

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA DE SÃO PAULO – IFSP

RODRIGO PALARIA DOS SANTOS

**PREVISÃO DE INSOLVÊNCIA COM REDES NEURAIS
ARTIFICIAIS AUTO-ORGANIZÁVEIS: ESTUDO
COMPARATIVO**

São Paulo

2013

RODRIGO PALARIA DOS SANTOS

**PREVISÃO DE INSOLVÊNCIA COM REDES NEURAIS
ARTIFICIAIS AUTO-ORGANIZÁVEIS: ESTUDO
COMPARATIVO**

Dissertação de Mestrado apresentado ao Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP como parte dos requisitos para obtenção do título de Mestre em Automação e Controle de Processos.

Orientador: Prof. Dr. Paulo Roberto Barbosa

São Paulo

2013

S238 Santos, Rodrigo Palaria dos.

Previsão de insolvência com redes neurais artificiais auto-organizáveis: estudo comparativo / Rodrigo Palaria dos Santos – 2013.

94 f.

Orientador: Prof. Dr. Paulo Roberto Barbosa

Dissertação (Mestrado Profissional em Automação e Controle de Processos) - Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, IFSP, 2013.

1. Previsão de insolvência 2. Redes neurais artificiais 3. Mapa auto - organizável 4. Análise discriminante I. Santos, Rodrigo Palaria dos. II. Instituto Federal de Educação, Ciência e Tecnologia de São Paulo. III Título

CDU 681.0



ATA DE EXAME DE DEFESA DE DISSERTAÇÃO

Nome do Programa: **Mestrado Profissional em Automação e Controle de Processos**

Nome do(a) Aluno(a): **Rodrigo Palaria dos Santos**

Nome do Orientador: **Prof. Dr. Paulo Roberto Barbosa**

Nome do Co-orientador:

Título do Trabalho: **'Previsão de Insolvência com Redes Neuraís Auto-Organizáveis: Estudo Comparativo'**

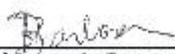
Abaixo o resultado de cada participante da Banca Examinadora

Nome completo dos Participantes Titulares da Banca	Sigla da Instituição	Aprovado / Não Aprovado
Prof. Dr. Paulo Roberto Barbosa – Orientador	IFSP – SPO	Aprovado
Prof. Dra. Iracema Hiroko Iramina Arashiro	IFSP – SPO	aprovado
Prof. Dr. Énio Carlos Segatto	IFSP – SPO	APROVADO
Nome completo dos Participantes Suplentes da Banca	Sigla da Instituição	Aprovado / Não Aprovado
Prof. Dr. Carlos Frajuca	IFSP – SPO	
Prof. Dra. Graziela Marchi Tiago	IFSP – SPO	

Considerando-se: APROVADO
 NÃO APROVADO

Assinaturas

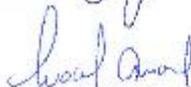
São Paulo, 16 de outubro de 2013



Presidente da Banca



Membro Interno



Membro Externo

Observações:

- Incluir citações de trabalhos similares, destacando contribuições
- Inserir revisão bibliográfica

DEDICATÓRIA

À Margarida, com amor, admiração e gratidão por sua compreensão, carinho, presença e constante apoio.

AGRADECIMENTOS

Aos meus pais pelo amor, carinho, dedicação e apoio constante para meu desenvolvimento pessoal, educacional e profissional.

Ao meu irmão que sempre acreditou nos meus objetivos e me ajudou a ver as coisas sob novas perspectivas.

Ao meu orientador Paulo Roberto Barbosa por sua compreensão, ensinamentos e orientação a este trabalho.

Aos amigos Alessandra Oliveira Soares, Dezidério Machado Lima, João Paulo Cabral da Silva, Natalia Motta Pinto e Ricardo Ballestero por sempre estarem presentes e tornarem a vida mais divertida.

Ao Instituto Federal de Educação, Ciência e Tecnologia de São Paulo pela oportunidade de realização do curso de mestrado profissional.

Aos demais amigos, parentes e professores que, direta ou indiretamente, contribuíram para a conclusão deste trabalho.

RESUMO

Frente às alterações no cenário econômico a avaliação e o entendimento da situação econômica das empresas tornam-se fundamentais para os investidores e acionistas. Com a utilização de variáveis e índices extraídos dos demonstrativos contábeis foram desenvolvidos dois modelos de previsibilidade, sendo um utilizando os métodos convencionais (análise discriminante) e outro o não convencional (redes neurais artificiais auto-organizáveis). As empresas utilizadas no estudo declararam insolvência entre 2004 e 2012, e as empresas solventes foram selecionadas a partir do índice Ibovespa em 2012. Devido ao elevado número de variáveis disponíveis foi realizada a redução de dimensionalidade através da análise de componentes principais. Avaliando os resultados nota-se que os modelos neurais apresentam resultados superiores e mantêm a estabilidade por um período maior de tempo para a previsão de insolvência de empresas.

Palavras chaves: Previsão de Insolvência, redes neurais artificiais, mapa auto-organizável, análise discriminante.

ABSTRACT

Considering the changes in the economic environment, understand the economic situation of companies become crucial for investors and shareholders. With the use of variables and indices extracted from the financial reports were developed two models of predictability, one using conventional methods (discriminant analysis) and other non-conventional (self-organizing neural networks). The companies used in the study declared bankruptcy between 2004 and 2012, and solvent companies were selected from the Bovespa index in 2012. Due the high number of available variables was performed a dimensionality reduction through principal component analysis. Evaluating the results we note that the neural models show superior results and maintain stability for a longer period of time for predicting corporate insolvency.

Key words: Insolvency prediction, artificial neural network, self-organized map, discriminant analysis.

SIGLAS E ABREVIações

BM&F	Bolsa de mercadoria e futuros
BP	Balanço patrimonial
CORR	Correlação
COV	Covariança
CP	Componentes principais
CTE	Constante
DET	Determinante
DISCR	Análise discriminante
DRE	Demonstrativo do resultado do exercício
RNA	Redes neurais artificiais
SOM	<i>Self-organazing map</i>

LISTA DE ILUSTRAÇÕES

Figura 1.	Evolução da demanda de crédito para empresas de grande porte	15
Figura 2.	Evolução do número de falências e recuperação judicial	15
Figura 3.	Representação de uma função de análise discriminante linear	22
Figura 4.	Fluxo metodológico de análise de dados.....	28
Figura 5.	Descrição da construção de gráfico <i>Boxplot</i>	36
Figura 6.	Topologias existentes para redes auto-organizáveis.....	43
Figura 7.	Arquitetura de uma rede auto-organizável.....	45
Figura 8.	<i>Boxplot</i> referente à variável B4.....	47
Figura 9.	<i>Boxplot</i> referente à variável C2.....	48
Figura 10.	<i>Boxplot</i> referente à variável A3.....	48
Figura 11.	<i>Boxplot</i> referente à variável B2.....	49
Figura 12.	Topologia do modelo de rede neural artificial auto-organizável.....	57
Figura 13.	Distância entre os neurônios após o treinamento da rede.....	58
Figura 14.	Avaliação do impacto das variáveis na alocação dos neurônios.	59
Figura 15.	Alocação de empresas por neurônio	61
Figura 16.	Posição dos neurônios após o treinamento.....	62
Figura 17.	Assertividade de modelo por metodologia	65
Figura 18.	Assertividade dos modelos considerando as empresas solventes	66
Figura 19.	Assertividade dos modelos considerando as empresas insolventes	66
Figura 20.	<i>Boxplot</i> referente à variável A1.....	75
Figura 21.	<i>Boxplot</i> referente à variável A2.....	75
Figura 22.	<i>Boxplot</i> referente à variável A3.....	76
Figura 23.	<i>Boxplot</i> referente à variável A4.....	76
Figura 24.	<i>Boxplot</i> referente à variável A5.....	77
Figura 25.	<i>Boxplot</i> referente à variável B1.....	77
Figura 26.	<i>Boxplot</i> referente à variável B2.....	78

Figura 27.	<i>Boxplot</i> referente à variável B3.....	78
Figura 28.	<i>Boxplot</i> referente à variável B4.....	79
Figura 29.	<i>Boxplot</i> referente à variável C1.....	79
Figura 30.	<i>Boxplot</i> referente à variável C2.....	80
Figura 31.	<i>Boxplot</i> referente à variável C3.....	80
Figura 32.	<i>Boxplot</i> referente à variável C4.....	81
Figura 33.	<i>Boxplot</i> referente à variável C5.....	81
Figura 34.	<i>Boxplot</i> referente à variável D1.....	82
Figura 35.	<i>Boxplot</i> referente à variável D2.....	82
Figura 36.	<i>Boxplot</i> referente à variável D3.....	83
Figura 37.	<i>Boxplot</i> referente à variável D4.....	83
Figura 38.	<i>Boxplot</i> referente à variável D5.....	84
Figura 39.	<i>Boxplot</i> referente à variável D6.....	84
Figura 40.	<i>Boxplot</i> referente à variável E1.....	85
Figura 41.	<i>Boxplot</i> referente à variável E2.....	85
Figura 42.	<i>Boxplot</i> referente à variável E3.....	86
Figura 43.	<i>Boxplot</i> referente à variável E4.....	86
Figura 44.	Avaliação de normalidade CP01.....	92
Figura 45.	Avaliação de normalidade CP02.....	92
Figura 46.	Avaliação de normalidade CP03.....	93
Figura 47.	Avaliação de normalidade CP04.....	93
Figura 48.	Avaliação de normalidade CP05.....	94
Figura 49.	Avaliação de normalidade CP06.....	94
Figura 50.	Avaliação de normalidade CP07.....	95
Figura 51.	Avaliação de normalidade CP08.....	95
Figura 52.	Avaliação de normalidade CP09.....	96
Figura 53.	Avaliação de normalidade CP10.....	96

LISTA DE QUADROS

Quadro 1.	Comparativo entre as metodologias para previsibilidade de dados.....	26
Quadro 2.	Empresas insolventes amostradas para o estudo	31
Quadro 3.	Empresas solventes amostradas para o estudo	31
Quadro 4.	Modelos desenvolvidos para predição utilizando análise discriminante	33
Quadro 5.	Representação das componentes principais.....	40
Quadro 6.	Empresas selecionadas para validação	41
Quadro 7.	Empresas selecionadas para treinamento.....	41
Quadro 8.	Classificação dos tipos de erros existentes	46
Quadro 9.	Autovalores resultantes da matriz de covariância.....	50
Quadro 10.	Coeficientes para constituição das componentes principais.....	51
Quadro 11.	Coeficientes resultantes do modelo de análise discriminante	54
Quadro 12.	Resultados do modelo de análise discriminante (treinamento)	55
Quadro 13.	Resultados do modelo de análise discriminante (validação)	55
Quadro 14.	Influência das variáveis para classificação nos neurônios.....	60
Quadro 15.	Caracterização dos neurônios	61
Quadro 16.	Resultados do modelo de redes neurais artificiais (treinamento)	63
Quadro 17.	Resultados do modelo de redes neurais artificiais (validação)	64
Quadro 18.	Detalhamento de assertividade por modelo e por empresa	67
Quadro 19.	Matriz de correlação para os dados mais recentes (ano 5).	87
Quadro 20.	Matriz de correlação para os dados do ano 4.....	88
Quadro 21.	Matriz de correlação para os dados do ano 3.....	89
Quadro 22.	Matriz de correlação para os dados do ano 2.....	90
Quadro 23.	Matriz de correlação para os dados do ano 1.....	91

SUMÁRIO

1.	Introdução	14
1.1.	Situação-problema.....	14
1.2.	Objetivos.....	16
2.	Conceituação teórica.....	17
2.1.	Conceito de insolvência	17
2.2.	Avaliação financeira de empresas	18
3.	Revisão bibliográfica	21
3.1.	Métodos convencionais: estatísticos	21
3.2.	Métodos não convencionais: redes neurais artificiais.....	23
3.3.	Estudo comparativo de metodologias.....	25
4.	Metodologia.....	28
4.1.	Amostragem dos dados	29
4.1.1.	Metodologia de amostragem.....	29
4.2.	Exploração dos dados	32
4.2.1.	Definição de variáveis	32
4.2.2.	Análise descritiva	35
4.3.	Modificação dos dados	37
4.3.1.	Análise de componentes principais.....	37
4.4.	Modelagem dos dados	40
4.4.1.	Desenvolvimento.....	40
4.4.2.	Análise discriminante	41
4.4.3.	Redes neurais artificiais auto-organizáveis.....	43
4.5.	Validação de modelos	45
5.	Análise dos resultados	47
5.1.	Análise descritiva.....	47

5.2.	Redução de dimensionalidade	50
5.3.	Modelos de previsão.....	52
5.3.1.	Modelo utilizando análise discriminante.....	53
5.3.2.	Modelo utilizando redes neurais artificiais.....	57
5.4.	Comparação entre as metodologias	64
6.	Considerações finais.....	68
	Referências.....	70
	Apêndice A: Glossário contábil	73
	Anexo 01: Análise descritiva detalhada	75
	Anexo 02: Matriz de correlação.....	87
	Anexo 03: Gráficos de avaliação de normalidade	92

1. INTRODUÇÃO

Frente às alterações no cenário econômico, a avaliação e o entendimento da situação econômica das empresas tornam-se fundamentais para os investidores e acionistas. Esses dados são especialmente relevantes quando analisados com as oscilações na estabilidade dos mercados externos e alterações nas condições econômicas nacionais, como a redução das taxas de juros.

As demonstrações contábeis que retratam a situação financeira corroboram para a avaliação econômica das empresas e análise dessas pode reduzir os casos de inadimplência. Isso porque, a partir das variáveis disponíveis, torna-se possível a elaboração de um modelo de previsão de insolvência de empresas, que minimiza as concessões a empresas suscetíveis a situações contingenciais, tais como a recuperação judicial e a falência.

1.1. SITUAÇÃO-PROBLEMA

Segundo levantamento realizado pelo Instituto Brasileiro de Planejamento Tributário (IBPT) é estimado que 574.385 novos empreendimentos públicos e privados tenham sido abertos entre janeiro e abril de 2012 em todo o Brasil, sendo os estados de São Paulo e Minas Gerais os que mais se destacam, representando respectivamente, 159.106 e 61.512 novos estabelecimentos (*Infomoney*, 2012).

O reflexo desse crescimento pode ser observado diretamente na demanda por crédito pelas empresas, conforme verificado na Figura 1, que indica um constante aumento na demanda por crédito pelas empresas de grande porte quando analisado os anos de 2011, 2012 e 2013.

Em paralelo ao crescimento da demanda de crédito, observa-se na Figura 2 um aumento no número de recuperações judiciais das empresas. Na qual, estudos realizados pela empresa Serasa Experian indicam que houve alta de 16,5% nas recuperações judiciais requeridas em relação ao mesmo período de 2012, em que ocorreram 395 pedidos frente aos 460 em 2013. Com relação aos pedidos de falência, houve queda de 1,5% em relação ao mesmo período de 2012, redução de 975 em 2012 para 885 em 2013. Porém é necessário destacar que as falências

decretadas não refletem a conjuntura, uma vez que muitas decisões judiciais demoram até dois anos.

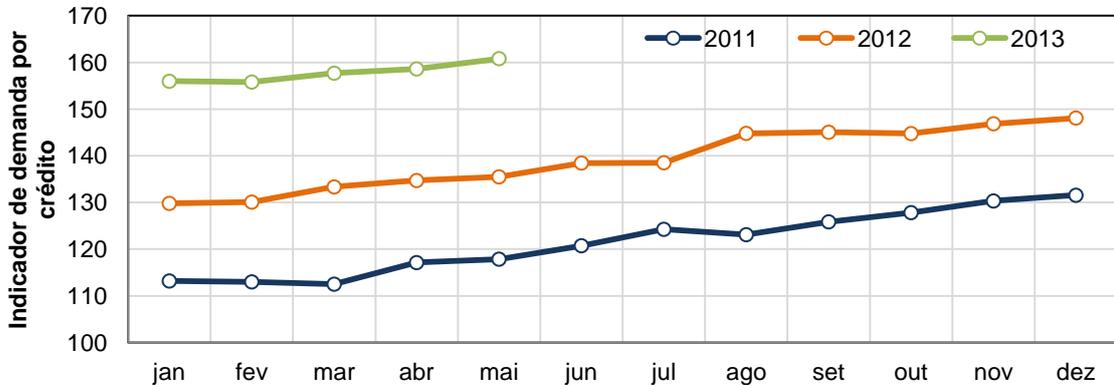


Figura 1. Evolução da demanda de crédito para empresas de grande porte (Fonte: Serasa Experian, 2013)

Desta forma, durante o processo de definição metodológica, é fundamental avaliar as especificidades e características envolvidas em cada abordagem, sendo necessário o estudo comparativo para a tomada de decisão, avaliando as vantagens, desvantagens e diferenciais atrelados a cada metodologia, de forma a entender previamente a viabilidade para utilização em estudos preditivos.

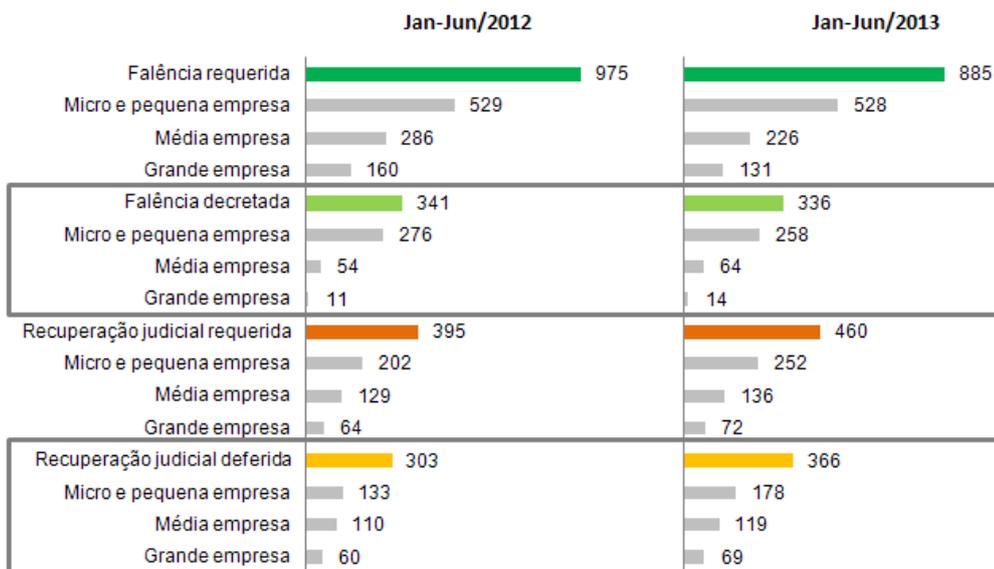


Figura 2. Evolução do número de falências e recuperação judicial (Fonte: Serasa Experian, 2013)

Devido à grande quantidade de informação disponível durante a análise, diversos métodos matemáticos podem ser implementados para prever a situação econômica destas empresas. Comumente, os métodos estatísticos são empregados como metodologia para previsão de dados. No entanto, as redes neurais artificiais podem constituir-se como um método alternativo para este fim.

1.2. OBJETIVOS

O objetivo principal deste trabalho é a proposição de um modelo de redes neurais artificiais auto-organizáveis para a predição de insolvência de empresas que apresente resultados comparáveis aos modelos baseados na metodologia estatística de análise discriminante.

Além disso, os objetivos secundários serão:

- Estudar as diferenças conceituais entre as metodologias para a previsão de insolvência baseado nos diferenciais envolvidos;
- Avaliar a sensibilidade dos modelos desenvolvidos ao longo do tempo, verificando se os modelos mantêm-se estáveis ou não.

A fim de atingir os objetivos em questão, realizou-se um levantamento de modelos desenvolvidos para a previsão de insolvência baseados em métodos estatísticos durante sua elaboração, para posteriormente elaborar novos modelos a partir das variáveis já utilizadas e assim realizar a comparação dos resultados entre as metodologias.

2. CONCEITUAÇÃO TEÓRICA

Com relação à insolvência de empresas, é importante conceituar que são consideradas empresas insolventes aquelas que se encontram em recuperação judicial ou em falência. No caso de recuperação judicial, o responsável pela empresa permanece na gestão dos negócios com supervisão do poder judiciário, enquanto que na falência, o devedor é afastado das atividades e o poder judiciário passa a conduzir o processo de liquidação das dívidas.

O Balanço Patrimonial (BP) e o Demonstrativo de Resultado do Exercício (DRE) destacam-se como as principais fontes de avaliação da situação econômica das empresas pois ambos são constituídos por diversos índices contábeis das empresas.

2.1. CONCEITO DE INSOLVÊNCIA

Sob o ponto de vista jurídico, o termo insolvência exprime a impossibilidade do devedor honrar integralmente seus débitos, dado a insuficiência de seu patrimônio para a geração de recursos e meios necessários para pagamento. Dentre os quesitos observados, são avaliados a incapacidade de pagar, a cassação de pagamento, a impontualidade e a enumeração ou indicação de fatos concretos regulamentados por lei que desabone aquele a quem se deve dinheiro ou outro bem (Campinho, 2012).

No Brasil, a insolvência foi inicialmente regulamentada pelo Decreto Lei n 7.661 de 1945, conhecida como a lei de falência, porém, em 2005, foi constituída a Lei 11.101, que adaptou a legislação à nova ordem jurídica, incorporando características recentes das empresas. Essa nova legislação aumentou a abrangência e a flexibilidade nos processos de recuperação de empresas da condição de insolventes e define critérios para a recuperação judicial, a recuperação extrajudicial e a falência (Gusmão, 2009).

Apesar da abrangência do normativo, este não é aplicado às empresas públicas e sociedades de economia mista, assim como às instituições financeiras públicas ou privadas, cooperativa de crédito, consócio, entidade de previdência

complementar, sociedade operadora de plano de saúde, sociedade seguradora, sociedade de capitalização e outras entidades legalmente equiparadas.

Conceitualmente, a recuperação apresenta um somatório de providências de ordem econômico financeiras, organizacionais e jurídicas; podendo ser classificada em recuperação judicial e extra-judicial. A recuperação judicial busca conciliar os interesses dos credores e devedores através de um processo jurídico, conforme definido na legislação:

Art. 47. A recuperação judicial tem por objetivo viabilizar a superação da situação de crise econômico-financeira do devedor, a fim de permitir a manutenção da fonte produtora, do emprego dos trabalhadores e dos interesses dos credores, promovendo, assim, a preservação da empresa, sua função social e o estímulo à atividade econômica (Lei 11.101/2005).

Por outro lado, a recuperação extra-judicial corresponde a uma tentativa de negociação informal entre devedor e credores, a fim de celebrarem pactos e repactuações (Campinho, 2010), conforme definido na legislação:

Art. 162. O devedor poderá requerer a homologação em juízo do plano de recuperação extrajudicial, juntando sua justificativa e o documento que contenha seus termos e condições, com as assinaturas dos credores que a ele aderiram (Lei 11.101/2005).

A falência difere dos métodos de recuperação judicial e extrajudicial pois oferece aos credores maior segurança do desastre econômico do devedor comum, impedindo preferências injustas, abusos e fraudes; priorizando a imparcialidade durante a liquidação e quitação das dívidas. Desta forma, a decretação judicial da falência constata que a crise econômico-financeira não foi ultrapassada pela recuperação, judicial ou extrajudicial, impossibilitando o empresário de prosseguir com suas atividades (Gusmão, 2009).

2.2. AVALIAÇÃO FINANCEIRA DE EMPRESAS

Como forma de avaliar a condição financeira das empresas, e inclusive estabelecer a condição de insolvência, pode se utilizar os demonstrativos contábeis. No Brasil, os demonstrativos são regidos pela Lei nº 6.404/76 que, apesar de ser

uma lei relativa às sociedades por ações, suas regras estendem-se às demais sociedades sujeitas à tributação do imposto de renda (Assaf, 2003).

Existem três demonstrativos contábeis básicos que resumem a situação econômica-financeira da empresa:

- **Balanço patrimonial:** é considerado uma lista dos ativos e dos passivos da empresa em um determinado momento, sendo que a diferença entre o ativos e passivos é o valor líquido da empresa, também conhecido com patrimônio líquido (Macus *et al.*, 2007). De acordo com a legislação, empresas devem elaborar o balanço patrimonial do exercício discriminando as informações do ativo e do passivo, respeitando as classificações estabelecidas:

§ 1º No ativo, as contas serão classificadas nos seguintes grupos:

a) ativo circulante; b) ativo realizável a longo prazo; c) ativo permanente, dividido em investimentos, imobilizado, intangível e diferido.

