF e RENTABILIDADE. As variáveis PMRE e PMPF apresentaram os mesmos sinais encontrados nos modelos anteriores, e que diferem da análise tradicional de finanças. Esse fato reforça novamente a ideia de que as empresas em dificuldades financeiras tendem a queimar seus estoques para fazer caixa e também carregam uma carteira de fornecedores em atraso. O sinal da variável RENTABILIDADE está condizente com o esperado, ou seja, um aumento da rentabilidade das empresas reduz suas chances de insolvência.

Tabela 28 – Resultados do método *stepwise* para a amostra t – 2

| | | Variáveis na equação | | | | | |
|----------------------|---------------|----------------------|-------|-------|-------|---------|--------|
| | | B | S.E. | Wald | df | p-valor | Exp(B) |
| Etapa 1 ^a | GiroAtivo | -0,527 | 0,622 | 0,717 | 1,000 | 0,397 | 0,591 |
| | PMRV | 0,017 | 0,013 | 1,761 | 1,000 | 0,185 | 1,017 |
| | PMRE | -0,015 | 0,006 | 5,518 | 1,000 | 0,019 | 0,985 |
| | PMPF | 0,021 | 0,013 | 2,561 | 1,000 | 0,110 | 1,021 |
| | PLneg | 0,695 | 1,917 | 0,132 | 1,000 | 0,717 | 2,004 |
| | LIQUIDEZ | -0,586 | 0,827 | 0,502 | 1,000 | 0,478 | 0,557 |
| | RENTABILIDADE | -0,812 | 0,521 | 2,427 | 1,000 | 0,119 | 0,444 |
| | ENDIVCURTO | 1,186 | 0,996 | 1,416 | 1,000 | 0,234 | 3,273 |
| | ENDIVTOTAL | 0,842 | 0,595 | 2,003 | 1,000 | 0,157 | 2,322 |
| | Constante | -0,137 | 1,264 | 0,012 | 1,000 | 0,914 | 0,872 |
| Etapa 7 ^a | PMRE | -0,011 | 0,006 | 4,022 | 1,000 | 0,045 | 0,989 |
| | PMPF | 0,031 | 0,010 | 9,074 | 1,000 | 0,003 | 1,031 |
| | RENTABILIDADE | -1,020 | 0,415 | 6,032 | 1,000 | 0,014 | 0,360 |
| | Constante | -0,937 | 0,732 | 1,639 | 1,000 | 0,201 | 0,392 |

Fonte: Resultados da Pesquisa.

A regressão logística obtida a partir da amostra de dados t – 2 alcançou nível de acerto total de 82,8%. Do total de 32 empresas insolventes, 28 foram classificadas corretamente, ou 87,5%. No caso das empresas solventes, a acurácia do modelo foi de 78,1%, portanto, menor que a taxa das empresas insolventes. Esse foi o único modelo em que a taxa de acerto para as empresas insolventes superou a das solventes.

Tabela 29 – Resultados da classificação para a amostra t – 2 (*logit*)

| Resultados da classificação | | | | | |
|-----------------------------|----------|--------|------------------------------|------|-------|
| | | RecJud | Associação ao grupo prevista | | Total |
| | | | 0 | 1 | |
| Original | Contagem | 0 | 25 | 7 | 32 |
| | | 1 | 4 | 28 | 32 |
| | % | 0 | 78,1 | 21,9 | 100,0 |
| | | 1 | 12,5 | 87,5 | 100,0 |

Fonte: Resultados da Pesquisa.

4.3.3 Regressão logística para a amostra t – 3

O modelo final da regressão logística obtido a partir da sétima etapa do método *Backward stepwise* apresentou como variáveis significantes o Giro do Ativo, PMRE, LIQUIDEZ e RENTABILIDADE. A variável Giro do Ativo foi significativa apenas nesse modelo e apresentou sinal negativo de acordo com o esperado, pois quanto mais eficiente for a empresa na utilização de sua base de ativos, menores serão suas chances de falência. A variável PMRE manteve o sinal negativo, igualmente encontrado nas demais regressões, reforçando que as empresas em dificuldades tendem a ter redução do nível de estoque. As demais variáveis significantes no modelo, LIQUIDEZ e RENTABILIDADE, tiveram comportamento de acordo com o esperado na literatura de finanças.