§ 2º No passivo, as contas serão classificadas nos seguintes grupos:

a) passivo circulante; b) passivo exigível a longo prazo; c) resultados de exercícios futuros; d) patrimônio líquido, dividido em capital social, reservas de capital, ajustes de avaliação patrimonial, reservas de lucros, ações em tesouraria e prejuízos acumulados. (Art. 178 da Lei nº 6.404/76)

- **Demonstrativo de resultado do exercício:** tem como finalidade exclusiva apurar o lucro ou prejuízo do exercício, englobando as receitas, as despesas, os ganhos e as perdas, independentemente de seus pagamentos e recebimentos (Assaf, 2003). Desta forma, o demonstrativo reflete os resultados gerados pelas decisões de investimentos tomadas pelos gestores em um determinado período de tempo (Malaga, 2009). De acordo com a legislação, as empresas devem discriminar na Demonstração do Resultado do Exercício as seguintes informações:

“a receita bruta das vendas e serviços, as deduções das vendas, os abatimentos e os impostos; a receita líquida das vendas e serviços, o custo das mercadorias e serviços vendidos e o lucro bruto; as despesas com as vendas, as despesas financeiras, deduzidas das receitas, as despesas gerais e administrativas, e outras despesas operacionais; o lucro ou prejuízo operacional, as outras receitas e as outras despesas; o resultado do exercício antes do Imposto sobre a Renda e a provisão para o imposto; as participações de debêntures, empregados, administradores e partes beneficiárias, mesmo na forma de instrumentos financeiros, e de instituições ou fundos de assistência ou previdência de empregados, que não se

caracterizem como despesa; o lucro ou prejuízo líquido do exercício e o seu montante por ação do capital social.” (Art. 187 da Lei nº 6.404/76)

- Demonstrativo de fluxo de caixa: a demonstração de fluxos de caixa substitui o que antes era chamado de demonstração das mudanças na posição financeira ou demonstração do fluxo de fundos, correspondendo a um relatório do fluxo gerado pelas operações da empresa, dos investimentos e das atividades financeiras (Damodaran, 2004).

Todos os demonstrativos são compostos por inúmeras variáveis que subsidiam o entendimento econômico das instituições, sendo fundamental o entendimento dos princípios contábeis, tais como investimentos, financiamentos, receitas, custos e despesas, pois subsidiam uma análise financeira adequada e de qualidade.

Segundo Malaga (2009), as demonstrações financeiras fornecem informações para analisar o histórico e potencial de crescimento, riscos envolvidos, qualidade dos resultados e inclusive os impactos das tomadas de decisões realizadas. Estes são fatores essenciais para o entendimento e classificação de empresas em uma possível situação de insolvência.

3. REVISÃO BIBLIOGRÁFICA

Através dos dados financeiros contidos nos demonstrativos contábeis e de outras informações disponibilizadas pelas empresas, torna-se possível constituir modelos preditivos através de técnicas matemáticas.

Os métodos estatísticos apresentam características específicas e suposições atreladas a sua implementação, fator que estimula a utilização de outros métodos de previsibilidade, dentre os quais as redes neurais artificiais, que possuem características complementares aos métodos convencionais

3.1. MÉTODOS CONVENCIONAIS: ESTATÍSTICOS

Dentre os métodos estatísticos utilizados para técnicas de previsão a análise discriminante é utilizada por ser capaz de classificar ou segmentar indivíduos, empresas ou observações em grupos homogêneos de acordo com as características envolvidas (Hair *et al.*, 2007).

A análise discriminante baseada na forma de utilização atual foi primeiramente apresentada por Fisher em 1936, quando se estabeleceu como uma técnica estatística que permite estudar as características de dois ou mais grupos de objetos através da análise simultânea de suas características (McLachlan, 2004).

Neste método, o processo de classificação considera o conjunto de variáveis independentes selecionadas e prioriza sua relevância, para na sequência ponderar cada variável remanescente de acordo com sua influência na variável resposta, também denominada variável dependente. Através da aplicação desta técnica são gerados modelos estatísticos que podem ser utilizados no processo preditivo de classificação de novos elementos, tornando-se uma ferramenta de grande utilidade em diversas áreas do conhecimento (Cooper *et al.*, 2003).

Quando o objetivo é a previsão de insolvência, diversos autores priorizam a técnica de análise discriminante para predição, uma vez que sua variável resposta deve ser categórica e pode apresentar quantas categorias forem necessárias para o estudo (Johnson e Wichern, 2007).

Segundo Hair *et al* (2007), a análise discriminante utiliza-se do método de mínimos quadrados que busca minimizar a soma das diferenças quadráticas entre os valores reais e os previstos para a variável dependente.

A equação discriminante é uma função linear da forma:

$$D_i = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_n * X_n \quad (1)$$

Onde D_i corresponde a i -ésima função discriminante, b_0 ao intercepto da equação, b_p ao coeficiente da p -ésima variável independente e X_n a n -ésima variável utilizada no modelo. Assim, a quantidade de funções discriminantes geradas está relacionada ao número de categorias da variável resposta. Consequentemente, será gerada uma única função discriminante se a variável apresentar apenas duas categorias (ver Figura 3) . Porém, se existirem três ou mais categorias, será necessário calcular uma função discriminante para cada par de categorias no grupo.

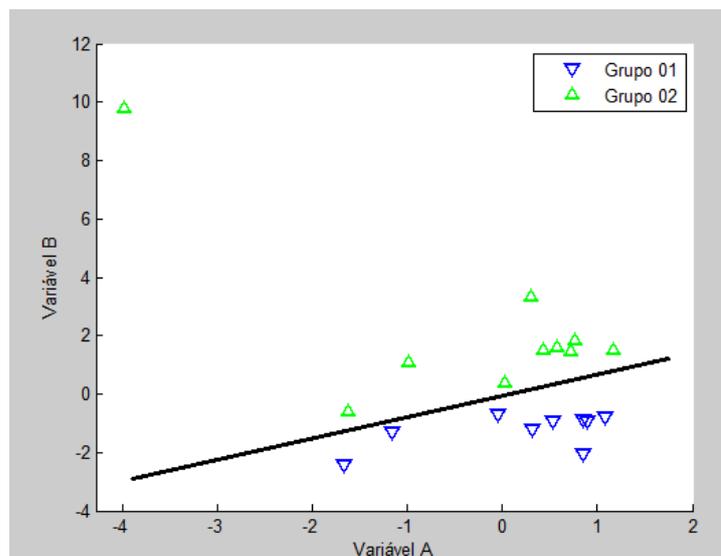


Figura 3. Representação de uma função de análise discriminante linear

A implementação da análise discriminante apresenta um conjunto de suposições a serem consideradas durante o processo de elaboração do modelo (Johnson e Wichern, 2007):

- normalidade multivariada das variáveis independentes, pois a não normalidade pode causar problemas na estimação da função;

- as variáveis, dependente e independente, devem apresentar apenas relações lineares, pois a função discriminante não absorve as relações não-lineares intrínsecas das variáveis;
- baixa correlação entre as variáveis independentes, pois a multicolinearidade pode acrescentar pouco poder explicativo à função discriminante.

Desta forma, é importante avaliar previamente se a amostra disponível para o modelo atende às suposições enumeradas para a implementação da metodologia em questão.

3.2. MÉTODOS NÃO CONVENCIONAIS: REDES NEURAIS ARTIFICIAIS

Redes neurais artificiais são comumente associadas ao sistema nervoso dos seres humanos, pois apresentam capacidade de aprender e adequar o conhecimento adquirido devido ao elevado número de conexões implementadas (Silva *et al.*, 2010).

Devido a complexidade envolvida durante sua constituição, os avanços tecnológicos auxiliam diretamente na viabilização de sua implementação. Neste sentido, Kovacs (2002) acredita que o apelo computacional é o que torna as redes neurais artificiais capazes de generalizar resultados a partir de conjuntos de dados destinados ao treinamento.

Históricamente, a primeira publicação relacionada a neurocomputação foi apresentada por McCulloch e Pitts, em 1943, a partir do qual foi proposto um modelo matemático associado ao neurônio biológico. Em 1949, Hebb, propôs o primeiro método de treinamento para as redes , fator que incentivou outros pesquisadores a prosseguirem os estudos e expandirem a aplicação das redes neurais artificiais até os dias de hoje.

Atualmente, as redes neurais artificiais são amplamente utilizadas com diversos propósitos, tais como controle de processos, segmentação de dados, otimização de sistemas, previsão de informações, entre outros.

Com relação à previsibilidade de dados, a área de finanças e economia tem empregado amplamente as redes neurais artificiais na previsão de debêntures (Curi, 2008), no cálculo de capital econômico para risco operacional (Ueno, 2010), na previsão de índices microeconômicos como o INCC – Índice Nacional de Construção Civil (Barbosa et al., 2011) e na previsão da taxa de câmbio (Santos, 2013). Porém, é também comumente utilizada em diversas outras áreas do conhecimento para previsão, tais como faturamento de lojas (Alves et al, 2013), alocação de recursos (Rosas et al, 2013), área da saúde relacionada a diabetes (Yano, 2011) e, inclusive, em eventos comportamentais (Canena, 2013).

Além disso, vale ressaltar que as redes neurais artificiais já foram empregadas para a previsão de insolvência, como apresentado por Almeida e Matias (2001), Castro Junior (2003) e Santana Carneiro (2004). Entretanto, nestes trabalhos, o método de rede neural artificial empregado foi o de retropropagação, que apresenta treinamento supervisionado, sendo esta a principal diferença entre este tipo de rede e as redes neurais artificiais auto-organizáveis.

Os fatores de sua ampla aplicação estão relacionados a suas características complementares aos demais métodos de previsão (Silva *et al.*, 2010), tais como:

- Não lineares: Durante todas as etapas associadas à construção da rede é possível identificar pontos de não linearidade;
- Capacidade de aprendizado: Os algoritmos de treinamento são constituídos de forma a aplicar de maneira distinta e exaustiva diversas combinações da amostra selecionada para minimizar o erro das estimativas, fator que se assemelha à inferência estatística não paramétrica;
- Adaptabilidade: Pode ser facilmente calibrada e ajustada, inclusive durante o funcionamento, tornando a rede estável e obtendo resultados mais robustos, por continuamente adquirir conhecimento por experimentação;
- Resposta a evidências: Pode ser utilizada para segmentar o conjunto de dados com base nas variáveis de entrada, mas também pode indicar assertividade sobre as tomadas de decisões já realizadas, pois viabilizam o agrupamento de padrões e peculiaridades da amostra;
- Informação contextual: As entradas interagem internamente na rede e otimizam os resultados através de operações matemáticas elementares;

- Tolerância a falhas: devido ao elevado número de conexões implementadas sofre pouca interferência de dados discrepantes e em casos com dados faltantes.

Segundo Hair *et al.* (2007), as redes neurais artificiais podem ser consideradas uma ferramenta estatística para regressões não lineares, o que requer atenção especial quanto a seleção da amostra durante a preparação metodológica. A partir da amostra disponível, os dados foram divididos em dois grupos: um grupo será utilizado para o treinamento durante a elaboração da rede e o outro será utilizado para avaliação do modelo desenvolvido a partir da amostra de treinamento.

As redes neurais artificiais auto-organizáveis destacam-se dentre as principais redes neurais artificiais utilizadas pois, diferentemente da maioria das redes, possuem método de treinamento não supervisionado, isto é, o treinamento é realizado apenas através do conjunto de dados de entrada. Assim, são utilizados métodos de treinamento competitivos que são capazes de detectar as similaridades, padrões e correlações do conjunto de informações disponíveis, para posteriormente classificar os registros em grupos (clusters) homogêneos de acordo com suas características (Silva *et al.*, 2010).

As redes auto-organizáveis foram propostas por Teuvo Kohonen (1982) e por isso também são conhecidas como redes de Kohonen ou SOM (self-organizing maps) e são associadas ao córtex cerebral, onde a ativação de uma região específica corresponde à resposta a um determinado estímulo sensorial.

3.3. ESTUDO COMPARATIVO DE METODOLOGIAS

As metodologias apresentadas como técnicas de previsão possuem características específicas, desta forma torna-se interessante realizar o estudo comparativo a fim de entender seus diferenciais, que serão subsídios para a definição metodológica em estudos de previsão.

Para os métodos de análise discriminante e redes neurais artificiais, foi realizada a comparação dos principais pontos envolvidos durante a construção de um modelo preditivo, permitindo evidenciar os diferenciais existentes entre as

metodologias tais como os métodos de estimação, suposições e característica das variáveis de entrada e saída, conforme observado no Quadro 1.

Quadro 1. Comparativo entre as metodologias para previsibilidade de dados

Item de comparação	Métodos convencionais (Análise discriminante)	Métodos não convencionais (Redes neurais artificiais)
Variável de saída	Uma variável categórica, sem limitação de categorias	Uma ou mais variáveis, sem restrição de classificação ou domínio
Variável de entrada	Variáveis utilizadas devem ser métricas	Variáveis utilizadas podem ser métricas ou não-métricas
Relação entre as variáveis de entrada e saída	Linear	Linear e não-linear
Métodos de estimação	Utiliza-se do método de mínimos quadrados	Utiliza-se do método de competição entre os neurônios (cálculo de distâncias)
Suposições	Normalidade multivariada das variáveis de entrada e não correlação entre as variáveis	Não há
Subjetividade na constituição	Não subjetivo	Subjetivo
São impactados pela ausência de dados	Sim, porém podem ser utilizadas técnicas específicas para minimização dos efeitos	Não
Os modelos são adaptados durante a utilização	Não	Sim

Frente as características apresentadas podemos destacar o uso das redes neurais artificiais como método alternativo devido a:

- viabilidade de relação não-linear entre as variáveis dependentes e independentes, permitindo maior flexibilidade com relação à escolha de variáveis para a previsão;
- subjetividade na constituição, por permitir agregar visão de especialistas e informações complementares às obtidas pelas variáveis utilizadas; e
- a possibilidade de realizar previsões com ausência de observações, que é considerado um fator de sensibilidade em métodos de previsão convencionais.

Vale destacar que apesar das vantagens apresentadas pelas redes neurais artificiais a adaptabilidade não implica diretamente na robustez do modelo. Desta forma, pode ser necessário fazer a seleção de um número suficiente de informações para seu treinamento e avaliar qualitativamente os resultados obtidos.

4. METODOLOGIA

Dado o objetivo principal deste trabalho, que corresponde à proposição de um modelo de redes neurais artificiais para a predição de insolvência de empresas que apresente resultados superiores aos modelos construídos baseado em metodologias estatísticas, tais como apresentados na Seção 2.3 – Metodologias usadas em previsão, a metodologia será detalhada nos subcapítulos subsequentes conforme o racional proposto por Cooper e Schindler (2003) delimitando cinco etapas, como apresentado no fluxo da Figura 4:

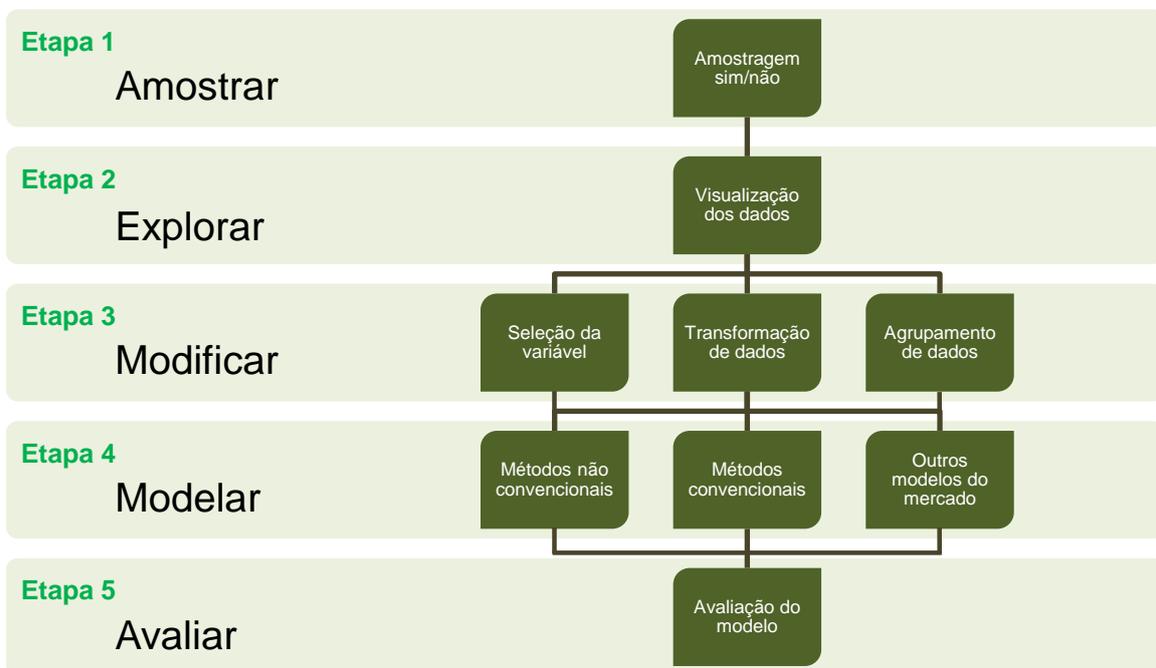


Figura 4. Fluxo metodológico de análise de dados
(Fonte: Cooper e Schindler, 2003, pp. 241)

Para o desenvolvimento do trabalho serão utilizados os softwares:

- Microsoft Excel versão 2007: Para a constituição dos índices, tratamento dos dados, consolidação das informações e elaboração de gráficos;
- Microsoft Word versão 2007: Para a consolidação dos resultados e constituição da dissertação;
- MatLab, versão 2010: Para a realização da análise descritiva, redução de dimensionalidade à elaboração de modelos de redes neurais artificiais e análise discriminante.

4.1. AMOSTRAGEM DOS DADOS

Durante a etapa de Amostrar são decididas as características da seleção da amostra do estudo em questão, inclusive a metodologia e período de seleção dos dados para estudo.

O trabalho de pesquisa versará sobre o uso de dados históricos de demonstrativos financeiros - balanço patrimonial e demonstração de resultados do exercício - de empresas de capital aberto. As empresas de capital aberto correspondem às sociedades anônimas com ações disponíveis para negociação na bolsa de valores ou mercado de balcão, sendo normalmente empresas de grande porte e de natureza comercial (Assaf, 2008).

Serão consideradas as empresas de capital aberto que possuem ou possuíram suas ações na bolsa BM&F Bovespa, que é considerada a principal instituição brasileira de intermediação para operações do mercado de capitais. Desta forma, devido ao grande número de informações disponíveis, torna-se necessário amostrar os dados da população, que atualmente corresponde a mais de 500 empresas solventes listadas na BM&F Bovespa (Site Institucional BM&F Bovespa, 2012¹).

4.1.1.METODOLOGIA DE AMOSTRAGEM

Para a seleção dos dados foi utilizada a metodologia de amostragem não probabilística, que, segundo Cooper e Schindler (2003), atende satisfatoriamente aos objetivos de amostragem, além de apresentar como vantagens o custo e a otimização do tempo.

A amostragem não probabilística apresenta diversos métodos de seleção de dados, porém será utilizado o método denominado amostragem intencional por quotas. Este método é normalmente utilizado para melhorar a representatividade dos elementos na amostra, possibilitando que certas características relevantes sejam identificadas na amostra, viabilizando representar uma parcela da população

¹ Acesso em 15/09/2012.

desejada e ser pertinente ao estudo em questão, neste caso a representatividade de empresas solventes e insolventes.

Neste sentido, a fim de evitar interferência no número de empresas solventes sobre empresas insolventes, serão selecionadas vinte empresas de cada situação econômica financeira, dado que a representatividade de empresas solventes de capital aberto é muito superior ao número de empresas denominadas insolventes, independentemente do segmento de atuação.

O período de corte para a amostra em questão será os anos de 2004 a 2012, a fim de minimizar potenciais alterações nos demonstrativos financeiros disponibilizados pelas empresas devido à mudança no panorama econômico, além de permitir maior diversidade durante a seleção de empresas na categoria insolvente.

A identificação de insolvência de empresas será realizada pelos relatórios de inadimplência, divulgados pela Bovespa, que identifica as empresas que se encontram em recuperação judicial ou extrajudicial, assim como empresas que decretaram falência no período estabelecido.

As empresas insolventes serão selecionadas aleatoriamente conforme divulgação nos últimos relatórios de inadimplência. Posteriormente, serão obtidos os dados históricos dos demonstrativos financeiros dos últimos cinco anos antes da declaração de falência. Portanto, se a falência ou recuperação judicial foi decretada em 2012, serão obtidos dados de 2011 (denominado ano 5), 2010 (denominado ano 4), 2009 (denominado ano 3), 2008 (denominado ano 2) e 2007 (denominado ano 1).

O uso de cinco anos de base histórica visa avaliar a eficiência de previsões em até cinco anos antes de declaração de falência, verificando a eficácia do modelo de predição ao longo do tempo e a estabilidade do modelo ao longo do período de análise.

Com relação às empresas solventes, serão selecionadas, aleatoriamente, vinte empresas de elevada representatividade e de diferentes segmentos do mercado brasileiro de ações com o intuito de evitar priorizar uma empresa que possa torna-se insolvente em um curto espaço de tempo.

Assim, obtemos uma amostra com quarenta empresas, vinte empresas insolventes (ver Quadro 2) e vinte empresas solventes (ver Quadro 3), nas quais serão destacados os anos de captura dos dados de cada empresa.

Quadro 2. Empresas insolventes amostradas para o estudo

Código	Empresa	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5
AGEN	Agrenco S.A.	2004	2005	2006	2007	2008
BOBR	Bombril S.A.	2001	2002	2003	2004	2005
BUET	Buettner S.A.	2004	2005	2006	2007	2008
CCHI	Cerâmica Chiarelli S.A.	2005	2006	2007	2008	2009
CELP	Centrais Elet Do Para S.A. - Celpa	2007	2008	2009	2010	2011
COBE	Construtora Beter S.A.	2006	2007	2008	2009	2010
EUCA	Eucatex S.A.	2000	2001	2002	2003	2004
FTRX	Fabrica Tecidos Carlos Renaux S.A.	2007	2008	2009	2010	2011
IGBR	IGB Eletrônica S.A.	2004	2005	2006	2007	2008
LCSA	Parmalat Brasil S.A	1999	2000	2001	2002	2003
RCSL	Recrosul S.A.	2001	2002	2003	2004	2005
REDE	REDE ENERGIA S.A.	2007	2008	2009	2010	2011
SCLO	Cia Industrial Schlosser S.A.	2006	2007	2008	2009	2010
SJOS	Fiação Tec São Jose S.A.	2005	2006	2007	2008	2009
SNSY	Sansuy S.A.	1999	2002	2003	2004	2005
STRP	Botucatu Têxtil S.A.	2005	2006	2007	2008	2009
TCNO	Tecnosolo Engenharia S.A.	2005	2006	2007	2008	2009
TEKA	Teka-Tecelagem Kuehnrich S.A.	2007	2008	2009	2010	2011
VAGV	Varig Transportes Aéreos S.A.	2000	2001	2002	2003	2004
VPTA	VASP - Viação Aérea São Paulo S.A.	1999	2000	2001	2002	2003

Quadro 3. Empresas solventes amostradas para o estudo

Código	Empresa	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5
ALLL	All America Latina Logística S.A.	2007	2008	2009	2010	2011
AMBV	Ambev S.A.	2007	2008	2009	2010	2011
BRKM	Braskem S.A.	2007	2008	2009	2010	2011
CESP	CESP S.A.	2007	2008	2009	2010	2011
CRUZ	Souza Cruz S.A.	2007	2008	2009	2010	2011
CYRE	Cyrela Brazil Realty S.A.	2007	2008	2009	2010	2011
DASA	Diagnósticos Da America S.A.	2007	2008	2009	2010	2011
DTEX	Duratex S.A.	2007	2008	2009	2010	2011
ELPL	Eletropaulo Metrop. São Paulo S.A.	2007	2008	2009	2010	2011
FIBR	Fibria Celulose S.A.	2007	2008	2009	2010	2011
GGBR	Gerdau S.A.	2007	2008	2009	2010	2011
GOLL	Gol Linhas Aéreas Inteligentes S.A.	2007	2008	2009	2010	2011
HGTX	Cia Hering S.A.	2007	2008	2009	2010	2011
LREN	Lojas Renner S.A.	2007	2008	2009	2010	2011
MMXM	Mmx Mineração E Metálicos S.A.	2007	2008	2009	2010	2011
NATU	Natura Cosméticos S.A.	2007	2008	2009	2010	2011
PCAR	Cia Brasileira De Distribuição S.A.	2007	2008	2009	2010	2011
PETR	Petrobras - Petróleo Brasileiro S.A.	2007	2008	2009	2010	2011
USIM	Usinas Sid De Minas Gerais S.A.	2007	2008	2009	2010	2011
VALE	Vale S.A.	2007	2008	2009	2010	2011

4.2. EXPLORAÇÃO DOS DADOS

Nesta etapa é identificada a relação entre os dados e caracterizada a amostra em questão de acordo com as variáveis consideradas durante o estudo. As informações a serem exploradas serão baseadas na lista de empresas amostradas (Quadro 2 e 3) e nos demonstrativos financeiros das quarenta empresas para os cinco anos estabelecidos.

Adicionalmente, serão avaliadas as principais características das variáveis selecionadas e suas inter-relações para entender o impacto para o estudo em questão.

4.2.1. DEFINIÇÃO DE VARIÁVEIS

Para o estudo em questão, serão consideradas as variáveis previamente utilizadas em trabalhos de previsão de insolvência baseados em métodos estatísticos, sendo um pré-requisito a utilização de dados do balanço patrimonial (BP) e do demonstrativo do resultado do exercício (DRE) para a elaboração do modelo, uma vez que métodos estatísticos já são empregados para avaliação de problemas em Instituições financeiras (Oliveira, 2010).

Com base nas condições apresentadas, foram selecionados aleatoriamente cinco modelos que se utilizaram de técnicas estatísticas de análise discriminante para previsão de insolvência em diferentes períodos: Altman (1968), Kanitz (1978), Sanvicente (1998), Silva (2001) e Guimarães (2008).

Para cada modelo mencionado foram detalhados os modelos propostos, as variáveis consideradas e a equação resultante do trabalho em questão, conforme apresentado no Quadro 4.

Através dos modelos descritos, foram observados 24 índices, todos construídos a partir de dados dos demonstrativos financeiros. Para auxiliar na interpretação dos índices apresentados, o conceito das variáveis contábeis consideradas nos 24 índices foi detalhado no Apêndice A.

Quadro 4. Modelos desenvolvidos para predição utilizando análise discriminante

Autor (ano)	Modelo proposto	Variáveis utilizadas	Análise do modelo
Kanitz (1978)	$A = 0,05 \times A_1 + 1,65 \times A_2 + 3,55 \times A_3 + 1,06 \times A_4 - 0,33 \times A_5$	<p>A₁: Lucro Líquido / Patrimônio Líquido</p> <p>A₂: (Ativo circulante + Realizável a LP) / Passivo circulante + Exigível a LP)</p> <p>A₃: (Ativo circulante – Estoques) / Exigível a CP</p> <p>A₄: Ativo circulante / Passivo circulante</p> <p>A₅: Exigível total / Patrimônio Líquido</p>	<p>A < -3: Indicativo de falência ²</p> <p>A > -3 e A < 0: Intervalo de incerteza</p> <p>A > 0: Indicativo de baixa probabilidade de falência</p>
Altman <i>et al.</i> (1979)	$B = -1,44 + 4,03 \times B_1 + 2,25 \times B_2 + 0,14 \times B_3 + 0,42 \times B_4$	<p>B₁: Lucros acumulados / Ativo total</p> <p>B₂: Lucro operacional antes de juros e IR / Ativo total</p> <p>B₃: Patrimônio líquido / exigível total</p> <p>B₄: Vendas líquidas / Ativo total</p>	<p>B > 0: Indicativo de solvência</p> <p>B < 0: Indicativo de insolvência</p>

² Excepcionalmente, neste modelo proposto, a previsão não detecta a insolvência que, conforme apresentado na seção 2.1, é caracterizada pela falência ou recuperação judicial. A recuperação judicial pode ser considerada um estágio preliminar à falência, a fim de remediar a situação da falência.

Quadro 4. Modelos desenvolvidos para predição utilizando análise discriminante (continuação)

Autor (ano)	Modelo proposto	Variáveis utilizadas	Análise do modelo
Sanvincente e Minardi (1998)	$C = -0,042 - 0,209 \times C_1 - 0,875 \times C_2 + 3,636 \times C_3 + 0,172 \times C_4 + 0,029 \times C_5$	<p>C₁: (Ativo circulante – Passivo total) / Ativo total</p> <p>C₂: (Patrimônio líquido – Capital Social) / Ativo total</p> <p>C₃: (Lucro operacional – Despesas finan.+ Receitas finan.) / Ativo total</p> <p>C₄: Patrimônio líquido / (Exigível total)</p> <p>C₅: Lucro operacional antes de juros e IR / Despesas financeiras</p>	<p>C > 0: Indicativo de solvência</p> <p>C < 0: Indicativo de insolvência</p>
Silva (2001)	$D = 0,722 - 5,124 \times D_1 + 11,016 \times D_2 - 0,342 \times D_3 - 0,048 \times D_4 + 8,605 \times D_5 - 0,004 \times D_6$	<p>D₁: Duplicatas descontadas / Duplicatas a receber</p> <p>D₂: Estoques/ Custo da mercadoria vendida</p> <p>D₃: Fornecedores / Vendas</p> <p>D₄: Estoque / Custo da mercadoria vendida x 360</p> <p>D₅: (Lucro operacional + Despesas finan.) / (Ativo total – Investimentos)</p> <p>D₆: Exigível total / (Lucro líquido + 0,1 x Imobilizado)</p>	<p>D > 0: Indicativo de solvência</p> <p>D < 0: Indicativo de insolvência</p>
Guimarães e Moreira (2008)	$E = 0,526 + 1,6791 \times E_1 - 1,8301 \times E_2 + 0,012 \times E_3 + 2,0812 \times E_4$	<p>E₁: Capital de giro líquido / Ativo total</p> <p>E₂: (Passivo Circulante + Exigível total) / (Passivo total)²</p> <p>E₃: ((Patrimônio líquido – Ativos de baixa liquidez)^{1/3})/100</p> <p>E₄: Lucro operacional antes de juros e IR / Passivo corrente + exigível a LP</p>	<p>E > 0: Indicativo de solvência</p> <p>E < 0: Indicativo de insolvência</p>

4.2.2. ANÁLISE DESCRITIVA

Baseado nas variáveis utilizadas, mencionadas no Quadro 4 (total de 24 índices), será realizada a análise descritiva destes índices, sendo agrupada pela sua situação financeira (solvente ou insolvente). Durante a avaliação, pretende-se observar seu comportamento através de medidas de posição e de dispersão, tais como:

- Média: corresponde ao valor esperado para um conjunto de dados;
- Mediana: corresponde a uma medida de posição central para os dados ordenados, dividindo a amostra em 50% com valores superiores e 50% com valores inferiores ao valor da mediana;
- Desvio padrão: corresponde a medida de dispersão dos dados em relação ao valor médio;
- Amplitude dos dados: corresponde à diferença entre o valor máximo e mínimo observado da variável em questão;
- Assimetria observada: corresponde ao grau de afastamento da distribuição com relação ao seu eixo central.

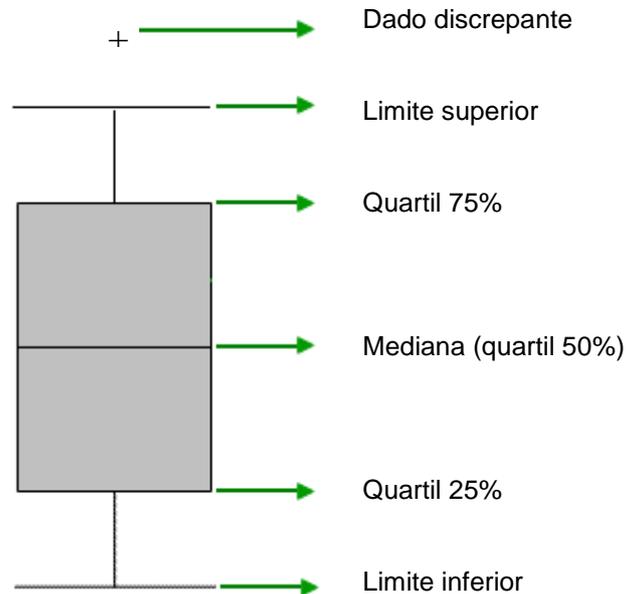
Além das medidas apresentadas, os dados podem ser sumarizados graficamente através do *boxplot*, que corresponde a um desenho esquemático que permite analisar a dispersão dos dados, assimetria e dados discrepantes. O *boxplot* é composto pelas informações mediana, quartil 25% (valor que corresponde a 25% da amostra ordenada), quartil 75% (valor que corresponde a 75% da amostra ordenada) e os limites superior e inferior conforme apresentado na Figura 5 (Bussab e Morettin, 2004).

O limite superior e o inferior são calculados considerando o intervalo interquartil, que corresponde à diferença entre quartil 75% e quartil 25%. Portanto, os valores que não estiverem entre os limites, inferior e superior, serão considerados dados discrepantes.

$$\text{Limite superior: } \text{Quartil } 75\% + 1,5 * (\text{Quartil } 75\% - \text{Quartil } 25\%)$$

$$\text{Limite inferior: } \text{Quartil } 25\% - 1,5 * (\text{Quartil } 75\% - \text{Quartil } 25\%) \quad (2)$$

Figura 5. Descrição da construção de gráfico *Boxplot*



Para complementar a análise descritiva realizada, serão construídas matrizes de correlações entre as variáveis, uma vez que a matriz permite avaliar a dependência entre as variáveis. Vale ressaltar que a matriz corresponde a um dos pré-requisitos a ser respeitado para utilização da metodologia da análise discriminante, de forma que esta não apresente variáveis com alta correlação.

A correlação é calculada para cada par de variáveis conforme apresentado em (3), sendo que os valores são limitados entre -1 e 1. Durante a interpretação dos resultados é importante avaliar que quanto mais próxima a correlação for de 1, indica maior correlação linear positiva entre as variáveis, isto é, quando uma variável aumenta a outra também aumenta. Por outro lado quanto mais próxima a correlação for de -1, indica maior correlação linear negativa entre as variáveis, isto é, quando uma variável aumenta a outra diminui. Por fim, quando mais próxima a correlação for de 0 indica ausência de correlação linear entre as variáveis. No entanto, é possível que exista uma relação não linear entre as variáveis.

$$\text{corr}(x, y) = \rho_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

4.3. MODIFICAÇÃO DOS DADOS

Na etapa Modificar, são transformados os dados em índices ou em novas variáveis a fim de otimizar e minimizar potenciais desvios causados pelas variáveis consideradas. A modificação torna-se necessária em estudos com elevado número de variáveis disponíveis e a existência de correlação entre as variáveis.

Assim, a análise de componentes principais, corresponde a uma possível técnica a ser utilizada para a redução de dimensionalidade, pois transforma as variáveis originais em novas variáveis, formadas pela combinação linear das variáveis originais e retirando a dependência previamente existente.

4.3.1. ANÁLISE DE COMPONENTES PRINCIPAIS

Segundo Haykin (2001), a análise de componentes principais constituiu-se como uma ferramenta efetiva para o reconhecimento estatístico de padrões. Ao passo que Jolliffe (2002) explica que corresponde a uma técnica de análise multivariada que tem sido utilizada em diversas áreas do conhecimento, tais como: biologia, química, economia, ecologia, psicologia, controle de qualidade e conclui que pode ser empregada facilmente em qualquer outra área ainda não explorada.

O valor prático da análise de componentes principais é fornecer uma técnica efetiva para redução de dimensionalidade, uma vez que podemos reduzir o número de variáveis necessárias retendo apenas as variáveis que apresentam variâncias grandes (Haykin, 2001, pp. 433). Além disso, auxilia na identificação de relação entre as variáveis não identificadas inicialmente e auxilia na interpretação que não pode ser normalmente realizada devido ao grande número de variáveis (Johnson e Wichern, 2007).

Matematicamente, a técnica de componentes principais é definida como uma transformação linear ortogonal que transforma os dados em novas coordenadas (Jolliffe, 2002). São pré-requisitos para a utilização da ferramenta que as variáveis sejam quantitativas (dados apenas numéricos) e recomendadas que o problema a ser estudado apresente uma dimensão elevada de variáveis.

Para iniciar a transformação dos dados, é necessário que todas as variáveis apresentem média zero. Para isso, é necessário subtrair a média observada de cada uma das variáveis ou padronizar todas as variáveis, padronização esta encontrada pela subtração da média dos valores da amostra e pela divisão do desvio padrão da amostra:

$$z_{pad} = \frac{(x - \mu_x)}{\sigma_x} \quad (4)$$

Onde μ corresponde à média dos valores observados da variável em questão, σ ao desvio padrão dos valores observados para a variável em questão, x à variável a ser padronizada.

Uma vez que as variáveis foram padronizadas, é possível calcular a covariância entre as variáveis, que corresponde a uma medida de tendência linear entre duas variáveis. Covariância com valores positivos indica que as variáveis variam conjuntamente na mesma direção, enquanto valores negativos indicam que as variáveis variam em sentidos opostos. Além disso, quanto maior os valores observado em módulo, mais forte é a relação entre as variáveis.

$$\text{cov}(x_1, x_2) = \frac{\sum_{i=1}^j (x_{1,i} - \bar{x}_1) \times (x_{2,i} - \bar{x}_2)}{(n-1)} \quad (5)$$

Desta forma, é possível construir a matriz de covariância que avalia a relação entre todas as variáveis utilizadas, calculada para cada combinação de variáveis, gerando uma matriz simétrica.

$$COV = \begin{pmatrix} \text{cov}(x_1, x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_n) \\ \text{cov}(x_2, x_1) & \text{cov}(x_2, x_2) & \dots & \text{cov}(x_2, x_n) \\ \dots & \dots & \dots & \dots \\ \text{cov}(x_n, x_1) & \text{cov}(x_n, x_2) & \dots & \text{cov}(x_n, x_n) \end{pmatrix} \quad (6)$$

Onde n corresponde a n -ésima variável considerada no estudo e a covariância entre a mesma variável corresponde a variância desta variável.

A partir da matriz de covariância, é possível calcular os autovalores, que apresentarão a mesma dimensão do número de variáveis. Desta forma, será gerado

um conjunto de números λ_i tal que, se for subtraído de cada entrada da diagonal da matriz de covariância, converte a matriz em uma matriz singular, que por definição apresenta determinante nulo:

$$\det(COV_{n \times n} - \lambda \times I_n) = 0 \quad (7)$$

A soma dos autovalores é igual à soma dos termos da diagonal da matriz de covariância, que corresponde à variância de cada uma das variáveis, sendo que a contribuição de cada autovalor pode ser expressa em porcentagem, dividindo a variância de cada componente pela variância total.

Assim, é possível avaliar a representatividade da variância que cada autovalor representa e definir a nova dimensão a ser considerada para o estudo em questão, podendo desconsiderar aquelas com baixa significância:

$$representatividade = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad (8)$$

Definida a nova dimensão, é necessário calcular as componentes principais, que serão obtidas a partir da multiplicação dos autovetores dos autovalores selecionados pela base de dados original transposta. Os autovetores são calculados individualmente a partir do autovalor selecionado:

$$(COV_{n \times n} - \lambda_i \times I_n) \times \hat{v}_i = 0 \quad (9)$$

Onde $\hat{v}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,n})$ e corresponde ao autovetor de dimensão n, λ_i corresponde ao *i-ésimo* autovalor e I_n a matriz identidade de ordem n.

Cada componente principal é apresentada em ordem de representatividade de variância. Desta forma, as primeiras componentes apresentam a maior variação presente das variáveis originais em relação à variação total (Jolliffe, 2002). Assim sendo, as componentes principais geradas, correspondem a uma combinação linear das variáveis originais, porém com uma dimensão reduzida em relação à dimensão original, uma vez que as componentes principais foram priorizadas (Ver Quadro 5).

Quadro 5. Representação das componentes principais

Elemento	Componente principal
1	$CP_1 = v_{1,1}X_{1,1} + v_{1,2}X_{2,1} + \dots + v_{1,n}X_{n,1}$
2	$CP_2 = v_{2,1}X_{1,2} + v_{2,2}X_{2,2} + \dots + v_{2,n}X_{n,2}$
...
k	$CP_k = v_{k,1}X_{1,k} + v_{k,2}X_{2,k} + \dots + v_{k,n}X_{n,k}$

4.4. MODELAGEM DOS DADOS

Corresponde à etapa de desenvolvimento do modelo com base no objetivo em questão. Neste sentido, é necessária a especificação da metodologia a ser adotada e as características a serem consideradas durante a construção do modelo.

Para viabilizar a comparação de assertividade entre as metodologias utilizadas para predição de insolvência, a partir das variáveis transformadas utilizando-se da técnica de componentes principais, serão desenvolvidos dois modelos, um pela metodologia de métodos convencionais (análise discriminante) e outro de métodos não convencionais (redes neurais artificiais auto-organizáveis).

4.4.1. DESENVOLVIMENTO

A fim de permitir a comparação dos resultados preditivos dos modelos desenvolvidos, a elaboração dos modelos será baseada em 75% da amostra, a fim de utilizar os 25% restantes para testes de verificação do modelo construído. Para isto, foram selecionadas, aleatoriamente, 30 empresas para o treinamento, sendo 15 solventes e 15 insolventes, mantendo as 10 empresas restantes da amostra para validação independente dos resultados. As empresas selecionadas para treinamento encontram-se no Quadro 6 e as empresas selecionadas para validação, encontram-se no Quadro 7.

Quadro 6. Empresas selecionadas para validação

Solventes		Insolventes	
Código	Empresa	Código	Empresa
AMBV	Ambev S.A.	CCHI	Ceramica Chiarelli S.A.
CRUZ	Souza Cruz S.A.	CELP	Centrais Elet Do Para S.A. - Celpa
DTEX	Duratex S.A.	LCSA	Parmalat Brasil S.A
NATU	Natura Cosméticos S.A.	REDE	REDE ENERGIA S.A.
VALE	Vale S.A.	VPTA	VASP - Viação Aérea São Paulo S.A.

Quadro 7. Empresas selecionadas para treinamento

Solventes		Insolventes	
Código	Empresa	Código	Empresa
ALLL	All America Latina Logística S.A.	AGEN	Agrenco S.A.
BRKM	Braskem S.A.	BOBR	Bombriil S.A.
CESP	CESP S.A.	BUET	Buettner S.A.
CYRE	Cyrela Brazil Realty S.A.	COBE	Construtora Beter S.A.
DASA	Diagnósticos Da America S.A.	EUCA	Eucatex S.A.
ELPL	Eletropaulo Metrop. Sao Paulo S.A.	FTRX	Fabrica Tecidos Carlos Renaux S.A.
FIBR	Fibria Celulose S.A.	IGBR	IGB Eletrônica S.A.
GGBR	Gerdau S.A.	RCSL	Recrosul S.A.
GOLL	Gol Linhas Aereas Inteligentes S.A.	SCLO	Cia Industrial Schlosser S.A.
HGTX	Cia Hering S.A.	SJOS	Fiação Tec São Jose S.A.
LREN	Lojas Renner S.A.	SNSY	Sansuy S.A.
MMXM	Mmx Mineração E Metálicos S.A.	STRP	Botucatu Têxtil S.A.
PCAR	Cia Brasileira De Distribuição S.A.	TCNO	Tecnosolo Engenharia S.A.
PETR	Petrobras - Petróleo Brasileiro S.A.	TEKA	Teka-Tecelagem Kuehnrich S.A.
USIM	Usinas Sid De Minas Gerais S.A.	VAGV	Varig Transportes Aéreos S.A.

4.4.2. ANÁLISE DISCRIMINANTE

O resultado da análise discriminante é popularmente denotado como função discriminante ou função linear de Fisher, sendo gerada a partir da combinação linear das variáveis originais, com o objetivo de maximizar a variância entre os grupos e minimizar a variância dentro dos grupos (Hair *et al.*, 2007).

Para o caso de duas populações, o cálculo da função discriminante parte da média das k variáveis disponíveis calculadas por grupo, sendo necessário também calcular a diferença entre as médias populacionais.

$$\begin{aligned} \mu_{grupoA} &= \overline{(x_1, x_2, \dots, x_k)} \\ \mu_{grupoB} &= \overline{(x_1, x_2, \dots, x_k)} \\ \mu &= \overline{(\mu_{grupoA} - \mu_{grupoB})} \end{aligned} \quad (10)$$

Na sequencia, também são calculadas as matrizes de covariância para cada grupo, sendo necessário o calculo da covariância conjunta entre todos os grupos (representado por g), pois um dos pré-requisitos para a realização da análise discriminante é que as matrizes de covariâncias sejam homogêneas, devendo ser considerado o número de elementos em cada grupo (representado por n).

$$COV_{geral} = \frac{\sum_{i=1}^g (n_i - 1) \cdot COV_{g_i}}{\sum_{i=1}^g (n_i - 1)} \quad (11)$$

Assim, a equação será obtida a partir das diferenças entre as médias e a matriz de covariância geral.

$$y_{discr} = \overline{(\mu_{grupoA} - \mu_{grupoB})} \cdot COV_{geral}^{-1} \cdot (x_1, x_2, \dots, x_k) \quad (12)$$

Para fins de utilização da função discriminante convencionou-se adicionar a constante na equação, desta forma a regra de alocação dos elementos torna-se mais clara.

$$cte = \frac{1}{2} \overline{(\mu_{grupoA} - \mu_{grupoB})} \cdot COV_{geral}^{-1} \cdot \overline{(\mu_{grupoA} + \mu_{grupoB})} \quad (13)$$

Portanto, quando aplicado as informações observadas de um determinado elemento na equação discriminante e resultar em valores positivos, deve-se classificar o elemento como pertencente ao grupo A, e quando resultar em valores negativos, deve-se classificar o elemento como pertencente ao grupo B.

$$\begin{aligned} y_{discr}'' &= \overline{(\mu_{grupoA} - \mu_{grupoB})} \cdot COV_{geral}^{-1} \cdot (x_1, x_2, \dots, x_k) - cte \\ \Rightarrow &\begin{cases} se \ y_{discr}'' \geq 0 \longrightarrow alocar \ no \ grupo \ A \\ se \ y_{discr}'' < 0 \longrightarrow alocar \ no \ grupo \ B \end{cases} \end{aligned} \quad (14)$$

Vale ressaltar que no caso de três ou mais populações serão geradas Z equações, sendo que Z será o valor mínimo entre o total de variáveis e o número de categorias subtraído em uma unidade.

4.4.3. REDES NEURAIS ARTIFICIAIS AUTO-ORGANIZÁVEIS

As rede neurais auto-organizáveis possuem um método de aprendizado não supervisionado, que possui como principio básico a concorrência entre os neurônios, uma vez que não apresenta saída desejada (Haykin, 2001).

A forma de competição entre os neurônios corresponde à proximidade entre o elemento de entrada e cada um dos neurônios envolvidos, normalmente realizado através da norma euclidiana do vetor de entrada e cada um dos vetores de peso dos neurônios, sendo priorizado sempre o neurônio que resultar na menor distância:

$$dist_j = \sqrt{\sum_{i=1}^n (x_i - w_i)^2} \quad (15)$$

Onde j corresponde ao j -ésimo neurônio existente, x corresponde aos pesos do elemento de entrada, w corresponde aos pesos do neurônio j e i cada uma das variáveis utilizadas para classificação. É necessário ressaltar que a posição inicial dos neurônios é definida de acordo com uma função de topologia estabelecida, por exemplo, uma grade (ou malha) retangular, hexagonal ou randômica (Figura 6).

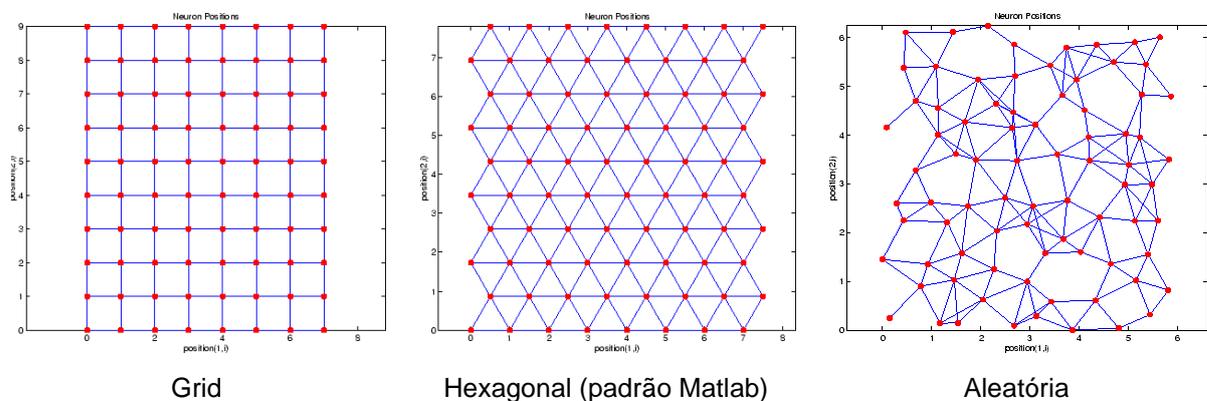


Figura 6. Topologias existentes para redes auto-organizáveis

O neurônio vencedor da competição terá seus pesos ajustados, proporcionalmente aos valores do padrão de entrada apresentado, buscando aperfeiçoar o seu estado para a próxima competição:

$$w_{atual}^{(v)} = w_{anterior}^{(v)} + \eta * (x^{(k)} - w_{anterior}^{(v)}) \quad (16)$$

Onde x corresponde aos pesos do elemento de entrada, w corresponde aos pesos do neurônio vencedor e η define a taxa de aprendizagem.

Desta forma, após a convergência do algoritmo cada um dos neurônios estará posicionado no centro dos grupos (clusters) que possuam características semelhantes entre si, porém a convergência torna-se mais eficiente uma vez que as variáveis de entradas sejam independentes entre si.