Tabela 30 – Resultados do método *stepwise* para a amostra t – 3

| Variáveis na equação | | | | | | | |
|----------------------|---------------|---------|----------|-------|---------|--------|-----------|
| | B | S.E. | Wald | df | p-valor | Exp(B) | |
| Etapa 1 ^a | GiroAtivo | -7,461 | 3,545 | 4,429 | 1 | 0,035 | 0,001 |
| | PMRV | -0,001 | 0,016 | 0,006 | 1 | 0,940 | 0,999 |
| | PMRE | -0,018 | 0,007 | 5,855 | 1 | 0,016 | 0,982 |
| | PMPF | 0,005 | 0,019 | 0,076 | 1 | 0,782 | 1,005 |
| | PL_Rec | -0,165 | 0,992 | 0,028 | 1 | 0,868 | 0,848 |
| | PLneg(1) | -33,978 | 9782,130 | 0,000 | 1 | 0,997 | 0,000 |
| | LIQUIDEZ | -0,284 | 1,078 | 0,069 | 1 | 0,792 | 0,753 |
| | RENTABILIDADE | -1,876 | 1,155 | 2,637 | 1 | 0,104 | 0,153 |
| | ENDIVCURTO | 1,340 | 1,293 | 1,075 | 1 | 0,300 | 3,821 |
| | ENDIVTOTAL | 0,083 | 0,795 | 0,011 | 1 | 0,917 | 1,087 |
| | Constante | 39,184 | 9782,132 | 0,000 | 1 | 0,997 | 1,041E+17 |
| Etapa 7 ^a | GiroAtivo | -2,699 | 0,974 | 7,679 | 1 | 0,006 | 0,067 |
| | PMRE | -0,016 | 0,007 | 4,918 | 1 | 0,027 | 0,984 |
| | LIQUIDEZ | -1,413 | 0,654 | 4,666 | 1 | 0,031 | 0,243 |
| | RENTABILIDADE | -2,529 | 0,830 | 9,288 | 1 | 0,002 | 0,080 |
| | Constante | 2,733 | 0,974 | 7,876 | 1 | 0,005 | 15,379 |

Fonte: Resultados da Pesquisa.

O modelo de regressão linear obtido para a amostra $t - 3$ apresentou taxa de acerto total de 83,3%. Da amostra de 30 empresas solventes, o modelo classificou corretamente 26, apresentando uma taxa de acerto de 86,7%. Em relação à amostra de empresas insolventes, a acurácia do modelo foi de 80%, sendo uma elevada taxa quando se considera tamanho distanciamento temporal do evento pedido de recuperação judicial.

Tabela 31 – Resultados da classificação para a amostra $t - 3$ (*logit*)

| Resultados da classificação | | | | | |
|-----------------------------|----------|--------|------------------------------|------|-------|
| | | RecJud | Associação ao grupo prevista | | Total |
| | | | 0 | 1 | |
| Original | Contagem | 0 | 26 | 4 | 30 |
| | | 1 | 6 | 24 | 30 |
| | % | 0 | 86,7 | 13,3 | 100,0 |
| | | 1 | 20,0 | 80,0 | 100,0 |

Fonte: Resultados da Pesquisa.

O modelo de regressão logística apresentou maior índice de acerto total na predição de falência na amostra $t - 1$, sendo que o resultado para as amostras $t - 2$ e $t - 3$ foram muito próximos, ao redor de 83%, com os resultados de $t - 3$ ligeiramente acima, em função do alto grau de acerto em $t - 3$ para as empresas solventes. No caso das empresas insolventes, principal objeto do estudo, a acurácia dos períodos $t - 1$ e $t - 2$ foi parecida, com a taxa de acerto de $t - 2$ ligeiramente acima, o que é um fato não esperado, mas pode estar relacionado ao baixo nível de acerto do modelo $t - 2$ para as empresas solventes. Por fim, o modelo $t - 3$ apresentou um nível inferior, mas satisfatório, de acertos para as empresas insolventes, 80%, acarretando um erro de apenas 20% para um período tão distante do evento.

Tabela 32 – Resumo dos resultados da classificação (*logit*)

| Resultados da classificação | | | |
|-----------------------------|------|------|-------|
| Período | 0 | 1 | Total |
| n-1 | 89,7 | 86,2 | 87,9 |
| n-2 | 78,1 | 87,5 | 82,8 |
| n-3 | 86,7 | 80,0 | 83,3 |

Fonte: Resultados da Pesquisa.

5 CONCLUSÃO

O principal objetivo deste trabalho foi elaborar modelos de predição de falência para empresas do setor sucroalcooleiro brasileiro por meio da Regressão Logística e Análise Discriminante Linear para os três anos anteriores ao evento do pedido de recuperação judicial, e avaliar a acurácia desses modelos.