Segundo Silva *et al.* (2010, pp. 237), a dinâmica de treinamento das redes SOM pode ser sistematizada em cinco passos:

- I. Cada neurônio da rede computa o nível de proximidade de seu vetor de pesos em relação ao padrão de entrada;
- II. Um mecanismo de competição entre os neurônios é aplicado com o objetivo de escolher o vencedor;
- III. A partir da definição do neurônio vencedor, resgata-se, por meio do mapa topológico, o conjunto de vizinhança que informa sobre quais são os neurônios vizinhos ao vencedor.
- IV. Os pesos do neurônio vencedor e de seus vizinhos são incrementados com o objetivo de aumentar o nível de proximidade com a respectiva entrada;
- V. Após a convergência, é possível identificar regiões do mapa que correspondem às classes do problema.

Assim sendo, a Figura 7 representa a arquitetura de um mapa auto-organizável, porém não existe a utilização de “bias”, normalmente associado à camada de entrada para proporcionar uma melhor adaptação dos dados:

Onde P corresponde ao vetor de entrada, IW corresponde ao vetor peso, N_i corresponde ao módulo de $IW - P$ e, por fim, C representa a função de transferência competitiva que, como descrito anteriormente, retorna o vetor de saída A com o valor 1 para o neurônio vencedor e 0 para os demais.

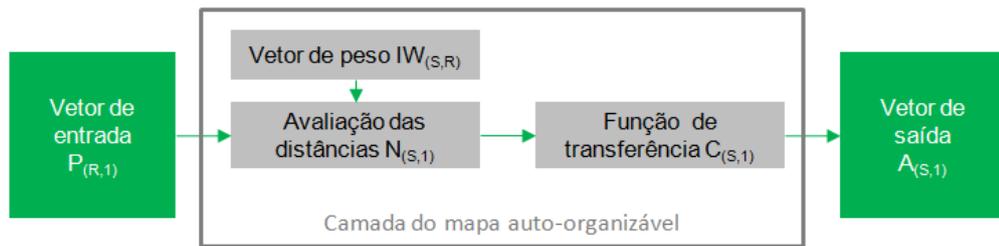


Figura 7. Arquitetura de uma rede auto-organizável adaptado de Barbosa *et al.* (2010, pp. 17).

Assim, Kohonen (1982) indica que as redes auto-organizáveis constituem-se como uma ferramenta efetiva para a classificação de dados, devido a duas características principais: visualização (consolidação das informações disponibilizadas de forma visual) e abstração (identificar relações dificilmente observadas durante a análise dos dados, devido ao grande número de variáveis envolvidas).

4.5. VALIDAÇÃO DE MODELOS

Corresponde à etapa de avaliação dos modelos desenvolvidos com relação a sua assertividade das predições, fator que viabilizar a comparação da eficiência entre os modelos.

A comparação de assertividade entre as metodologias utilizadas para predição de insolvência será realizada em duas etapas:

- Etapa 1: avaliação dos resultados considerando apenas os dados utilizados para a elaboração dos modelos, denominada amostra de treinamento (total de 30 empresas);
- Etapa 2: avaliação dos resultados considerando apenas os dados não utilizados durante o desenvolvimento dos modelos, denominada amostra de validação (total de 10 empresas).

Adicionalmente, em cada etapa, será realizada a comparação entre os tipos de erro. Como citado por Bussab e Morettin (2004), é possível cometer erros denominados Erros Tipo 1 e Erro Tipo 2, sendo que a denominação dos erros estão

associados a classificação incorreta do elemento quando comparado ao seu resultado real esperado, conforme apresentado no Quadro 8.

Quadro 8. Classificação dos tipos de erros existentes

Classificação real	Classificação estimada	
	Solvente	Insolvente
Solvente	Sem erro	Erro tipo 2
Insolvente	Erro tipo 1	Sem erro

Durante a análise dos resultados é esperado que os Erros do Tipo 1 e do Tipo 2 apresentem resultados próximos, pois o desbalanceamento entre os erros pode resultar em uma classificação tendenciosa para alguma das categorias (Bussab e Morettin, 2004), fator este que pode indicar que o modelo não está calibrado, ou seja, que está classificando incorretamente os elementos.

Por fim, será avaliada a assertividade geral de cada modelo, isto é, a proporção de acerto, considerando os elementos solventes e insolventes classificados corretamente com relação às amostras utilizadas.

5. ANÁLISE DOS RESULTADOS

5.1. ANÁLISE DESCRITIVA

Durante a análise dos resultados, para facilitar a avaliação dos *boxplots* desenvolvidos, será apresentado um gráfico por grupo de características. Para a visualização detalhada dos *boxplots* elaborados para as 24 variáveis, sugere-se observar o Anexo 1.

As variáveis serão descritas pelos códigos conforme apresentado previamente no Quadro 4 na Seção 3.2.1 - Definição de variáveis, na qual foi definida sua forma de cálculo. Além disso, o eixo X apresenta a situação da empresa: Insolvente (ins.) e solvente (solv.) e o ano de análise da variável de Ano 5 (ano mais recente) até o ano 01.

Deste modo, através dos *boxplots* elaborados, pode-se observar que para as variáveis B3, B4, C1, C4, D2, D3, D4 e D6 os valores observados para as empresas solventes e insolventes são próximos, sendo que ao longo do tempo a diferença entre as variáveis torna-se mais perceptível, conforme observado na Figura 8.

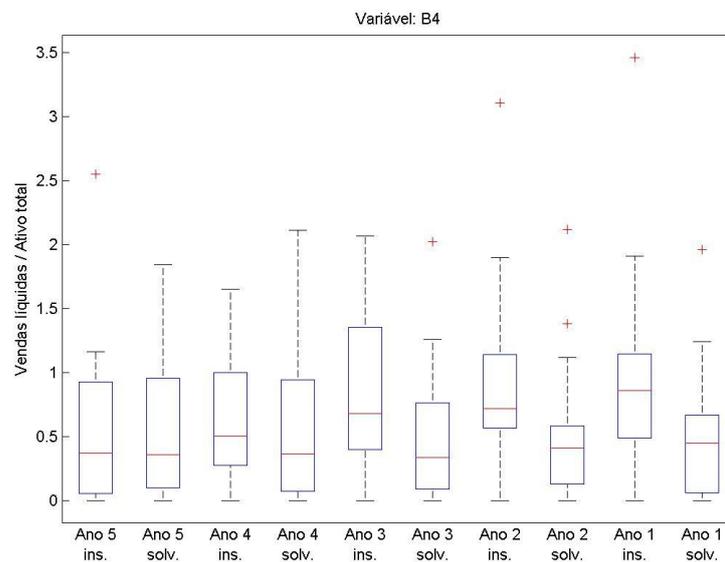


Figura 8. *Boxplot* referente à variável B4.

Para as variáveis A1, A5, C2, C3, D5 e E1 verificam-se que as empresas insolventes apresentam maior amplitude para os valores observados e normalmente

valores inferiores quando comparado com as empresas solventes, conforme observado na Figura 9.

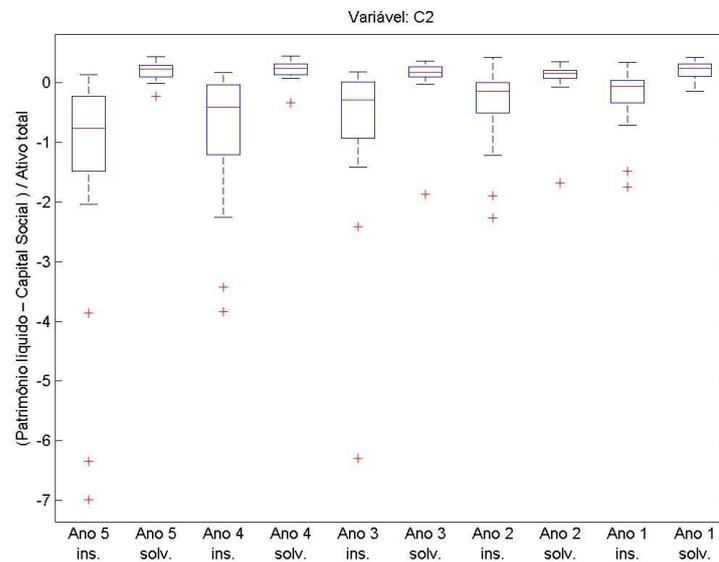


Figura 9. *Boxplot* referente à variável C2

Já para as variáveis A2, A3, A4, B1 e E4, observa-se que as empresas solventes apresentam maior amplitude para os valores observados e normalmente valores superiores quando comparado com as empresas insolventes, conforme observado na Figura 10.

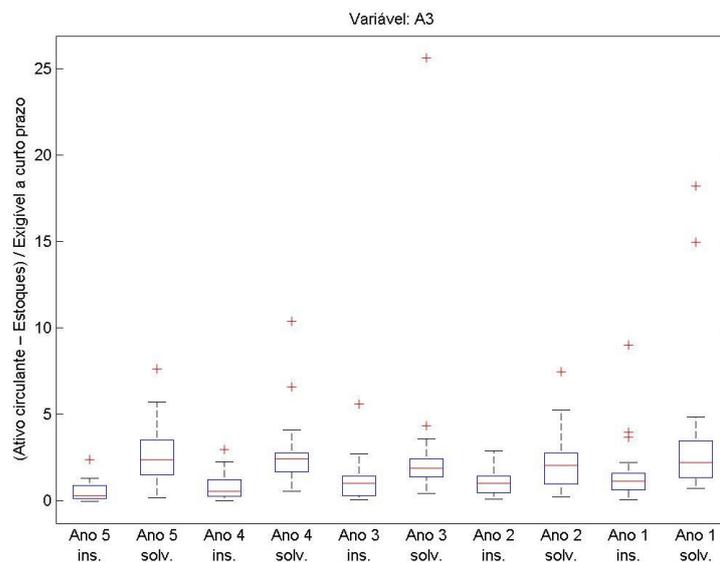


Figura 10. *Boxplot* referente à variável A3.

Para a variável B2 nota-se pequena variação na amplitude entre as variáveis para empresas solventes e insolventes e, apesar de empresas solventes

apresentarem valores superiores as insolventes, a diferença se reduz ao longo do tempo, conforme Figura 11.

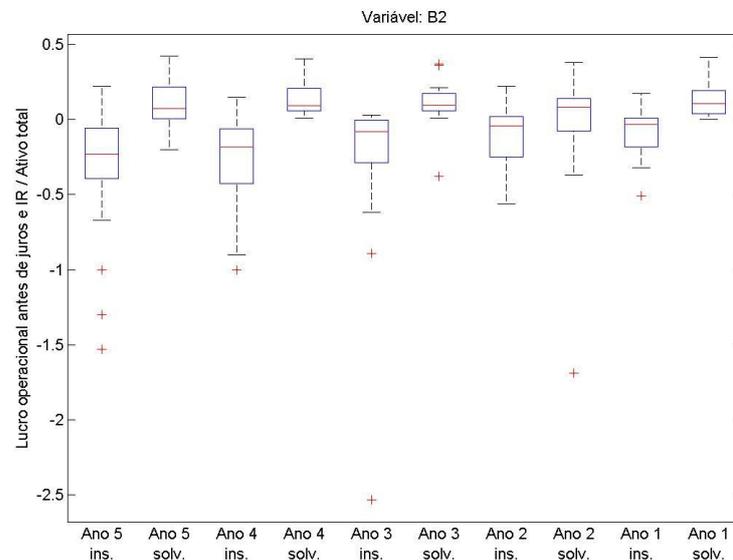


Figura 11. *Boxplot* referente à variável B2.

Para a variável E2, verifica-se que as empresas insolventes apresentam maior amplitude nos dados quando comparada as empresas solventes, sendo que a diferença se reduz ao longo do tempo, enquanto a variável E3 apresenta comportamento oposto.

Vale destacar que as variáveis C5 e D1 apresentaram um elevado número de dados faltantes, respectivamente pela ausência de despesas financeiras e de duplicatas a receber, fator que prejudica a utilização destas variáveis no modelo.

As matrizes de correlação encontram-se no Anexo 2, nas quais é possível observar que certas variáveis apresentam correlações próximas a 1 ou -1. O Quadro 18 apresenta as correlações das variáveis para o ano mais recente de informação disponível. A partir da análise, observa-se um elevado grau de correlação entre o conjunto de variáveis A2, A3, A4, B3 e E4, quando avaliadas em pares, e também, o conjunto de variáveis B2, C2, C3 e E1, quando avaliadas em pares.

As correlações observadas para o ano mais recente de informação disponível (ano 5), mantiveram quando avaliado os demais períodos disponíveis para estudo, conforme verificado nos Quadro 20, Quadro 21, Quadro 22 e Quadro 23 e ressaltados em verde.

5.2. REDUÇÃO DE DIMENSIONALIDADE

Baseada nas 24 variáveis originais para o ano mais recente de informação disponível (considerando empresas solventes e insolventes), os dados foram padronizados e, posteriormente, foi construída a matriz de covariância, para que assim fossem calculados os autovalores e realizada a avaliação do poder de explicação de cada componente principal.

Quadro 9. Autovalores resultantes da matriz de covariância

Autovalor	Peso	Representatividade	Representatividade acumulada
1	5,28199	22,97%	22,97%
2	3,54100	15,40%	38,36%
3	2,35822	10,25%	48,61%
4	2,03234	8,84%	57,45%
5	1,74408	7,58%	65,03%
6	1,68937	7,35%	72,38%
7	1,27014	5,52%	77,90%
8	1,05006	4,57%	82,47%
9	0,91619	3,98%	86,45%
10	0,89075	3,87%	90,32%
11	0,67850	2,95%	93,27%
12	0,56535	2,46%	95,73%
13	0,40734	1,77%	97,50%
14	0,24573	1,07%	98,57%
15	0,18546	0,81%	99,38%
16	0,05913	0,26%	99,63%
17	0,03366	0,15%	99,78%
18	0,02193	0,10%	99,87%
19	0,01509	0,07%	99,94%
20	0,01157	0,05%	99,99%
21	0,00212	0,01%	100,00%
22	0,00000	0,00%	100,00%
23	0,00000	0,00%	100,00%
24	0,00000	0,00%	100,00%

Avaliando os resultados do Quadro 9, observou-se que as 10 primeiras componentes principais capturam 90,32% da variabilidade total dos dados, que corresponde a um valor de alta precisão, dado que reduz a informação de 24 variáveis em apenas 10.

Portanto, é necessário realizar a transformação das variáveis originais através da combinação linear pelos pesos estabelecidos das componentes principais. Através dos pesos atribuídos, é possível avaliar as variáveis que serão mais representativas nas novas variáveis, conforme destacado em vermelho no Quadro 10.

Quadro 10. Coeficientes para constituição das componentes principais

Variável	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10
A1	0,048	-0,046	0,290	0,459	-0,009	0,442	0,018	-0,036	-0,034	0,004
A2	-0,123	-0,498	-0,062	-0,030	0,054	0,020	0,072	-0,017	-0,019	-0,048
A3	-0,228	-0,045	-0,090	0,340	0,421	-0,251	-0,016	0,034	0,140	-0,003
A4	-0,244	-0,028	-0,110	0,327	0,418	-0,225	-0,030	0,024	0,118	-0,005
A5	-0,025	0,044	-0,294	-0,448	0,003	-0,446	-0,107	0,025	0,064	0,052
B1	-0,197	0,017	0,140	0,000	-0,121	-0,085	-0,277	-0,444	-0,255	0,249
B2	-0,363	0,114	-0,050	-0,093	-0,043	0,110	0,254	0,013	-0,118	-0,233
B3	-0,108	-0,503	-0,070	-0,046	0,037	0,041	0,058	-0,016	-0,026	-0,043
B4	0,068	0,126	-0,439	-0,129	0,076	0,380	0,074	-0,015	-0,125	-0,066
C1	-0,028	0,130	-0,432	-0,039	0,214	0,367	-0,047	-0,024	-0,152	0,086
C2	-0,379	0,064	0,040	-0,032	-0,180	0,073	-0,223	-0,006	-0,069	0,137
C3	-0,329	0,086	-0,034	-0,041	0,041	0,014	0,162	-0,107	-0,245	-0,197
C4	-0,108	-0,503	-0,070	-0,046	0,037	0,041	0,058	-0,016	-0,026	-0,043
C5	0,044	-0,008	-0,062	0,014	0,133	0,058	-0,585	0,188	-0,225	-0,616
D1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
D2	0,036	-0,025	-0,393	0,375	-0,412	-0,145	0,031	-0,012	-0,023	-0,023
D3	0,150	-0,043	0,043	0,063	0,128	-0,215	0,071	-0,437	-0,648	-0,011
D4	0,036	-0,025	-0,393	0,375	-0,413	-0,145	0,031	-0,013	-0,023	-0,024
D5	-0,302	0,083	0,007	-0,101	-0,035	0,054	0,498	-0,018	-0,004	-0,100
D6	-0,007	-0,020	0,070	0,055	0,022	-0,110	0,091	0,729	-0,544	0,346
E1	-0,375	0,059	-0,040	-0,010	-0,113	0,129	-0,257	0,081	0,071	0,131
E2	0,386	-0,016	-0,064	-0,002	0,154	-0,040	0,235	-0,030	-0,033	-0,136
E3	0,084	-0,031	-0,255	-0,002	0,298	0,186	-0,104	-0,138	0,049	0,504
E4	-0,066	0,408	0,022	0,181	0,195	-0,156	0,115	-0,039	0,011	-0,067

Assim as variáveis transformadas, através da combinação linear das variáveis originais, podem ser caracterizadas por:

- Componente 1: explicada positivamente, em especial pela variável E2 e negativamente, em especial pelas variáveis B2, C2, C3, D5 e E1;
- Componente 2: explicada positivamente, em especial pela variável E4 e negativamente, em especial pelas variáveis A2, B3 e C4;
- Componente 3: explicada negativamente, em especial pelas variáveis B4, C1, D2 e D4;
- Componente 4: explicada positivamente, em especial pelas variáveis A1, A3, A4, D2, D4 e negativamente, em especial pela variável A5;
- Componente 5: explicada positivamente, em especial pelas variáveis A3 e A4, e negativamente, em especial pelas variáveis D2 e D4;
- Componente 6: explicada positivamente, em especial pelas variáveis A1, B4 e C1, e negativamente, em especial pela variável A5;
- Componente 7: explicada positivamente, em especial pela variável D5 e negativamente, em especial pela variável C5;
- Componente 8: explicada positivamente, em especial pela variável D6 e negativamente, em especial pelas variáveis B1 e D3;
- Componente 9: explicada negativamente, em especial pelas variáveis D3 e D6;
- Componente 10: explicada positivamente, em especial pelas variáveis D6 e E3, e negativamente, em especial pela variável C5.

A partir da redução de dimensionalidade, espera-se facilitar a identificação de informações relevantes e facilitar a interpretação dos resultados dos modelos a serem desenvolvidos.

5.3. MODELOS DE PREVISÃO

Posteriormente a redução de dimensionalidade foram desenvolvidos os dois modelos pré-estabelecidos para efeito de comparação, um modelo de análise discriminante e um modelo de redes neurais artificiais auto-organizável. A

ferramenta utilizada para o desenvolvimento de ambos os modelos foi o MatLab, versão 2010.

5.3.1. MODELO UTILIZANDO ANÁLISE DISCRIMINANTE

Previamente à elaboração do modelo, foram avaliadas as suposições estabelecidas para o uso da metodologia de análise discriminante. Primeiramente, com relação à dependência entre as variáveis, a suposição foi atendida após a modificação das variáveis via componentes principais, pois esta é uma das adaptações realizadas pela técnica.

Com relação à normalidade das variáveis, foram realizados gráficos de avaliação de normalidade para cada uma das dez componentes principais. No Anexo 3, observa-se que as variáveis atendem parcialmente às suposições pois apresentaram leves desvios dos dados próximos às extremidades. Esta análise fica evidenciada pelo distanciamento de alguns dados da reta estabelecida no gráfico, a qual representa a distribuição normal esperada.

Apesar das suposições serem atendidas parcialmente, o modelo de análise discriminante desenvolvido, utilizando-se de três entradas com relação a base de dados, sendo 1) base de validação (10 linhas e 10 colunas); 2) base de treinamento (30 linhas e 10 colunas); e 3) saídas esperadas para base de treinamento (30 linhas e 1 coluna).

Vale ressaltar que nas dimensões apresentadas a coluna representa uma variável, enquanto cada linha corresponde aos resultados observados por cada elemento amostral (empresa avaliada).

Definiu-se a utilização da distribuição linear para o desenvolvimento do modelo discriminante, uma vez que esta distribuição ajusta distribuições normais multivariadas para cada um dos grupos especificados (solvente e insolvente) através da matriz de covariância.

Assim, o modelo desenvolvido apresentou um coeficiente para cada uma das dez componentes principais (CP) utilizadas durante a elaboração além de um coeficiente para a constante, representados no Quadro 11.

Quadro 11. Coeficientes resultantes do modelo de análise discriminante

Variável	Coeficiente
Constante	-1,3602
CP1	4,6972
CP2	21,9161
CP3	7,2140
CP4	-4,7702
CP5	5,2747
CP6	-2,8510
CP7	5,9517
CP8	-4,6865
CP9	6,9231
CP10	10,0989

Matematicamente, a equação resultante do modelo de análise discriminante pode ser representada por:

$$Y_{\text{Discr.}} = -1,37 + 4,70 \times CP1 + 21,92 \times CP2 + 7,21 \times CP3 - 4,77 \times CP4 + 5,27 \times CP5 - 2,85 \times CP6 + 5,95 \times CP7 - 4,69 \times CP8 + 6,92 \times CP9 + 10,10 \times CP10 \quad (17)$$

Com auxílio da distribuição, é possível avaliar a representatividade de erro de classificação para a base de treinamento. Isto é, o percentual de observações da base de treinamento que está sendo classificada incorretamente nos grupos esperados, sumarizados no Quadro 12.

A partir da aplicação do modelo na base de treinamento, o erro de estimação observado foi de 3,33%, que indica que apenas um elemento da base de treinamento está sendo classificado incorretamente, para a base de dados utilizada mais recente (ano 5).

Desta maneira, o próximo passo foi aplicar o modelo à base de treinamento ao longo dos cinco anos de base histórica, fator que permitiu observar que os melhores índices de assertividade encontram-se para as data bases próximas àquela utilizada para construção do modelo. Neste sentido, nota-se que o modelo reduz sua aderência ao ser utilizado em base de dados de períodos anteriores aos

utilizados para treinamento, reduzindo a assertividade de 97% para 77%, quando comparado o ano 5 e o ano 1 respectivamente.

Quadro 12. Resultados do modelo de análise discriminante (treinamento)

Ano base	Solventes		Insolventes		Assertividade	
	Corretas	Incorretas	Corretas	Incorretas	Corretas	Incorretas
Ano 1	10 (67%)	5 (33%)	13 (87%)	2 (13%)	23 (77%)	7 (23%)
Ano 2	11 (73%)	4 (27%)	11 (73%)	4 (27%)	22 (73%)	8 (27%)
Ano 3	12 (80%)	3 (20%)	13 (87%)	2 (13%)	25 (83%)	5 (17%)
Ano 4	12 (80%)	3 (20%)	14 (93%)	1 (7%)	26 (87%)	4 (13%)
Ano 5 (mais recente)	14 (93%)	1 (7%)	15 (100%)	0 (0%)	29 (97%)	1 (3%)

Posteriormente, foi aplicado o modelo à base de validação para avaliar a real assertividade do modelo para situações de previsão de insolvência ou solvência, uma vez que as informações não foram utilizadas para construção do modelo:

Quadro 13. Resultados do modelo de análise discriminante (validação)

Ano base	Solventes		Insolventes		Assertividade	
	Corretas	Incorretas	Corretas	Incorretas	Corretas	Incorretas
Ano 1	1 (20%)	4 (80%)	4 (80%)	1 (20%)	5 (50%)	5 (50%)
Ano 2	1 (20%)	4 (80%)	5 (100%)	0 (100%)	6 (60%)	4 (40%)
Ano 3	1 (20%)	4 (80%)	4 (80%)	1 (80%)	5 (50%)	5 (50%)
Ano 4	2 (40%)	3 (60%)	5 (100%)	0 (100%)	7 (70%)	3 (30%)
Ano 5 (mais recente)	3 (60%)	2 (40%)	4 (80%)	1 (20%)	7 (70%)	3 (30%)

Conforme esperado, ao avaliar o Quadro 13, nota-se que os índices de assertividade para a amostra de validação foram menores do que os observados com a utilização da base de treinamento, redução de 97% de assertividade (base de

treinamento) para 70% de assertividade (base de validação), quando comparados os resultados do ano 5. Porém ainda, os resultados obtidos são considerados positivos e aplicáveis, especialmente pelo fato de que no ano mais recente a taxa de assertividade de insolvente e solvente é muito próxima e indica que o modelo não apresenta viés de classificação para alguma categoria.

No entanto, através dos resultados da base de validação para o Ano 1 (previsão com 5 anos de antecedência utilizando-se o mesmo modelo), há indícios que o modelo encontra-se não calibrado uma vez que classifica a base em insolvente com alta probabilidade, pois 80% dos dados foram classificados insolventes, sendo que 50% corretamente e 50% incorretamente.