A princípio foram consideradas vinte variáveis explicativas, que posteriormente foram reduzidas a dez, por meio da técnica de análise de componentes principais, a fim de eliminar problemas de multicolinearidade, uma vez que algumas variáveis continham características informacionais semelhantes. Os modelos foram construídos inicialmente a partir das variáveis originais GiroAtivo, PMRV, PMRE, PMPF, PL_Rec e PLneg e das componentes principais do grupo de variáveis de liquidez, rentabilidade, endividamento de curto prazo e endividamento total. Por fim, dentre essas variáveis disponíveis foram selecionadas para os modelos finais as variáveis significantes pelo método *stepwise*.

As variáveis LIQUIDEZ e RENTABILIDADE foram significantes para o período das demonstrações financeiras imediatamente anteriores ao pedido de recuperação judicial, tanto pelo método da regressão linear quanto para a análise discriminante linear. Para esse período, o modelo *logit* apresentou taxa de acerto das empresas insolventes superior à análise discriminante linear em aproximadamente dez pontos percentuais, com 86,2% contra 75,9%. No modelo $t - 2$, as variáveis significantes presentes em ambos os modelos foram PMPF e RENTABILIDADE, sendo que o modelo *logit* também foi superior à análise discriminante linear para as empresas insolventes, porém com uma diferença ainda maior, 12,5%, com taxas de acerto de 87,5% e 75%, respectivamente. No último período analisado, $t - 3$, novamente o modelo *logit* foi superior no grupo das insolventes, contudo com uma vantagem inferior, de apenas 3,3%, com 80% de acerto.

Conforme observado nos trabalhos desenvolvidos anteriormente, o grau de acurácia na predição das empresas insolventes tende a diminuir quanto mais distante do pedido de recuperação judicial a empresa se encontra. No presente estudo, o modelo *logit* apresentou esse tipo de comportamento tanto para as empresas insolventes quanto na análise conjunta, com taxas de acerto decrescentes à medida que se distancia do pedido de recuperação judicial. Já os modelos obtidos pela análise discriminante apresentaram comportamento praticamente constante da taxa de acerto para as empresas insolventes e decrescente do ponto de vista da análise conjunta. Vale ressaltar que o modelo *logit* também foi superior à análise discriminante

linear sob a óptica da análise conjunta nos três períodos em questão, com acurácia de 87,9%, 82,8% e 83,3%, respectivamente.

Para se atingir resultados satisfatórios foi observada neste trabalho a necessidade de se experimentar diversas variáveis explicativas, bem como seus formatos ideais para inserção nos modelos. As várias tentativas de construção dos modelos fazem parte do processo para obtenção de resultados sólidos e com bom poder explicativo para a predição de falência. Apesar dos esforços, o presente trabalho buscou apenas trazer pequena colaboração com o campo de estudos de falência e estimular o interesse de pesquisadores para novas contribuições, sabendo-se do enorme espaço nessa área de pesquisa.

Como recomendação para estudos futuros seria válida a utilização de novas técnicas estatísticas para a construção dos modelos, uma vez que se observou, ao longo da história, que o catalisador para obtenção de resultados melhores nesse campo de predição de falências foi a aplicação de novos modelos estatísticos. Também seria de grande valia para a evolução dos modelos de predição de falência que fossem estudados cada vez mais setores específicos da Economia, em vez de macrossetores como os apresentados nos trabalhos da seção de revisão bibliográfica. Com esses trabalhos seria possível entender com maior profundidade quais as variáveis mais sensíveis para a detecção de falência para cada setor.

No campo de aplicação prática, este trabalho contribui com a possibilidade de implementação dos modelos desenvolvidos em *softwares* de instituições financeiras a fim de proporcionar melhor assertividade na concessão de crédito, minimizando os níveis de inadimplência no setor sucroalcooleiro. Também seria de grande valia para as empresas de auditoria, que poderiam utilizá-los com o intuito de aprimorar seus métodos de julgamento da situação financeira das empresas canavieiras em função de suas demonstrações financeiras.

Por fim, este trabalho tem caráter inovador por ser o primeiro a elaborar modelos de predição de falência para o setor sucroalcooleiro no Brasil, enfrentando o grande desafio da coleta de dados de empresas majoritariamente de capital fechado sediadas em praticamente todos Estados do território nacional.

REFERÊNCIAS BIBLIOGRÁFICAS

ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, v. 29, p. 589-609, 1968.

ALTMAN, E. I.; HALDEMAN, R. G.; NARAYANAN, P. ZetaTM Analysis: a new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, p. 29-54, 1977.

ALTMAN, E. I.; BAYDIA, T. K. N.; DIAS, L. M. R. Previsão de Problemas Financeiros em Empresas. *Revista de Administração de Empresas*, p. 17-28, 1979.

ANFAVEA. *Anuário da indústria automobilística brasileira 2019*. Associação Nacional dos Fabricantes de Veículos Automotores, 2019. Disponível em: www.anfavea.com.br/anuario2019/anuario.pdf.