5.3.2. MODELO UTILIZANDO REDES NEURAIS ARTIFICIAIS

O modelo de redes neurais artificiais desenvolvido utilizou-se de duas entradas com relação à base de dados, sendo 1) base de validação (10 linhas e 10 colunas) e 2) base de treinamento (10 linhas e 30 colunas). Vale ressaltar que nas dimensões apresentadas a linha representa uma variável enquanto cada coluna corresponde aos resultados observados por cada elemento amostral.

Assim, foi desenvolvida uma rede neural artificial auto-organizável com o objetivo de utilizar o algoritmo competitivo para resolução do problema, contendo como características estruturais a topologia hexagonal, dimensão dos neurônios 3 por 2; a medida Euclidiana como forma de cálculo entre os neurônios. Isto posto, cada um dos 6 neurônios irão competir entre si através das relações permitidas visualizadas na Figura 12 pelas linhas em vermelho.

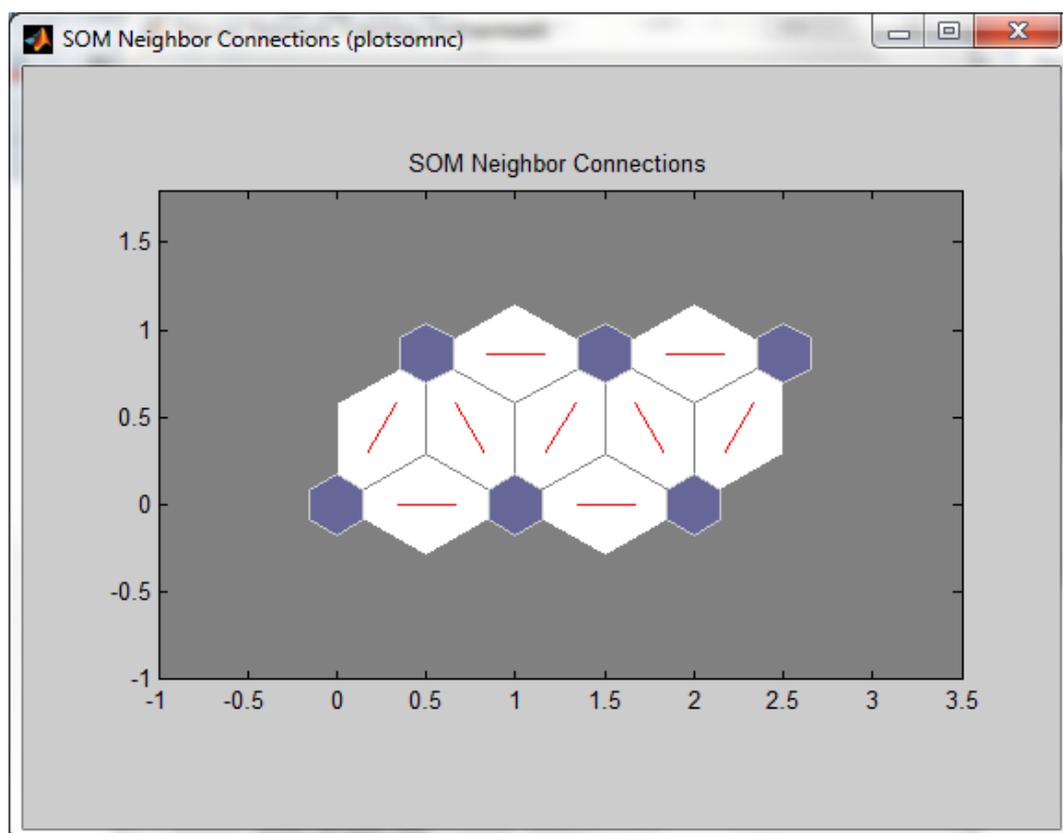


Figura 12. Topologia do modelo de rede neural artificial auto-organizável

Através do treinamento da rede com a base de 30 elementos (empresas), é possível avaliar a proximidade entre os neurônios, representado graficamente

através das cores que variam do preto até o amarelo (ver figura 13), sendo que quanto mais escuro o neurônio maior a sua proximidade e conseqüentemente mais semelhantes são suas características, fator este que auxiliará no agrupamento e caracterização destes neurônios posteriormente.

Interpretando a Figura 13, observa-se que os neurônios [1,1] (linha e coluna respectivamente) e [2,1] são os mais próximos, pois apresentam a cor preta, e os neurônios [1,2] e [1,3] são os mais distantes, pois apresentam a cor amarela mais expressiva.

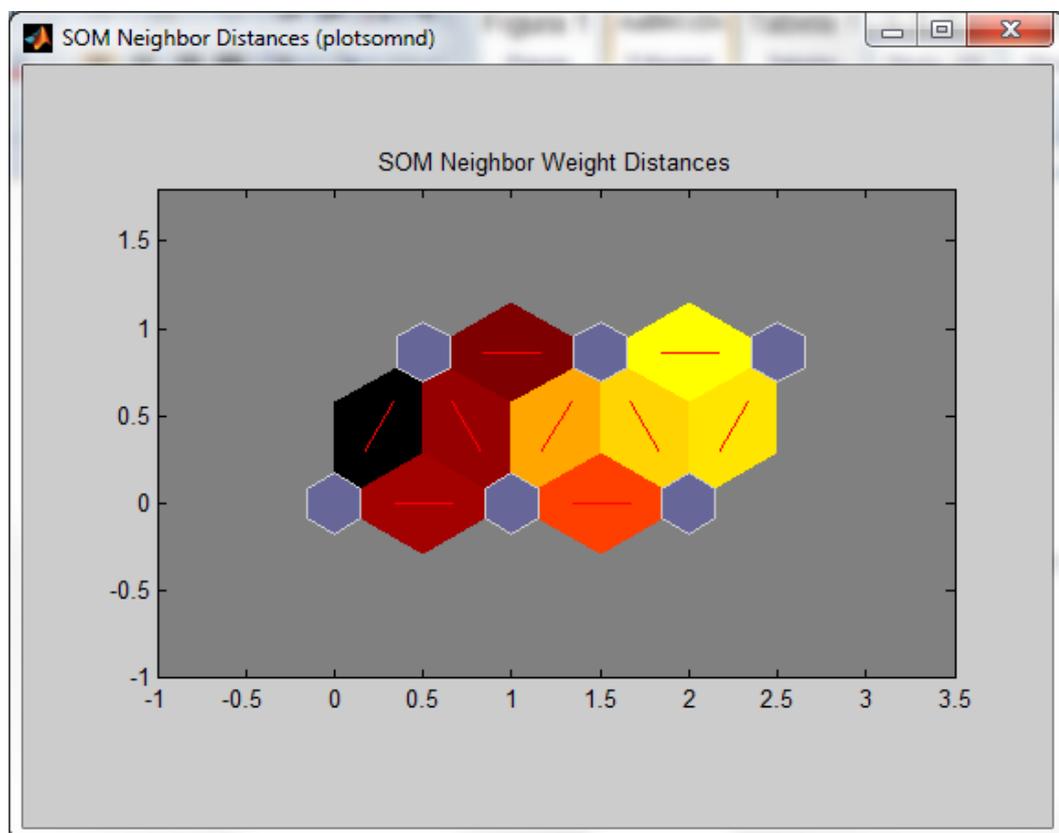


Figura 13. Distância entre os neurônios após o treinamento da rede

Na sequência, foi possível avaliar o quanto cada variável auxilia na classificação do elemento em um determinado neurônio, sendo representado também através do esquema de cores descritos anteriormente (ver Figura 14), onde cada componente principal utilizada foi denominada *input*. Assim, a componente principal 1 foi referenciada na figura como *input 1*, e sucessivamente para as demais componentes, totalizando os 10 *inputs*.

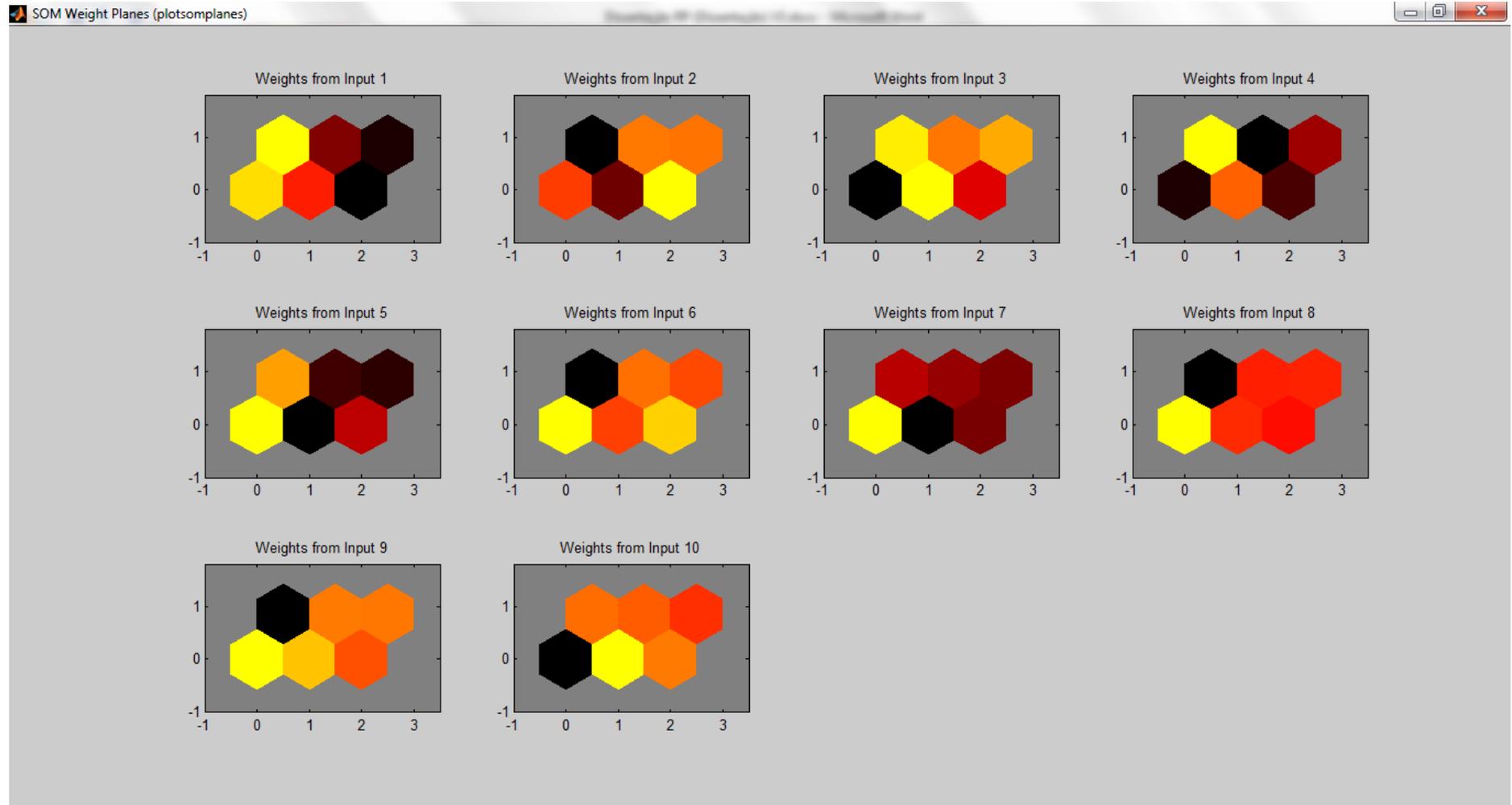


Figura 14. Avaliação do impacto das variáveis na alocação dos neurônios.

O Quadro 14 resume a análise da Figura 14, indicando quais os neurônios tornam-se mais próximos e quais se tornam mais distantes para cada uma das variáveis utilizadas para a construção do modelo.

Quadro 14. Influência das variáveis para classificação nos neurônios

Variável	Neurônios próximos	Neurônios distantes
CP 1	[1,3] e [2,3]	[1,1] e [2,1]
CP 2	[1,1] e [2,2]	[1,2] e [1,3]
CP 3	[2,1]	[1,1] e [2,2]
CP 4	[1,2] e [2,1]	[1,1] e [2,2]
CP 5	[2,2] e [1,2]	[1,1] e [2,1]
CP 6	[1,1]	[2,1] e [2,3]
CP 7	[2,2]	[2,1]
CP 8	[1,1]	[2,1]
CP 9	[1,1]	[2,1]
CP 10	[2,1]	[2,2]

Assim, baseado na influência de cada variável sobre o neurônio, os elementos utilizados para treinamento foram alocados no neurônio vencedor após a comparação com os demais. Avaliando a Figura 15, nota-se que os neurônios [1,3] e [1,2] apresentam o maior número de elementos, respectivamente 14 e 10 elementos, enquanto os neurônios [1,1], [2,1] e [2,2] apresentam apenas um elemento cada.

Com base nesta classificação, o próximo passo foi avaliar as características dos elementos que foram alocados em cada neurônio, verificando qual a representatividade de elementos solventes e insolventes, permitindo assim caracterizar cada neurônio de acordo com sua predominância.

A definição da característica predominante do neurônio auxiliará diretamente no processo preditivo quando aplicado à base de validação, pois após a caracterização dos neurônios, as alocações futuras serão associadas a estas características.

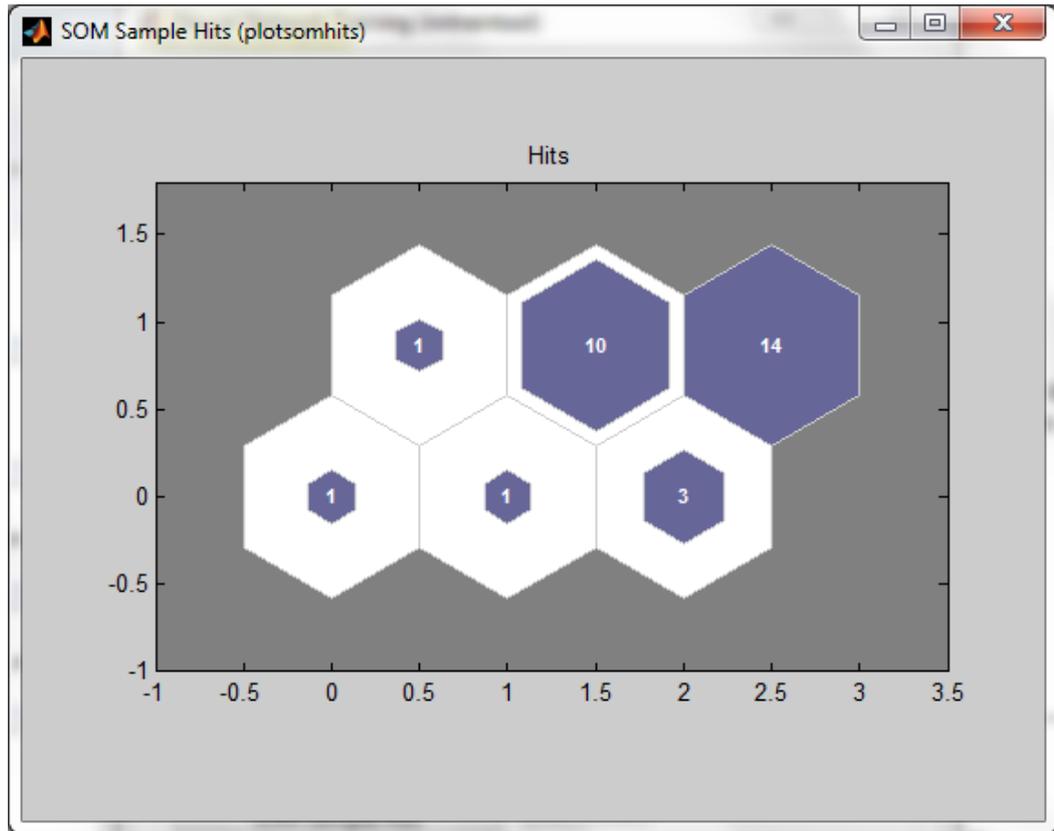


Figura 15. Alocação de empresas por neurônio

Através do Quadro 15, é possível verificar que os neurônios [1,3] e [2,3] são caracterizados por empresas solventes enquanto que os neurônios [1,1], [1,2], [2,1] e [2,2] são caracterizados por empresas insolventes. Vale ressaltar que o neurônio [1,3] é aquele que concentra a maior parte de empresas solventes (86% da amostra utilizada) e que o neurônio [1,2] é o que concentra a maior parte de empresas insolventes (67% da amostra utilizada).

Quadro 15. Caracterização dos neurônios

Neurônio	Solventes	Insolventes	Predominância
[1,1]	0	1	Insolvência
[1,2]	10	0	Insolvência
[1,3]	1	13	Solvência
[2,1]	0	1	Insolvência
[2,2]	0	1	Insolvência
[2,3]	1	2	Solvência

Assim, ao utilizar o modelo para previsão de insolvência, quando o elemento for alocado nos neurônios [1,1], [1,2], [2,1] ou [2,2], a empresa será denominada insolvente, no entanto quando o elemento for alocado nos neurônios [1,3] ou [2,3] a empresa será denominada solvente.

Pela Figura 16, é possível avaliar a posição de cada um dos elementos utilizados para treinamento, que estão representados pelos pontos em verde; e a posição dos neurônios após o treinamento da rede, que estão representados pelos pontos em azul. O retângulo laranja na figura destaca o conjunto de neurônios que apresentaram elevada concentração de elementos, enquanto o retângulo azul destaca o conjunto de neurônios que apresentam características distintas e por isso encontram-se distantes em relação aos demais .

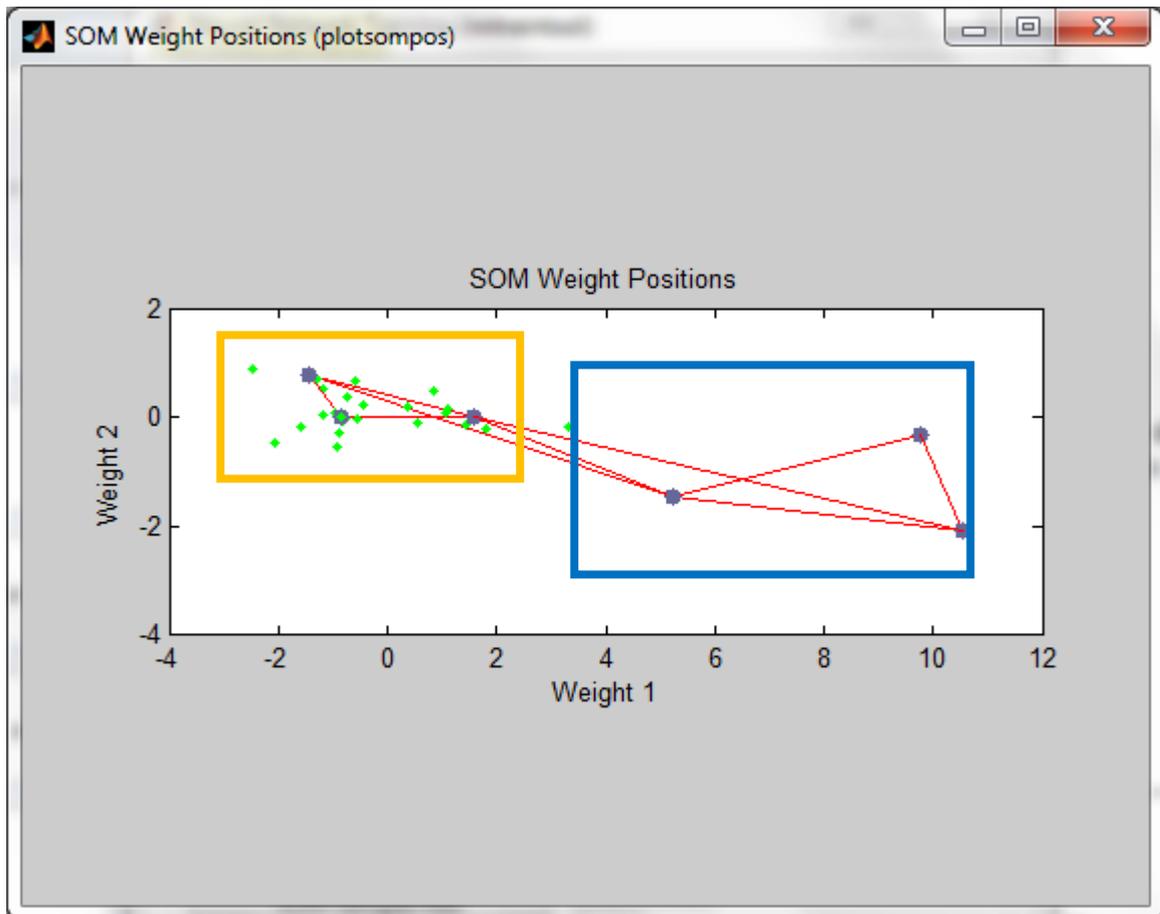


Figura 16. Posição dos neurônios após o treinamento.

Com o modelo desenvolvido, realizou-se a comparação dos resultados de assertividade para a base utilizada. Adicionalmente, o modelo foi aplicado às bases

de treinamento para os anos anteriores disponíveis, a fim de avaliar o comportamento dos resultados ao longo do ano.

Quadro 16. Resultados do modelo de redes neurais artificiais (treinamento)

Ano base	Solventes		Insolventes		Assertividade	
	Corretas	Incorretas	Corretas	Incorretas	Corretas	Incorretas
Ano 1	7 (47%)	8 (53%)	15 (100%)	0 (0%)	22 (73%)	8 (27%)
Ano 2	9 (60%)	6 (40%)	13 (87%)	2 (13%)	22 (73%)	8 (27%)
Ano 3	10 (67%)	5 (33%)	14 (93%)	1 (7%)	24 (80%)	6 (20%)
Ano 4	11 (73%)	4 (27%)	15 (100%)	0 (0%)	26 (87%)	4 (13%)
Ano 5 (mais recente)	13 (87%)	2 (13%)	15 (100%)	0 (0%)	28 (93%)	2 (7%)

A partir do Quadro 16, é possível avaliar que os melhores índices de assertividade encontram-se para as data bases próximas àquela utilizada para construção do modelo, variando de 93% para o ano mais recente e decrescendo até 73% o ano mais antigo.

Vale destacar que a classificação de empresas solventes parece não estar calibrada para o ano mais antigo, pois a proporção de acertos é de 47% para solventes e de 100% para insolventes, indicando que o modelo possivelmente está tendendo a classificar as empresas como insolventes.

Posteriormente, foi aplicado o modelo às bases de validação, sendo possível avaliar a real assertividade do modelo para situações de previsão de insolvência ou solvência, uma vez que as informações não foram utilizadas para construção do modelo.

A partir do Quadro 17, é possível avaliar que os melhores índices de assertividade também se encontram para as data bases próximas àquela utilizada para construção do modelo, e também é possível observar que o modelo reduz sua aderência ao ser utilizado em base de dados de períodos anteriores aos utilizados para treinamento.

No entanto, através dos resultados da base de validação para o Ano 1 (previsão com 5 anos de antecedência utilizando-se o mesmo modelo), há indícios que o modelo encontra-se não calibrado já que classifica a base em insolvente com alta probabilidade, pois 90% dos dados foram classificados insolventes, sendo que 55% corretamente e 45% incorretamente.

Quadro 17. Resultados do modelo de redes neurais artificiais (validação)

Ano base	Solventes		Insolventes		Assertividade	
	Corretas	Incorretas	Corretas	Incorretas	Corretas	Incorretas
Ano 1	1 (20%)	4 (80%)	5 (100%)	0 (0%)	6 (60%)	4 (40%)
Ano 2	2 (40%)	3 (60%)	5 (100%)	0 (0%)	7 (70%)	3 (30%)
Ano 3	3 (60%)	2 (40%)	5 (100%)	0 (0%)	8 (80%)	2 (20%)
Ano 4	3 (60%)	2 (40%)	5 (100%)	0 (0%)	8 (80%)	2 (20%)
Ano 5 (mais recente)	4 (80%)	1 (20%)	5 (100%)	0 (0%)	9 (90%)	1 (10%)

Assim, analisando os resultados, observa-se que o modelo apresenta boa capacidade preditiva de empresas insolventes com 1 ano de antecedência (90% de assertividade), porém o modelo reduz sua precisão quando é utilizado para prever a insolvência com 5 anos de antecedência (60% de assertividade) (Ver Quadro 17).

5.4. COMPARAÇÃO ENTRE AS METODOLOGIAS

Ao avaliar os resultados obtidos pelas metodologias de análise discriminante e redes neurais artificiais verificamos que os resultados apresentam características semelhantes com relação a sua assertividade e seu comportamento quando aplicados a amostra de treinamento e de validação.

Observando a Figura 17, é possível avaliar os resultados de assertividade dos modelos conjuntamente para classificação de empresas em solventes e insolventes, indicando que a aplicação do modelo no ano mais recente (ano 5) apresentou resultados significativos quando avaliamos a assertividade geral do modelo. Neste

caso, as redes neurais artificiais apresentaram proporção de acerto de 90% enquanto que a análise discriminante acerto de 70%. Nos anos subsequentes as redes neurais artificiais também apresentaram resultados superiores quando comparados análise discriminante. Vale destacar que para ambas as metodologias os resultados apresentaram diminuição na assertividade ao longo do tempo, pois para o ano mais antigo (ano 1) as redes neurais artificiais apresentaram proporção de acerto de 60% enquanto que para a análise discriminante acerto foi de 50%.

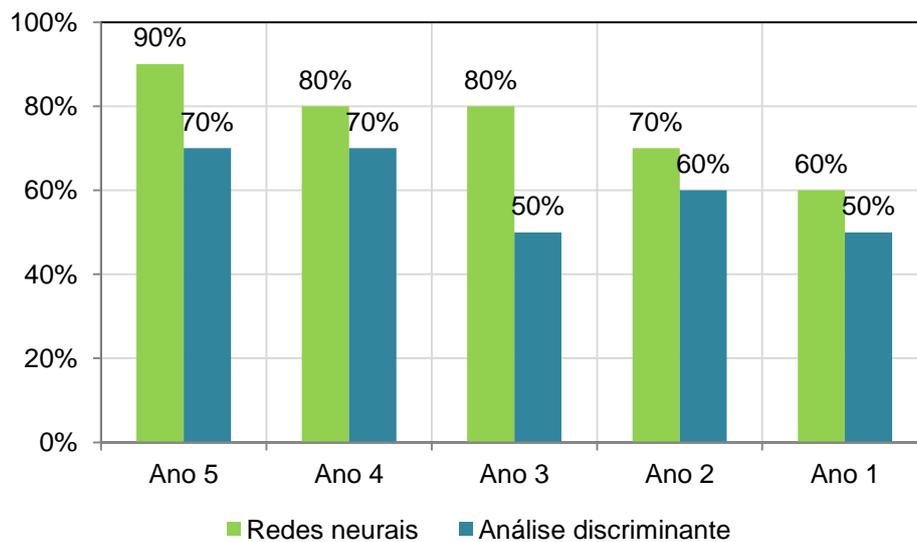


Figura 17. Assertividade de modelo por metodologia

Baseado nos resultados apresentados é importante avaliar a assertividade levando em consideração as classificações realizadas, especialmente para avaliar se os modelos estão apresentando algum desvio durante a classificação das empresas. Neste sentido, a Figura 18 apresenta a assertividade por metodologia para as empresas solventes, enquanto a Figura 19 apresenta a assertividade por metodologia para as empresas insolventes.

Sendo assim, ao avaliar os resultados quando aplicado base de dados mais recentes (ano 5) nota-se que a assertividade é próxima para ambas as metodologias e para ambas as categorias (solventes e insolventes), pois para redes neurais artificiais o percentual de acerto de solventes foi de 80% e para insolventes de 100%, enquanto para a análise discriminante o percentual de acerto de solvente foi de 60% e de insolventes de 80% (Ver Figura 18 e 19).

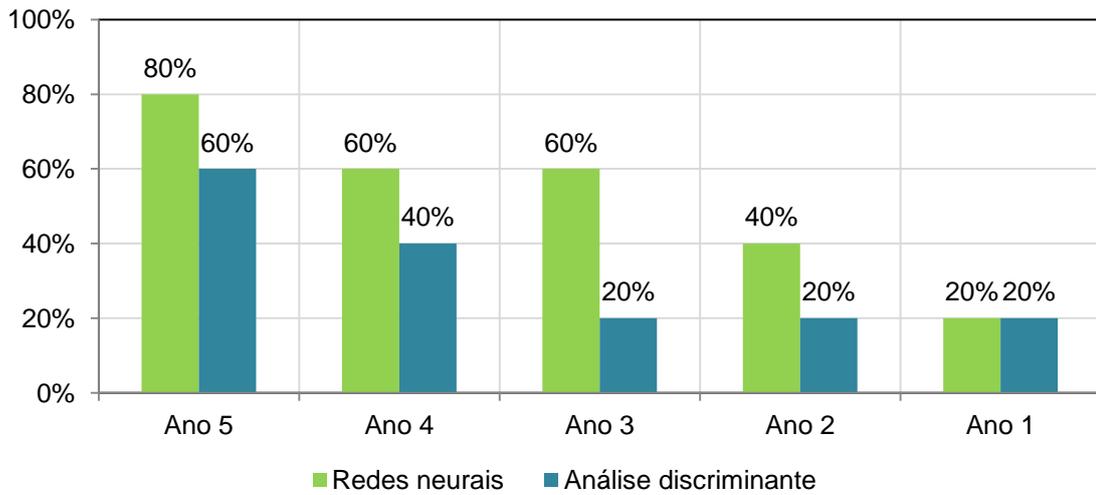


Figura 18. Assertividade dos modelos considerando as empresas solventes

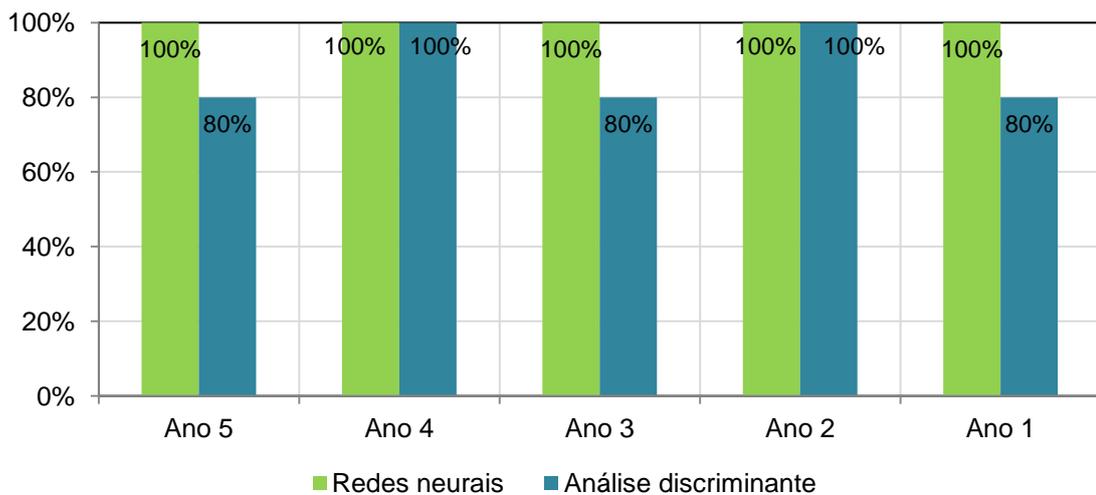


Figura 19. Assertividade dos modelos considerando as empresas insolventes

Para o ano mais antigo (ano 1) é possível observar que ambos os modelos não estão calibrados, uma vez que apresentam tendência para classificação das empresas em insolventes, podendo ser observado pelo alto percentual de acerto de classificação de insolventes e baixo percentual de acerto para solventes. Para a análise discriminante observou-se acerto para solventes de 20% enquanto para insolventes 80%. Já para o modelo de redes neurais artificiais observou acerto para solventes de 20% enquanto para insolventes de 100% (Ver Figura 18 e 19).

6. CONSIDERAÇÕES FINAIS

Baseado nas alterações do cenário econômico, as empresas tornaram-se sensíveis e suscetíveis à incapacidade financeira, fator que motivou a constituição de modelos preditivos para ampliar o entendimento sobre a insolvência de empresas.

Diversos modelos estatísticos já foram propostos com esta finalidade, porém através do estudo comparativo, foi possível avaliar que as redes neurais artificiais possuem características distintas que complementam as demais metodologias. Entre essas características, ressaltam-se a não exigência de relação linear entre a variável de entrada e saída, a sua capacidade de adaptação durante a construção da rede e que a ausência de alguma informação não invalida a realização de predição da variável resposta.

A fim de avaliar a viabilidade do uso de redes neurais artificiais para modelos de insolvência, foram selecionados na literatura cinco modelos construídos baseados em métodos estatísticos (conforme Quadro 4) que utilizam informações disponibilizadas nos demonstrativos de resultados financeiros das empresas, dispondo de um total de 24 variáveis a serem utilizadas.

Devido ao elevado número de variáveis, foi realizada uma redução de dimensionalidade através da análise de componentes principais que permitiu a transformação das 24 variáveis em 10 variáveis, mantendo ainda 90% da variabilidade das variáveis originais.

Desta forma, a partir das 10 variáveis, foram construídos dois modelos, sendo um utilizando a metodologia de análise discriminante e outro utilizando a metodologia de redes neurais artificiais, empregando-se uma amostra de 40 empresas, sendo 30 empresas selecionadas aleatoriamente a serem utilizadas durante a fase de treinamento (15 solventes e 15 insolventes) e as 10 empresas restantes a serem utilizadas como validação dos modelos desenvolvidos (5 solventes e 5 insolventes).

Baseado nos resultados observa-se que as redes neurais artificiais são viáveis para o desenvolvimento de modelos de previsão devido as suas características complementares aos métodos estatísticos. Para os dados considerados neste trabalho, compreendidos no período entre 2004 e 2012, os

modelos neurais apresentaram resultados superiores quando comparados aos resultados da análise discriminante.

O modelo de rede neural artificial desenvolvido, quando aplicados à base de dados históricas, anos anteriores à constatação da insolvência, apresentam redução na sua assertividade, reduzindo de 90% (ano anterior a constatação) para 60% (5 anos anterior a constatação), fato também observado quando utilizado a metodologia de análise discriminante. Porém, é possível observar que as redes neurais artificiais apresentam resultados superiores a 70% até 4 anos antes da constatação de insolvência, enquanto que a análise discriminante até 2 anos antes da constatação de insolvência, indicando que as redes apresentam resultados mais consistentes quando comparados com a análise discriminante para o período em questão. Vale ressaltar que a superioridade na assertividade observada das redes neurais artificiais pode ser decorrente do atendimento parcial das suposições esperadas das variáveis para uso da metodologia de análise discriminante, conforme observado no Anexo 03.

Na prática, o modelo proposto de redes neurais artificiais pode atender aos objetivos esperados, mesmo tendo apresentado uma distorção da assertividade ao longo do tempo. Esse resultado não colocaria a Instituição Financeira em elevada exposição ao risco, uma vez que, com o passar do tempo, o modelo passa a classificar como insolvente as empresas que seriam solventes e não o contrário, que poderia sugerir a concessão indevida de crédito.

Frente às conclusões, é possível considerar as redes neurais artificiais como um método alternativo para a realização de predições. Para futuras pesquisas sugere-se: 1) a utilização dos modelos auto-organizáveis para estudos preditivos em outros temas e em outras áreas do conhecimento e 2) estudos comparativos de previsão entre as metodologias de redes neurais artificiais auto-organizáveis e métodos estatísticos de análise de agrupamento, que apresenta classificação não supervisionada de dados.

REFERÊNCIAS

- ALMEIDA, F.C.; MATIAS, A. B. Avaliação de insolvência de bancos brasileiros por meio de redes neurais. São Paulo: Universidade de São Paulo. 2001
- ALTMAN, E. I. *Financial ratios: discriminant analysis and the prediction of corporate bankruptcy*. The Journal of Finance, Vol. 23, pp. 589-609, 1968.
- ALVES, L. A.; CHICRALLA, R. C.; LEITE, V. P.; MELLO, J. C. C.; BIONDI-NETO, L. B. *Previsão de faturamento para lojas do setor de varejo com redes neurais*. PODes - Revista Eletrônica Pesquisa Operacional para o Desenvolvimento. Rio de Janeiro, Vol. 5, pp. 1-13, 2013.
- ASSAF, A. N. *Finanças corporativas e valor*. São Paulo: Atlas, 2003.
- BARBOSA, P. R. ; BUENO, E.I. *Modelo neural com atraso de tempo para previsão do INCC*. Revista Sinergia (CEFETSP), Vol. 12, pp. 16-22, 2011.
- BARBOSA, P. R.; CRIVELARO, K. C. O.; SELEGHIM, P. *On the application of self-organizing neural networks in gas-liquid and gas-solid flow regime identification*. Journal of the Brazilian Society of Mechanical Sciences and Engineering (Impresso), Vol. 32, pp. 15-20, 2010.
- BM&FBOVESPA. *Informações periódicas e eventuais* Disponível em: <<http://siteempresas.bovespa.com.br/consbov/InfoPerEventuaisBuscData.asp?Site=C>> . Acesso em: 01 de fevereiro de 2013.
- BRASIL. Decreto-Lei nº 7.661, de 21 de junho de 1945. *Dispõe sobre a lei de falências*. Diário Oficial [da República Federativa do Brasil], Brasília, DF.
- BRASIL. Lei nº 11.101, de 09 de fevereiro de 2005. *Regula a recuperação judicial, a extrajudicial e a falência do empresário e da sociedade empresária*. Diário Oficial [da República Federativa do Brasil], Brasília, DF.
- BRASIL. Lei nº 6.404, de 17 de dezembro de 1976. *Dispõe sobre as Sociedades por Ações*. Diário Oficial [da República Federativa do Brasil], Brasília, DF.
- BUSSAB, W. O. ; MORETTIN, P. A. *Estatística básica*. São Paulo: Saraiva, 2004.
- CAMPINHO, S. *Falência e recuperação de empresa: O novo regime da insolvência empresarial*. Rio de Janeiro: Renovar, 2010.
- CANENA, A. C. *Redes neurais artificiais dinâmicas aplicadas a detecção automática de eventos comportamentais*. Universidade Federal de Santa Catarina, 2013.
- CASTRO JUNIOR, F. H. F. *Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais*. São Paulo: Universidade de São Paulo. 2003.
- COOPER, D.; SCHINDLER, P. *Métodos de pesquisa em administração*. Porto Alegre: Bookman, 2003.

- CURI, L. Z. *Aplicação de redes neurais na precificação de debêntures*. Escola de Economia de São Paulo, 2008.
- DAMODARAN, A. *Finanças corporativas: teoria e prática*. São Paulo: Bookman, 2004.
- GUIMARÃES, A.; MOREIRA, T. B. S. *Previsão de insolvência: um modelo baseado em índices contábeis com utilização da análise discriminante*. Revista de Economia Contemporânea. Rio de Janeiro, Vol. 12, pp. 151-179, 2009.
- GUSMÃO, M. *Lei de recuperação e falências*. Rio de Janeiro: Lumen Juris, 2009.
- HAIR, J. F.; TATHAM, R. L.; BLACK, W.; ANDERSON, R. E. *Análise Multivariada de Dados*. Porto Alegre: Bookman, 2007.
- HAYKIN, S. S. *Redes neurais: princípios e prática*. Porto Alegre: Bookman, 2001.
- HEBB, D. O. *The organization of behaviour: a neuropsychological theory*. New York: Willey, 1949.
- INFOMONEY. *Número de novos empreendimentos no País ultrapassa 574 mil em 2012*. Disponível em: <<http://www.infomoney.com.br/empreendedor/noticia/2436028-numero+novos+empreendimentos+pais+ultrapassa+574+mil+2012>>. Acesso em: 08 de junho de 2012.
- JOHNSON, R. A.; WICHERN D. W. *Applied multivariate statistical analysis*. New Jersey: Pearson Prentice Hall, 2007.
- JOLLIFFE, I. T. *Principal component analysis*. New York: Springer, 2002.
- KANITZ, S. C. *Como prever falências*. São Paulo: McGraw-Hill, 1978.
- KOHONEN, T. *Self-organizing maps*. Berlim: Springer, 2001.
- KOVACS, Z. L. *Redes Neurais Artificiais: Fundamentos e Aplicações*. São Paulo: Editora Livraria da Física, 2002.
- MALAGA, F. K. *Análise de demonstrativos financeiros e de performance empresarial: para empresas não financeiras*. São Paulo: Saint Paulo, 2009.
- MARCUS, A.; BODIE, Z.; KANE, A. *Fundamentos de investimentos*. Porto Alegre: Bookman, 2007.
- MCCULLOCH, W. S.; PITTS, W. *A Logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics. Vol. 5, pp. 198-211, 1943.
- MCLACHLAN, G. *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley, 2004.
- OLIVEIRA, R. B. S.; *Detecção de problemas em Instituições financeiras utilizando modelos estatísticos*. Dissertação de Mestrado. Universidade de São Paulo, 2010.
- ROSAS, M. A.; BEZERRA, A. F. B.; DUARTE-NETO, P. J. *Uso das redes neurais artificiais na aplicação de metodologia para alocação de recursos da saúde*. Rev. Saúde Pública. São Paulo, Vol. 47, pp. 128-136, 2013.

- SANTANA, G. A.; CARNEIRO, R. N. A. *Previsão de insolvência empresarial empregando redes neurais*. São Paulo: Instituto Tecnológico de Aeronáutica, 2004.
- SANTOS, A. A. P. *Previsão não-linear da taxa de câmbio real/dólar utilizando redes neurais e sistemas nebulosos*. Universidade Federal de Santa Catarina, 2005.
- SANVINCENTE, A. Z.; MINARDI, A. M. A. F. *Identificação de indicadores contábeis significativos para previsão de concordata de empresas*. 1998. Disponível em: <http://www.risktech.com.br/PDFs/indicadores_concordata.pdf>. Acesso em: 05 de junho de 2012.
- SERASA-EXPERIAN. *Indicador Serasa Experian de demanda das empresas por crédito*. Disponível em: <http://www.serasaexperian.com.br/release/indicadores/demanda_pj_credito.htm> . Acesso em: 01 de julho de 2013.
- SERASA-EXPERIAN. *Indicador Serasa Experian de falências e recuperações*. Disponível em: <http://www.serasaexperian.com.br/release/indicadores/falencias_concordatas.htm> . Acesso em: 01 de julho de 2013.
- SILVA, I. N. S.; SPATTI, D. H.; FLAUZINO, R. A. *Redes neurais artificiais: para engenharia e ciências aplicadas*. São Paulo: Artliber, 2010.
- SILVA, J. P. *Análise financeira das empresas*. São Paulo: Atlas, 2001.
- UENO, A. S. C. *Modelos causais no cálculo de capital para risco operacional: investigação do uso de redes neurais artificiais como modelo avançado de mensuração de capital*. Escola de Economia de São Paulo, 2010.
- YANO, V.; Zimmer, A.; Ferrari, G. L. *Aplicação de redes neurais artificiais no auxílio à determinação de diabetes através de pupilometria dinâmica*. V Latin American Congress on Biomedical Engineering CLAIB. Havana, pp. 16-21, 2011.

APÊNDICE A: GLOSSÁRIO CONTÁBIL

Ativo: São todos os bens, direitos e valores a receber de uma entidade. Contas do ativo têm saldos devedores, à exceção das contas retificadoras (como depreciação acumulada e provisões para ajuste ao valor de mercado).

Ativo circulante: Corresponde as disponibilidades, os direitos realizáveis no curso do exercício social subsequente e as aplicações de recursos em despesas do exercício seguinte.

Ativo imobilizado: Corresponde aos direitos que tenham por objeto bens corpóreos destinados à manutenção das atividades da companhia ou da empresa ou exercidos com essa finalidade, inclusive os decorrentes de operações que transfiram à companhia os benefícios, riscos e controle desses bens.

Ativo não circulante: É composto por ativo realizável a longo prazo, investimentos, imobilizado e intangível.

Ativo realizável a longo prazo: Corresponde aos direitos realizáveis após o término do exercício seguinte, assim como os derivados de vendas, adiantamentos ou empréstimos a sociedades coligadas ou controladas, diretores, acionistas ou participantes no lucro da companhia, que não constituírem negócios usuais na exploração do objeto da companhia.

Caixa: Dinheiro em caixa e depósitos à vista.

Capital social: É o valor previsto em contrato ou estatuto. Corresponde ao montante subscrito e, por dedução, a parcela ainda não realizada.

Despesas: São gastos incorridos para, direta ou indiretamente, gerar receitas. As despesas podem diminuir o ativo e/ou aumentar o passivo exigível, mas sempre provocam diminuições na situação líquida.

Duplicatas: Título de crédito cuja quitação prova o pagamento de obrigação oriunda de compra de mercadorias ou de recebimentos de serviços. É emitida pelo credor (vendedor da mercadoria) contra o devedor (comprador), pelo qual se deve ser remetida a este último para que a reconheça o valor devido.

Estoque: Representam os bens destinados à venda e que variam de acordo com a atividade da entidade.

Exigível a longo prazo: Até 04.12.2008, classificavam-se como exigibilidades com vencimento após o encerramento do exercício subsequente. A partir desta data, tais exigibilidades são denominadas "Passivo Não Circulante", no entanto, tal nomenclatura ainda é utilizada para fins de análise dos demonstrativos contábeis.

Intangível: Corresponde aos direitos que tenham por objeto bens incorpóreos destinados à manutenção da companhia ou exercidos com essa finalidade, inclusive o fundo de comércio adquirido.

Investimentos: Corresponde as participações permanentes em outras sociedades e os direitos de qualquer natureza, não classificáveis no ativo circulante, e que não se destinem à manutenção da atividade da companhia ou da empresa.

Lucro líquido: Corresponde ao resultado do exercício que remanescer depois de deduzidas as participações dos administradores e das partes beneficiárias.

Passivo circulante: Corresponde as obrigações da companhia, inclusive financiamentos para aquisição de direitos do ativo não circulante; quando se vencerem no exercício seguinte.

Passivo não circulante: Corresponde as obrigações da companhia, inclusive financiamentos para aquisição de direitos do ativo não circulante; se tiverem vencimento em prazo maior ao exercício seguinte.

Passivo: Obrigação presente da entidade, derivada de eventos já ocorridos,, cuja liquidação se espera resulte em saída de recursos capazes de gerar benefícios econômicos.

Patrimônio líquido: É composto pelos valores do capital social, reservas de capital, ajustes de avaliação patrimonial, reservas de lucros, ações em tesouraria e prejuízos acumulados.

Receitas: São entradas de elementos para o ativo da empresa, na forma de bens ou direitos que sempre provocam um aumento da situação líquida, correspondendo a um aumento de benefícios econômicos durante o período contábil na forma de entradas ou aumentos de ativos ou reduções de passivos que resultam em aumento no patrimônio líquido, com exceção daqueles relativos a contribuições de capital feitas por proprietários.

ANEXO 01: ANÁLISE DESCRITIVA DETALHADA

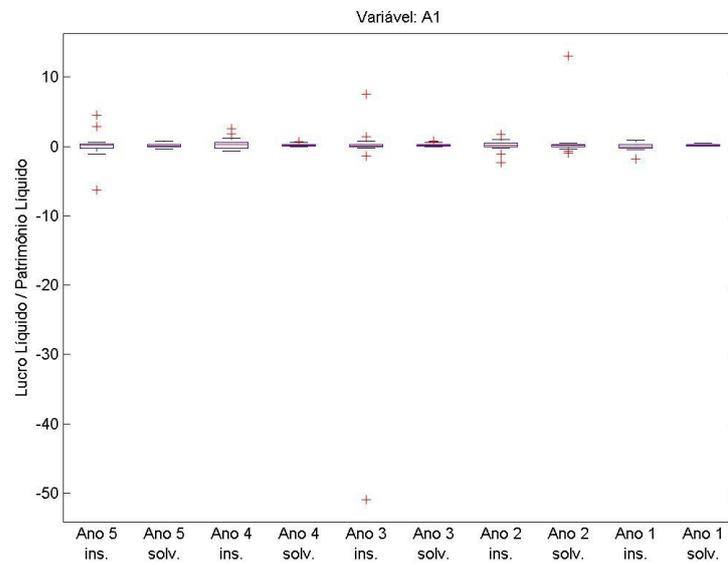


Figura 20. *Boxplot* referente à variável A1.

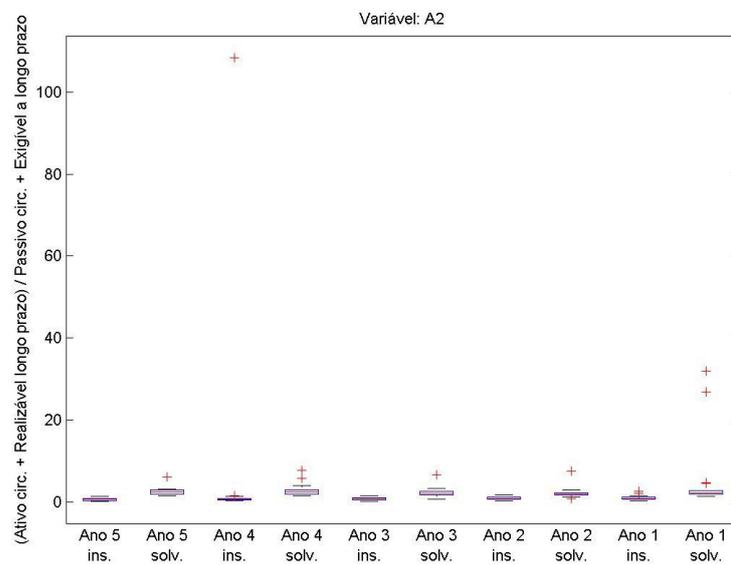


Figura 21. *Boxplot* referente à variável A2.