ANP. *Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis 2019*. Agência Nacional do Petróleo, 2019. Disponível em: www.anp.gov.br/arquivos/central-conteudos/anuario-estatistico/2019/2019-anuario-versao-impressao.pdf.

BEAVER, W. Financial ratios as predictors of failure. *Journal of Accounting Research*, v. 4, p. 71-111, 1966.

BERK, J. B.; DEMARZO, P. M. *Corporate Finance*. 4. ed. Boston: Pearson, 2017.

BODIE, Z.; KANE, A.; MARCUS, A. *Investimentos*. 10. ed. Porto Alegre: McGraw-Hill, 2015.

BOOTH, D.; HAUSER, R. P. Predicting Bankruptcy with Robust Logistic Regression. *Journal of Data Science*, p. 565-584, 2011.

BRIGHAM, E. F.; EHRHARDT, M. C.; GAPENSKI, L. C. *Administração Financeira: teoria e prática*. 9. ed. São Paulo: Atlas, 2001.

BRITO, G. S.; CORRAR, L. J.; NETO, A. A. Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil. *Revista Contabilidade e Finanças*, v. 20, p. 28-43, 2009.

BROOKS, C. *Introductory econometrics for finance*. 3. ed. Cambridge: University Printing House, 2014.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 8. ed. São Paulo: Saraiva, 2013.

CASTRO JUNIOR, F. H. F. *Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais*. Orientador: Ronaldo Zwicker. 2003. Dissertação (Mestrado em Administração) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003.

CHIEH, R. S. *Indicadores financeiros trimestrais para prever falências nos setores de mineração, óleo e gás*. Orientador: Joelson Oliveira Sampaio. 2018. Dissertação (Mestrado em Economia) – Escola de Economia de São Paulo, Fundação Getulio Vargas, 2018.

DEMIRKAN, S.; PLATT, H. Financial status, corporate governance quality, and the likelihood of managers using discretionary accruals. *Accounting Research Journal*, v. 22, p. 93-117, 2009.

EPE/MME. *Análise de conjuntura dos biocombustíveis*. Empresa de Pesquisa Energética, 2018. Disponível em: www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-402/An%C3%A1lise_de_Conjuntura_Ano%202018.pdf.

_____. *Anuário estatístico de energia elétrica*. Empresa de Pesquisa Energética, 2018. Disponível em: www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-160/topico-168/Anu%C3%A1rio_2019_WEB.pdf.

GITMAN, J. L. *Princípios da Administração Financeira*. 10. ed. São Paulo: Pearson, 2004.

HONGYU, K.; JUNIOR, G. O.; SANDANIELO, V. M. Análise de componentes principais: resumo teórico, aplicação e interpretação. *Engineering and Science*, v. 1, p. 83-90, 2016.

KANITZ, S. C. *Como prever falências*. 1. ed. São Paulo: McGraw-Hill, 1972.

LECLERE, M. J. Bankruptcy studies and ad hoc variable selection: a canonical correlation analysis. *Review of Accounting and Finance*, v. 5, p. 410-422, 2006.

MATIAS, A. B.; SIQUEIRA, J. O. Risco Bancário: modelo de previsão de insolvência de bancos no Brasil. *Revista de Administração da Universidade de São Paulo*, v. 31, p. 19-28, 1996.

OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, v. 18, p. 109-131, 1980.

REGAZZI, A. J. Análise multivariada. *Centro de Ciências Exatas e Tecnológicas*, Universidade Federal de Viçosa, 2001.

ROSS, S. A.; WESTERFIELD, R. W.; JAFFE, J. F. *Administração Financeira: Corporate Finance*. 7. ed. São Paulo: Atlas, 2008.

SANVICENTE, A. Z.; MINARDI, A. F. Identificação de indicadores contábeis significativos para a previsão de concordata de empresas. *Finance Lab Working Papers*, Ibmecc, 1998.

SILVA, J. P. Avaliação da saúde financeira das empresas. *Revista de Administração de Empresas*, p. 41-47, 1983.

UNICA. *Balanco de atividades 2012/13 a 2018/19*. União da Indústria de Cana-de-Açúcar, 2019. Disponível em: www.unica.com.br/wp-content/uploads/2019/06/Relatorio-Atividades-201213-a-201819.pdf.

USDA. *Sugar: world markets and trade*. United States Department of Agriculture, 2019. Disponível em: <https://apps.fas.usda.gov/psdonline/circulars/sugar.pdf>.

WOOLDRIDGE, J. M. *Introdução à Econometria*. 3. ed. São Paulo: Cengage Learning, 2016.