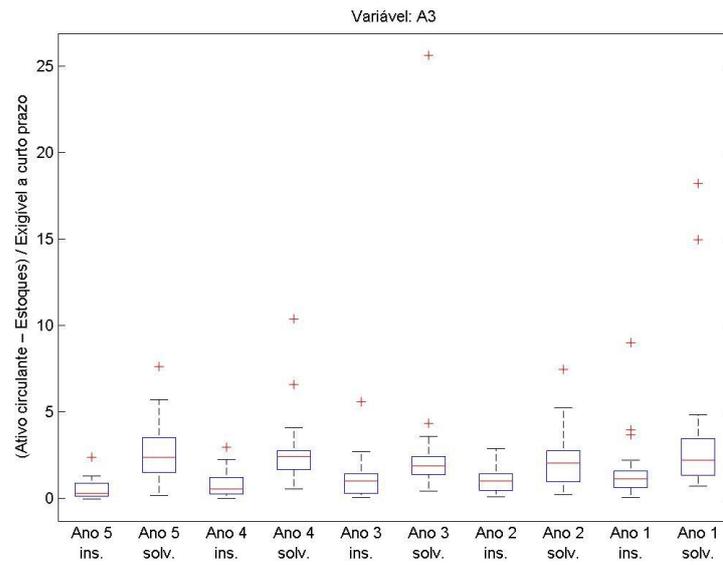


Figura 22. *Boxplot* referente à variável A3.

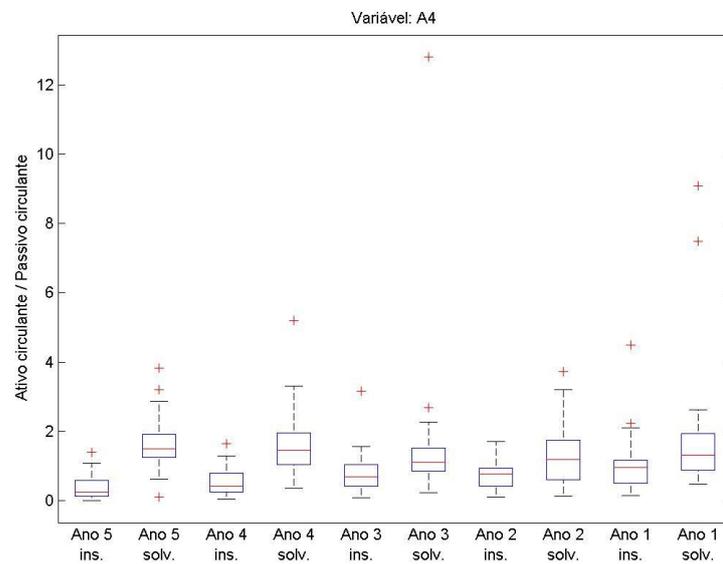


Figura 23. *Boxplot* referente à variável A4.

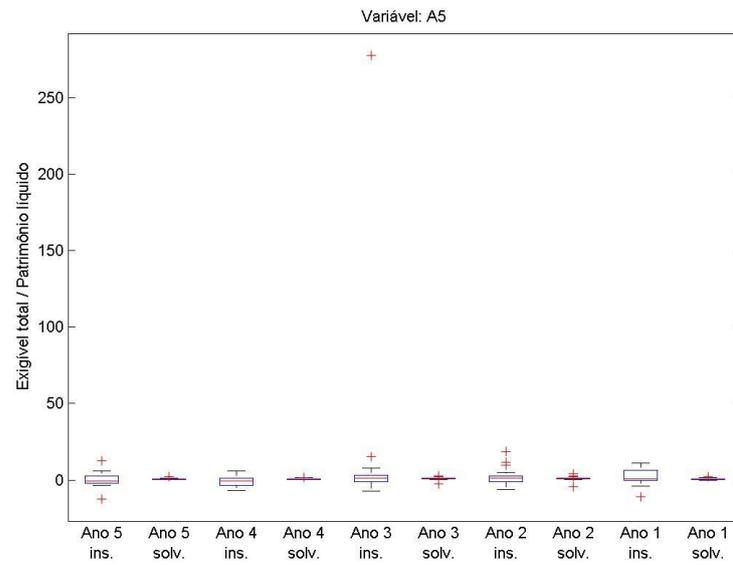


Figura 24. *Boxplot* referente à variável A5.

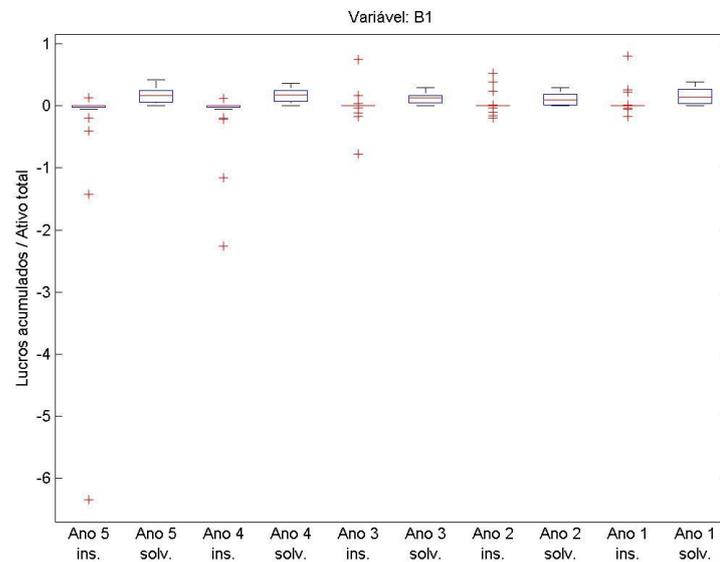


Figura 25. *Boxplot* referente à variável B1.

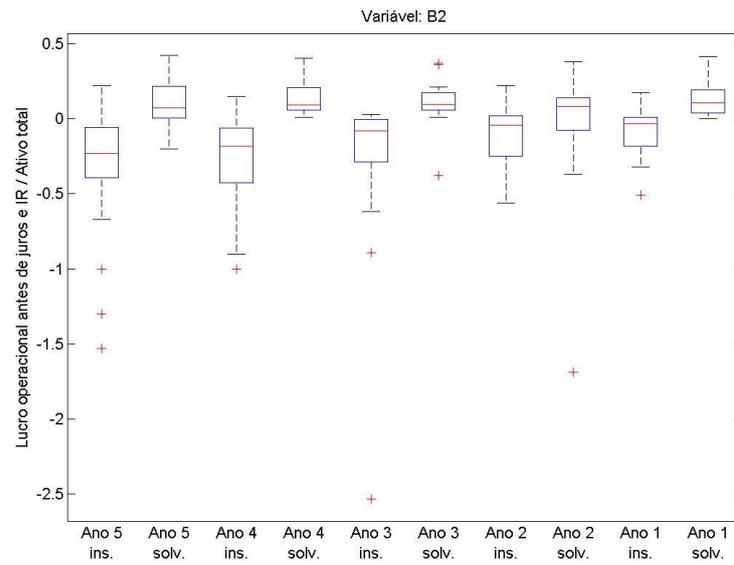


Figura 26. *Boxplot* referente à variável B2.

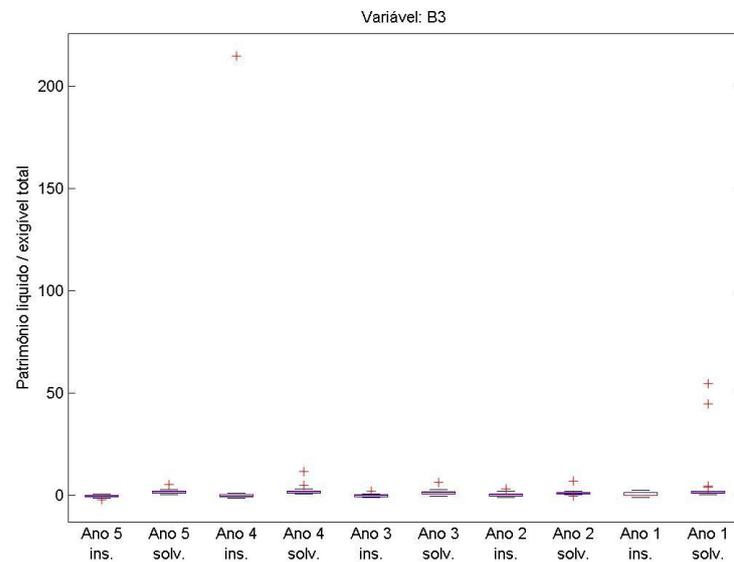


Figura 27. *Boxplot* referente à variável B3.

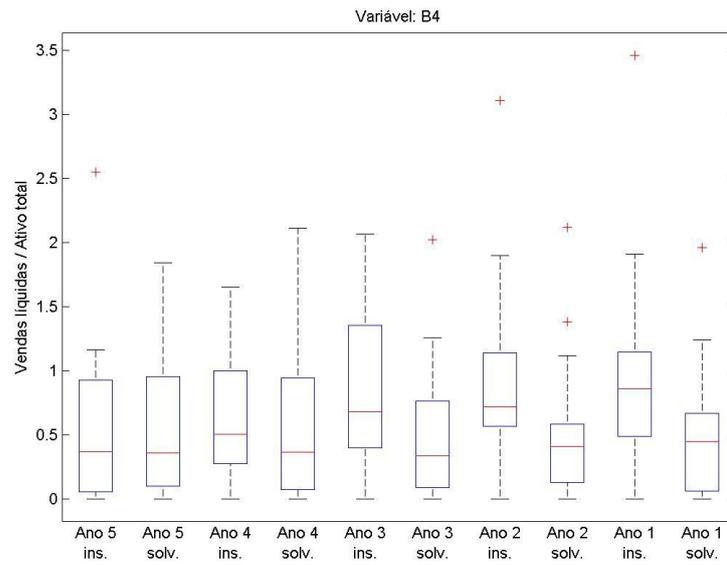


Figura 28. *Boxplot* referente à variável B4.

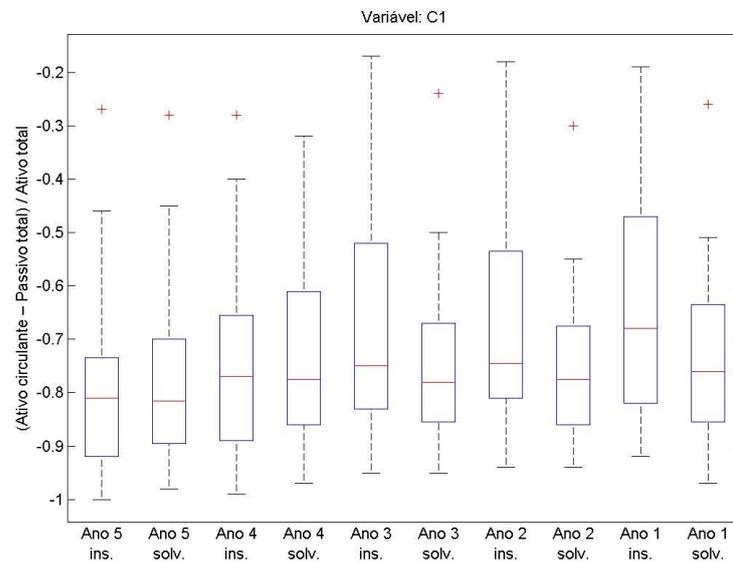


Figura 29. *Boxplot* referente à variável C1.

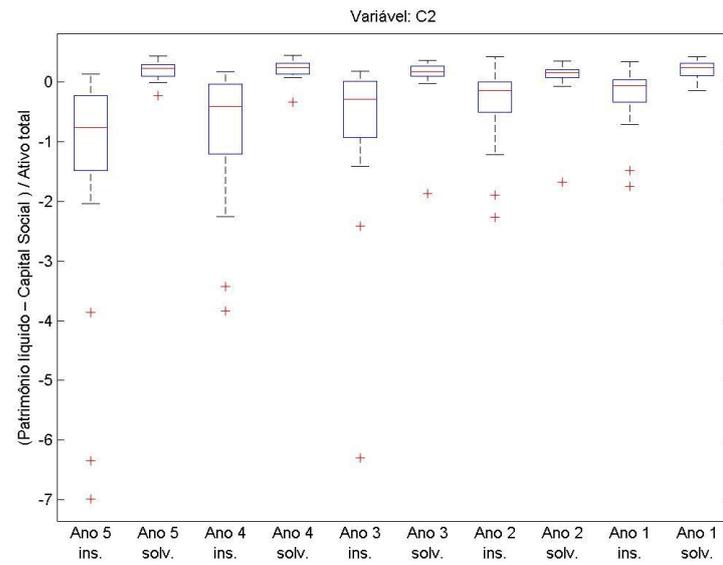


Figura 30. *Boxplot* referente à variável C2.

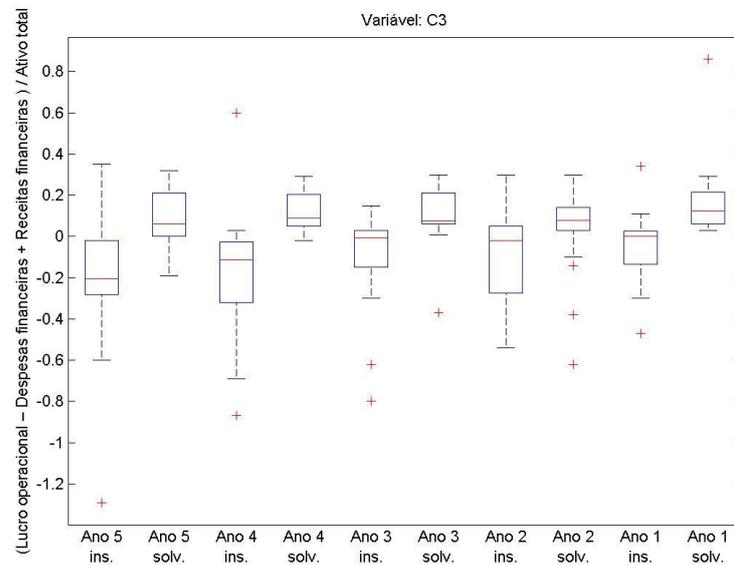


Figura 31. *Boxplot* referente à variável C3.

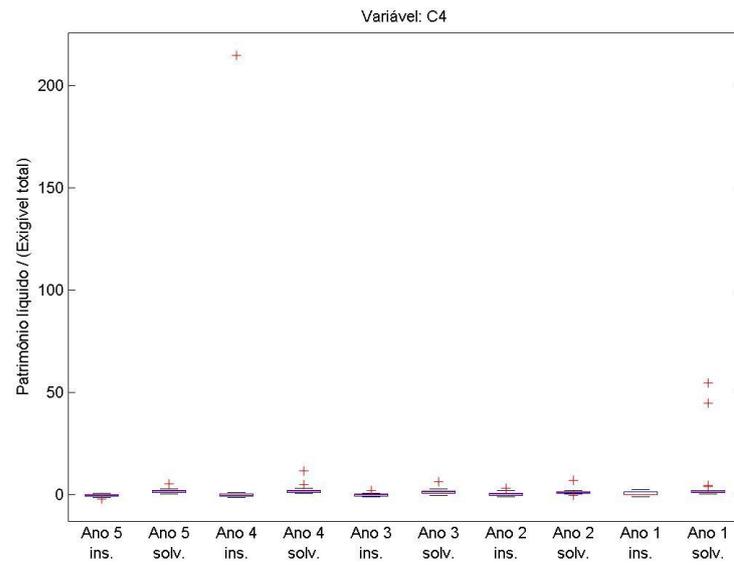


Figura 32. *Boxplot* referente à variável C4.

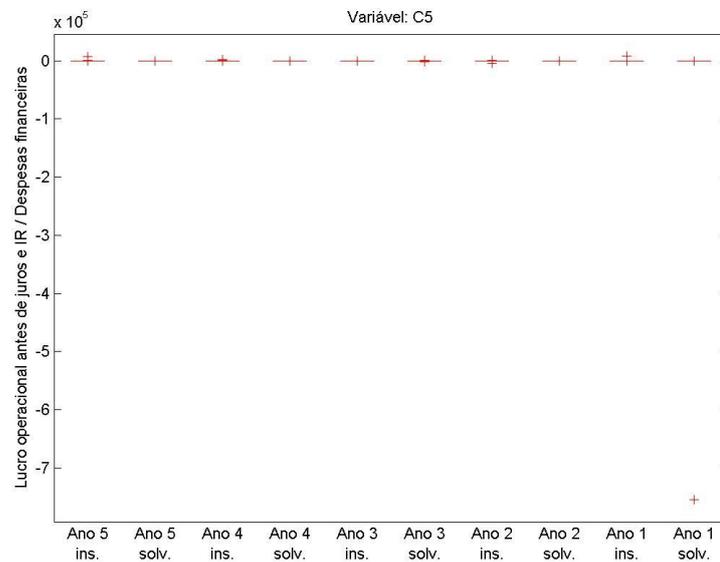


Figura 33. *Boxplot* referente à variável C5.

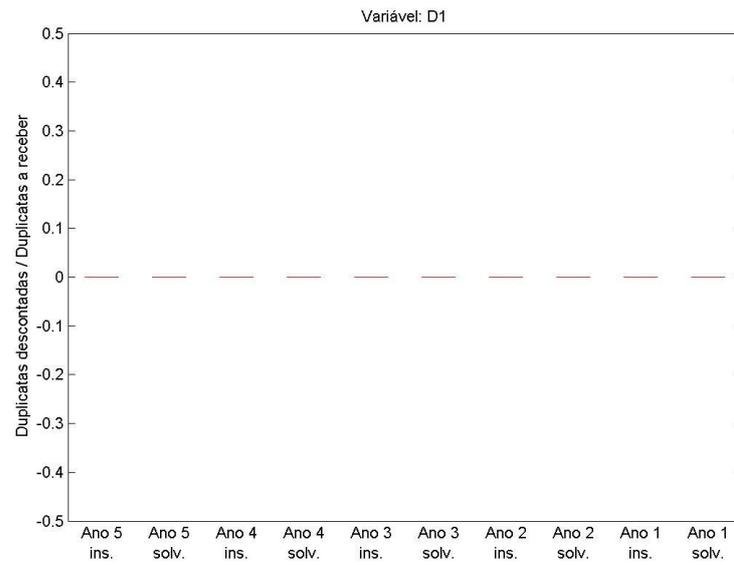


Figura 34. *Boxplot* referente à variável D1.

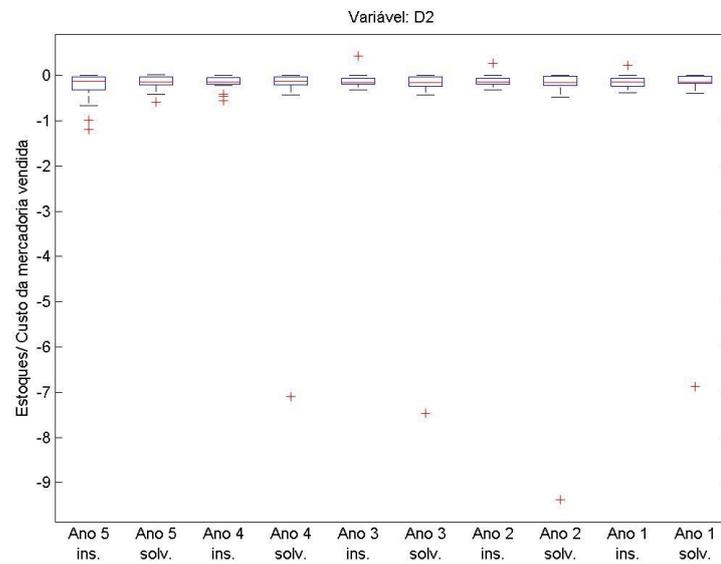


Figura 35. *Boxplot* referente à variável D2.

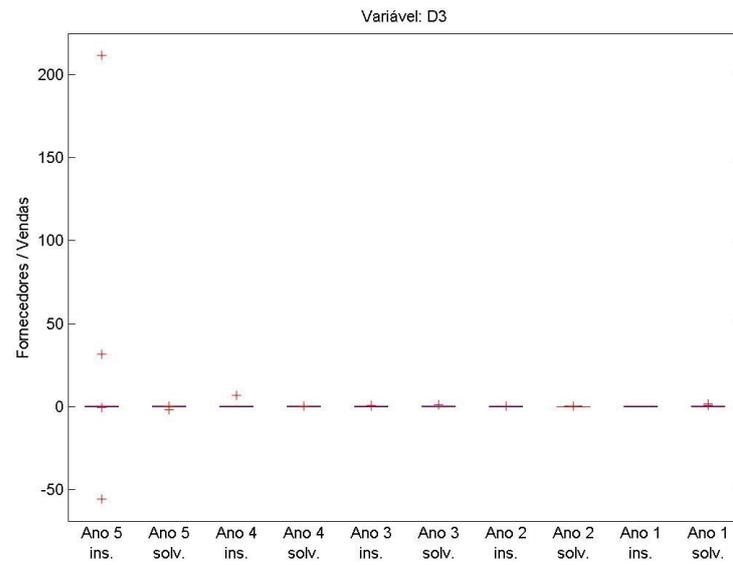


Figura 36. *Boxplot* referente à variável D3.

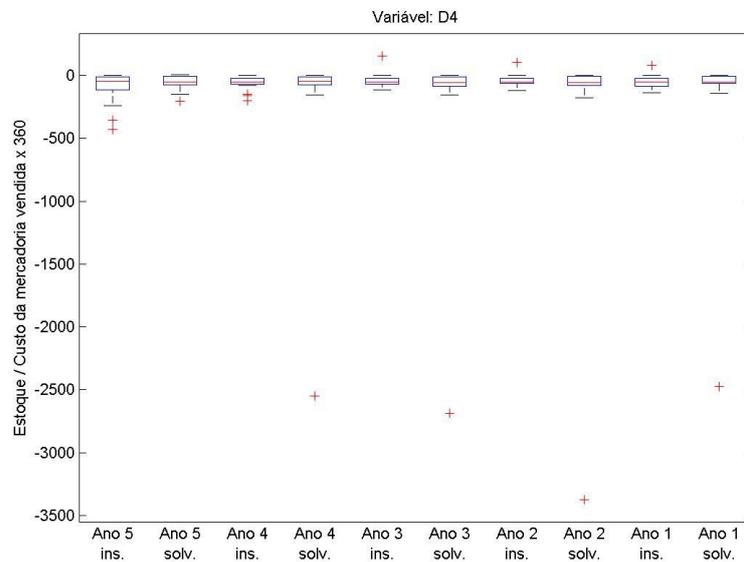


Figura 37. *Boxplot* referente à variável D4.

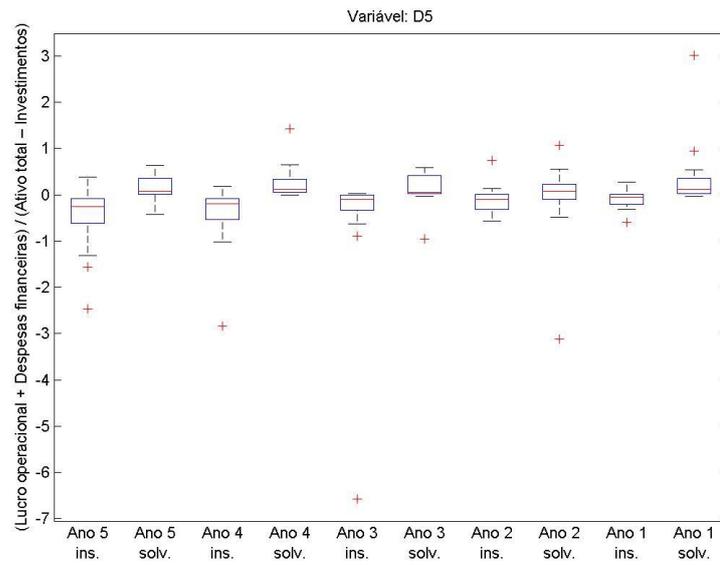


Figura 38. *Boxplot* referente à variável D5.

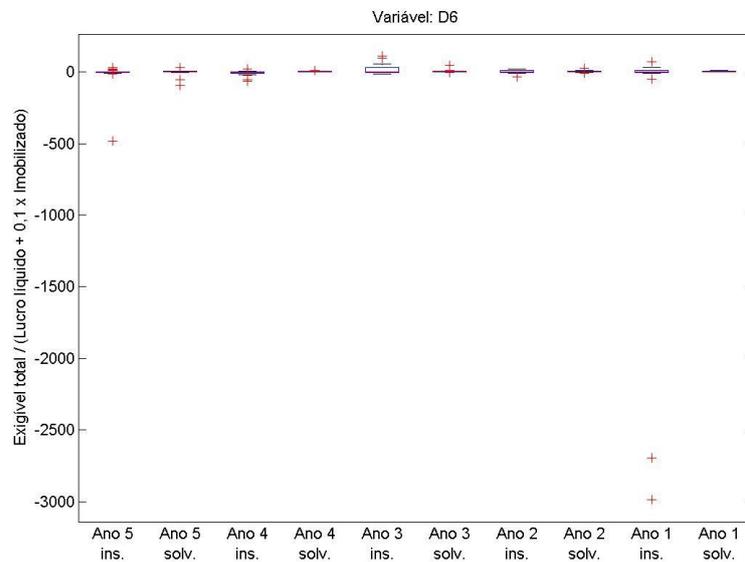


Figura 39. *Boxplot* referente à variável D6.

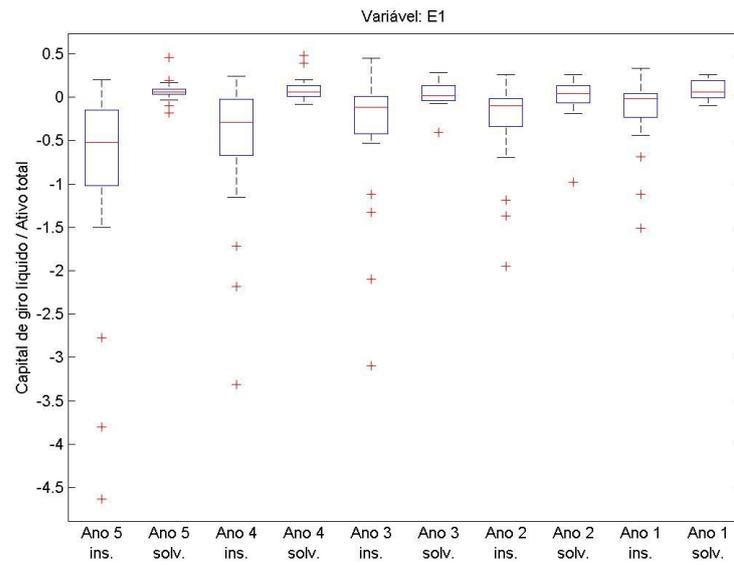


Figura 40. *Boxplot* referente à variável E1.

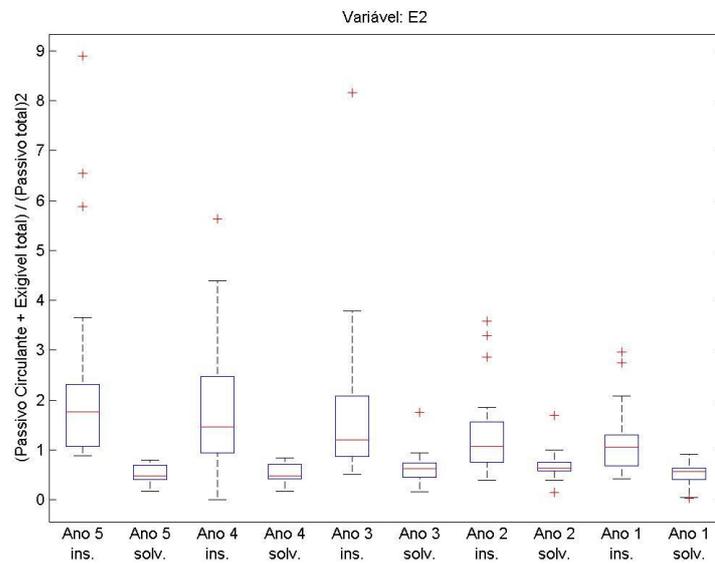


Figura 41. *Boxplot* referente à variável E2.

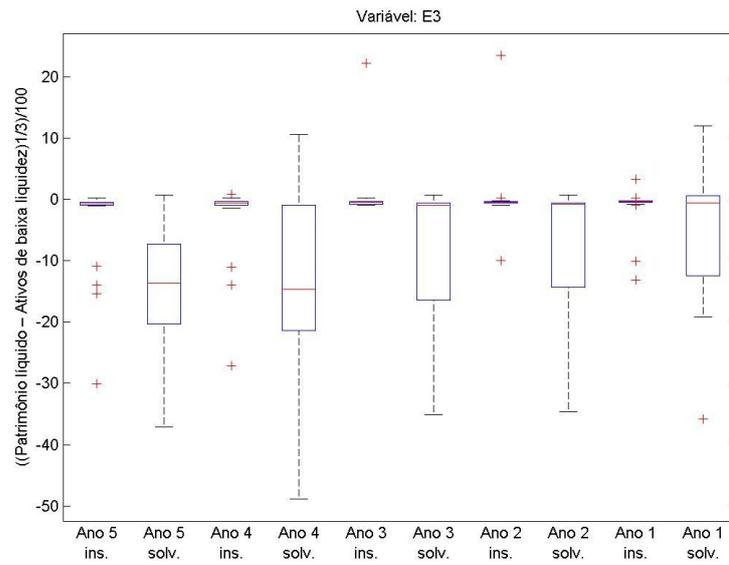


Figura 42. *Boxplot* referente à variável E3.

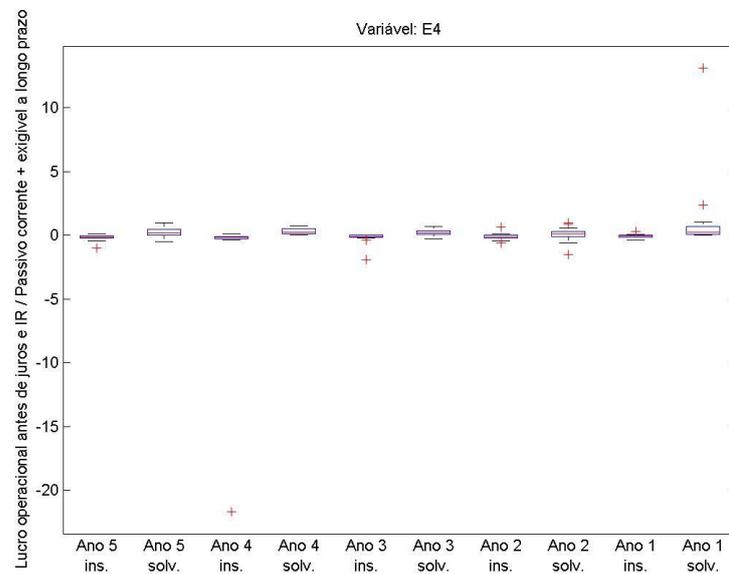


Figura 43. *Boxplot* referente à variável E4.

ANEXO 2: MATRIZ DE CORRELAÇÃO

Quadro 19. Matriz de correlação para os dados mais recentes (ano 5).

Ano 5	A1	A2	A3	A4	A5	B1	B2	B3	B4	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	D6	E1	E2	E3	E4		
A1	1,00																									
A2	0,15	1,00																								
A3	0,14	0,91	1,00																							
A4	0,13	0,90	0,99	1,00																						
A5	-0,40	-0,10	-0,09	-0,05	1,00																					
B1	0,14	0,41	0,32	0,31	0,00	1,00																				
B2	0,28	0,40	0,43	0,44	-0,01	0,47	1,00																			
B3	0,13	1,00	0,90	0,90	-0,10	0,39	0,38	1,00																		
B4	-0,05	-0,30	-0,32	-0,28	0,25	-0,09	-0,09	-0,28	1,00																	
C1	-0,09	-0,11	-0,05	0,00	0,18	-0,14	-0,03	-0,10	0,77	1,00																
C2	0,11	0,29	0,35	0,36	0,01	0,19	0,84	0,27	-0,19	-0,08	1,00															
C3	0,28	0,58	0,58	0,59	-0,02	0,51	0,92	0,57	-0,15	-0,02	0,70	1,00														
C4	0,13	1,00	0,90	0,90	-0,10	0,39	0,38	1,00	-0,28	-0,10	0,27	0,57	1,00													
C5	-0,08	-0,04	0,06	0,08	0,05	-0,16	-0,17	-0,01	0,18	0,24	-0,10	-0,12	-0,01	1,00												
D1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00											
D2	-0,06	0,03	0,06	0,06	0,06	-0,07	-0,07	0,05	0,15	0,09	-0,04	-0,05	0,05	-0,04	-	1,00										
D3	-0,02	-0,12	-0,14	-0,16	-0,06	-0,06	-0,13	-0,14	-0,21	-0,17	-0,12	-0,13	-0,14	0,10	-	-0,88	1,00									
D4	-0,06	0,03	0,06	0,06	0,06	-0,07	-0,07	0,05	0,15	0,09	-0,04	-0,05	0,05	-0,04	-	1,00	-0,88	1,00								
D5	0,20	0,34	0,25	0,24	-0,07	0,37	0,58	0,30	-0,26	-0,27	0,44	0,53	0,30	-0,87	-	-0,08	-0,05	-0,08	1,00							
D6	0,23	0,08	0,11	0,10	-0,40	-0,38	0,08	0,07	-0,23	-0,16	0,21	0,06	0,07	-0,04	-	-0,02	0,05	-0,02	0,09	1,00						
E1	-0,01	0,23	0,37	0,41	0,13	0,16	0,74	0,23	-0,03	0,21	0,89	0,64	0,23	-0,02	-	-0,01	-0,14	-0,01	0,31	0,03	1,00					
E2	-0,01	-0,39	-0,45	-0,46	-0,01	-0,33	-0,79	-0,38	0,34	0,18	-0,92	-0,70	-0,38	0,16	-	0,07	0,05	0,07	-0,50	-0,13	-0,87	1,00				
E3	-0,01	0,27	0,30	0,33	-0,06	0,19	0,12	0,28	0,23	0,49	0,03	0,16	0,28	0,14	-	0,02	-0,11	0,02	-0,07	-0,03	0,20	-0,10	1,00			
E4	0,19	0,86	0,79	0,78	-0,08	0,35	0,47	0,87	-0,22	-0,08	0,24	0,71	0,87	-0,04	-	0,03	-0,12	0,03	0,35	0,06	0,19	-0,32	0,22	1,00		

Quadro 20. Matriz de correlação para os dados do ano 4.

Ano 4	A1	A2	A3	A4	A5	B1	B2	B3	B4	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	D6	E1	E2	E3	E4	
A1	1,00																								
A2	-0,08	1,00																							
A3	-0,14	0,83	1,00																						
A4	-0,16	0,74	0,95	1,00																					
A5	-0,33	-0,06	0,01	0,10	1,00																				
B1	-0,10	0,24	0,14	0,12	-0,10	1,00																			
B2	-0,69	0,32	0,43	0,47	0,27	0,28	1,00																		
B3	-0,15	0,84	0,74	0,69	-0,06	0,29	0,41	1,00																	
B4	-0,12	-0,22	-0,10	0,00	0,22	-0,12	0,22	-0,16	1,00																
C1	-0,15	-0,14	0,09	0,20	0,23	-0,15	0,15	-0,14	0,83	1,00															
C2	-0,40	0,41	0,46	0,48	0,25	0,31	0,77	0,53	0,05	0,12	1,00														
C3	-0,32	0,40	0,50	0,54	0,23	0,33	0,87	0,53	0,17	0,11	0,84	1,00													
C4	-0,15	0,84	0,74	0,69	-0,06	0,29	0,41	1,00	-0,16	-0,14	0,53	0,53	1,00												
C5	0,01	-0,02	-0,02	0,03	-0,01	0,07	-0,10	0,02	0,18	0,22	-0,09	-0,09	0,02	1,00											
D1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00										
D2	0,04	-0,06	0,09	0,08	0,03	-0,13	-0,08	-0,03	0,17	0,15	-0,08	-0,07	-0,03	-0,04	-	1,00									
D3	-0,17	-0,33	-0,21	-0,21	-0,10	-0,22	-0,16	-0,33	-0,03	0,06	-0,36	-0,34	-0,33	0,19	-	0,15	1,00								
D4	0,04	-0,06	0,09	0,08	0,03	-0,13	-0,08	-0,03	0,17	0,15	-0,08	-0,07	-0,03	-0,04	-	1,00	0,15	1,00							
D5	-0,75	0,45	0,49	0,50	0,21	0,29	0,94	0,54	0,12	0,05	0,67	0,77	0,54	-0,07	-	-0,11	-0,08	-0,11	1,00						
D6	0,03	0,07	0,21	0,22	-0,14	0,21	0,22	0,12	0,08	0,08	0,24	0,32	0,12	-0,14	-	-0,05	-0,27	-0,05	0,14	1,00					
E1	-0,32	0,38	0,53	0,59	0,31	0,21	0,64	0,51	0,11	0,28	0,92	0,73	0,51	-0,06	-	-0,03	-0,39	-0,03	0,53	0,28	1,00				
E2	0,19	-0,50	-0,53	-0,54	-0,15	-0,33	-0,54	-0,67	0,15	0,01	-0,91	-0,71	-0,67	0,09	-	0,09	0,40	0,09	-0,48	-0,25	-0,90	1,00			
E3	-0,01	-0,13	0,03	0,05	-0,01	-0,02	-0,12	-0,03	0,40	0,53	-0,17	-0,16	-0,03	0,08	-	0,06	0,19	0,06	-0,10	0,03	-0,01	0,12	1,00		
E4	-0,46	0,53	0,57	0,59	0,12	0,34	0,88	0,56	0,16	0,03	0,57	0,82	0,56	-0,07	-	-0,06	-0,20	-0,06	0,89	0,17	0,43	-0,39	-0,10	1,00	

Quadro 21. Matriz de correlação para os dados do ano 3

Ano 3	A1	A2	A3	A4	A5	B1	B2	B3	B4	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	D6	E1	E2	E3	E4		
A1	1,00																									
A2	0,05	1,00																								
A3	0,03	0,40	1,00																							
A4	0,02	0,40	0,99	1,00																						
A5	-0,99	-0,08	-0,05	-0,03	1,00																					
B1	0,02	0,32	0,04	0,05	-0,04	1,00																				
B2	-0,12	0,38	0,12	0,14	0,03	0,27	1,00																			
B3	0,07	0,94	0,34	0,34	-0,10	0,34	0,41	1,00																		
B4	-0,23	-0,39	-0,20	-0,19	0,23	-0,26	-0,08	-0,35	1,00																	
C1	-0,13	-0,25	0,03	0,06	0,13	-0,25	0,08	-0,24	0,75	1,00																
C2	-0,03	0,43	0,18	0,20	0,03	0,19	0,52	0,52	-0,30	-0,02	1,00															
C3	-0,14	0,44	0,11	0,14	0,13	0,47	0,64	0,51	-0,21	-0,08	0,78	1,00														
C4	0,07	0,94	0,34	0,34	-0,10	0,34	0,41	1,00	-0,35	-0,24	0,52	0,51	1,00													
C5	-0,04	-0,18	0,00	-0,02	0,03	-0,11	-0,19	-0,17	-0,04	-0,07	-0,12	-0,25	-0,17	1,00												
D1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00											
D2	-0,02	-0,15	0,00	0,01	0,03	-0,15	-0,09	-0,14	0,17	0,07	-0,07	-0,11	-0,14	0,00	-	1,00										
D3	0,02	-0,13	-0,14	-0,16	-0,03	0,02	-0,04	-0,16	-0,17	-0,15	-0,08	-0,09	-0,16	0,11	-	-0,76	1,00									
D4	-0,02	-0,15	0,00	0,01	0,03	-0,16	-0,09	-0,14	0,17	0,07	-0,07	-0,11	-0,14	0,00	-	1,00	-0,76	1,00								
D5	-0,12	0,34	0,09	0,11	0,01	0,16	0,96	0,35	-0,02	0,09	0,32	0,43	0,35	-0,13	-	-0,10	0,03	-0,10	1,00							
D6	0,08	0,01	0,21	0,21	-0,07	-0,07	0,13	0,03	0,09	0,20	0,24	0,08	0,03	0,09	-	0,07	-0,19	0,07	0,11	1,00						
E1	-0,02	0,40	0,26	0,30	0,04	0,17	0,37	0,50	-0,29	0,10	0,91	0,66	0,50	-0,11	-	-0,06	-0,22	-0,06	0,16	0,26	1,00					
E2	0,00	-0,50	-0,24	-0,26	-0,01	-0,27	-0,44	-0,60	0,45	0,14	-0,96	-0,74	-0,60	0,10	-	0,09	0,11	0,09	-0,24	-0,20	-0,94	1,00				
E3	-0,03	-0,18	0,29	0,30	0,02	-0,08	-0,05	-0,14	0,30	0,54	-0,12	-0,09	-0,14	0,05	-	0,01	-0,02	0,01	-0,03	0,08	0,02	0,11	1,00			
E4	-0,06	0,50	0,08	0,09	-0,03	0,30	0,92	0,50	0,00	0,05	0,35	0,58	0,50	-0,31	-	-0,13	-0,06	-0,13	0,91	0,00	0,21	-0,28	-0,10	1,00		

Quadro 22. Matriz de correlação para os dados do ano 2

Ano 2	A1	A2	A3	A4	A5	B1	B2	B3	B4	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	D6	E1	E2	E3	E4	
A1	1,00																								
A2	-0,15	1,00																							
A3	-0,14	0,02	1,00																						
A4	-0,15	-0,01	0,97	1,00																					
A5	-0,64	0,03	0,18	0,22	1,00																				
B1	-0,18	0,03	0,23	0,28	0,20	1,00																			
B2	-0,29	-0,03	0,39	0,46	0,35	0,52	1,00																		
B3	-0,15	1,00	0,01	-0,02	0,02	0,02	-0,04	1,00																	
B4	0,12	-0,22	-0,22	-0,18	0,10	-0,21	-0,08	-0,20	1,00																
C1	0,23	-0,22	0,32	0,36	-0,05	-0,11	0,16	-0,21	0,58	1,00															
C2	-0,21	0,06	0,31	0,38	0,37	0,42	0,90	0,05	-0,19	0,06	1,00														
C3	-0,35	-0,02	0,32	0,39	0,25	0,53	0,91	-0,03	-0,06	0,10	0,78	1,00													
C4	-0,15	1,00	0,01	-0,02	0,02	0,02	-0,04	1,00	-0,20	-0,21	0,05	-0,03	1,00												
C5	-0,07	0,00	-0,12	-0,17	0,13	-0,07	-0,24	0,00	-0,11	-0,19	-0,06	-0,34	0,00	1,00											
D1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00										
D2	-0,02	0,04	0,13	0,12	-0,01	-0,12	-0,10	0,04	0,18	0,15	-0,11	-0,08	0,04	0,02	-	1,00									
D3	0,07	-0,06	-0,18	-0,21	-0,10	-0,02	-0,25	-0,06	-0,18	-0,16	-0,32	-0,28	-0,06	0,01	-	0,01	1,00								
D4	-0,02	0,04	0,13	0,12	-0,01	-0,12	-0,10	0,04	0,18	0,15	-0,11	-0,08	0,04	0,02	-	1,00	0,01	1,00							
D5	-0,24	0,02	0,29	0,35	0,24	0,33	0,83	0,00	-0,23	0,00	0,76	0,64	0,00	-0,28	-	-0,20	-0,11	-0,20	1,00						
D6	0,16	0,06	0,23	0,21	-0,23	0,13	0,14	0,04	0,11	0,15	0,07	0,16	0,04	-0,10	-	-0,04	0,00	-0,04	0,11	1,00					
E1	-0,15	0,10	0,47	0,53	0,24	0,41	0,75	0,09	-0,16	0,22	0,82	0,79	0,09	-0,11	-	-0,06	-0,46	-0,06	0,45	0,08	1,00				
E2	0,23	-0,21	-0,46	-0,51	-0,32	-0,50	-0,83	-0,19	0,34	0,01	-0,92	-0,79	-0,19	0,07	-	0,11	0,33	0,11	-0,65	-0,10	-0,92	1,00			
E3	0,04	0,27	0,06	0,04	-0,10	-0,16	-0,19	0,28	0,22	0,30	-0,18	-0,14	0,28	-0,06	-	0,05	0,11	0,05	-0,17	-0,06	-0,14	0,17	1,00		
E4	0,12	-0,99	0,08	0,12	0,02	0,04	0,14	-0,99	0,19	0,23	0,03	0,12	-0,99	-0,04	-	-0,05	0,03	-0,05	0,09	-0,01	-0,01	0,11	-0,29	1,00	

Quadro 23. Matriz de correlação para os dados do ano 1

Ano 1	A1	A2	A3	A4	A5	B1	B2	B3	B4	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	D6	E1	E2	E3	E4		
A1	1,00																									
A2	0,04	1,00																								
A3	-0,08	0,73	1,00																							
A4	-0,08	0,76	0,96	1,00																						
A5	-0,75	0,06	0,14	0,15	1,00																					
B1	0,01	0,30	0,20	0,23	0,08	1,00																				
B2	-0,16	0,48	0,40	0,46	0,13	0,27	1,00																			
B3	-0,05	0,96	0,70	0,75	0,16	0,40	0,60	1,00																		
B4	-0,18	-0,23	-0,09	-0,03	0,20	-0,53	0,24	-0,20	1,00																	
C1	-0,09	-0,07	0,12	0,21	0,18	-0,22	0,34	-0,02	0,78	1,00																
C2	-0,09	0,51	0,39	0,45	0,23	0,60	0,70	0,65	-0,16	0,05	1,00															
C3	-0,21	0,39	0,31	0,37	0,10	0,27	0,77	0,47	0,16	0,29	0,27	1,00														
C4	-0,05	0,96	0,70	0,75	0,16	0,40	0,60	1,00	-0,20	-0,02	0,65	0,47	1,00													
C5	-0,13	-0,06	-0,11	-0,14	0,14	0,01	-0,07	-0,07	-0,16	-0,14	0,03	-0,12	-0,07	1,00												
D1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00											
D2	-0,25	0,17	0,18	0,08	0,47	-0,10	0,05	0,16	0,15	-0,05	0,25	-0,21	0,16	0,13	-	1,00										
D3	0,04	-0,16	-0,13	-0,15	-0,03	0,01	-0,57	-0,24	-0,13	-0,10	-0,39	-0,23	-0,24	-0,02	-	0,12	1,00									
D4	-0,25	0,17	0,18	0,07	0,47	-0,10	0,05	0,16	0,15	-0,05	0,25	-0,21	0,16	0,13	-	1,00	0,12	1,00								
D5	-0,09	0,54	0,37	0,43	0,09	0,20	0,94	0,64	0,23	0,25	0,67	0,64	0,64	-0,25	-	0,09	-0,63	0,09	1,00							
D6	0,04	0,09	0,06	0,08	-0,08	-0,03	-0,01	0,02	0,04	0,06	-0,03	0,02	0,02	0,02	-	-0,04	0,02	-0,04	-0,01	1,00						
E1	-0,13	0,49	0,45	0,51	0,19	0,39	0,81	0,62	0,03	0,18	0,93	0,33	0,62	0,03	-	0,20	-0,57	0,20	0,80	-0,01	1,00					
E2	0,08	-0,56	-0,46	-0,51	-0,19	-0,48	-0,72	-0,69	0,12	-0,03	-0,97	-0,26	-0,69	-0,04	-	-0,27	0,45	-0,27	-0,72	0,04	-0,97	1,00				
E3	-0,09	-0,39	-0,17	-0,19	0,10	-0,12	-0,24	-0,34	0,25	0,32	-0,24	-0,16	-0,34	-0,01	-	0,06	0,12	0,06	-0,27	-0,08	-0,25	0,30	1,00			
E4	-0,02	0,52	0,33	0,42	0,05	0,18	0,77	0,57	0,32	0,46	0,35	0,82	0,57	-0,17	-	-0,08	-0,17	-0,08	0,69	0,07	0,39	-0,33	-0,16	1,00		

ANEXO 3: GRÁFICOS DE AVALIAÇÃO DE NORMALIDADE

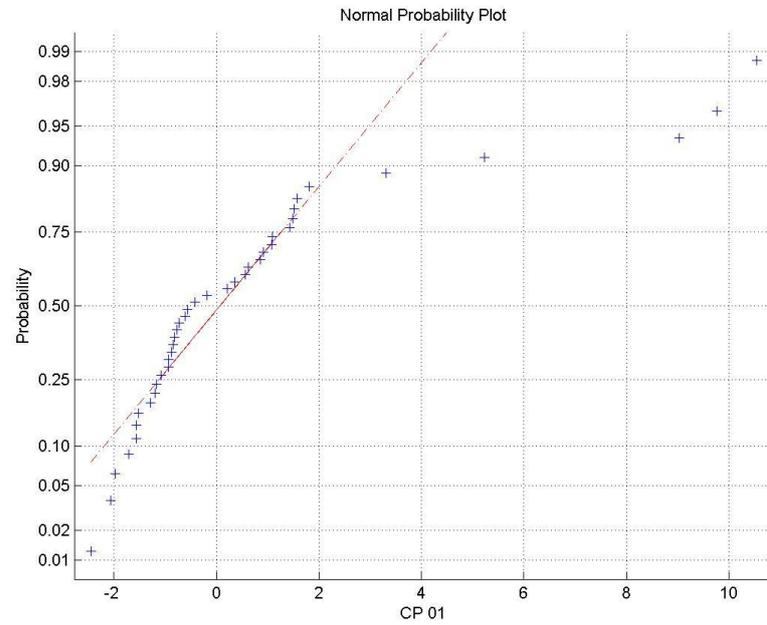


Figura 44. Avaliação de normalidade CP01.

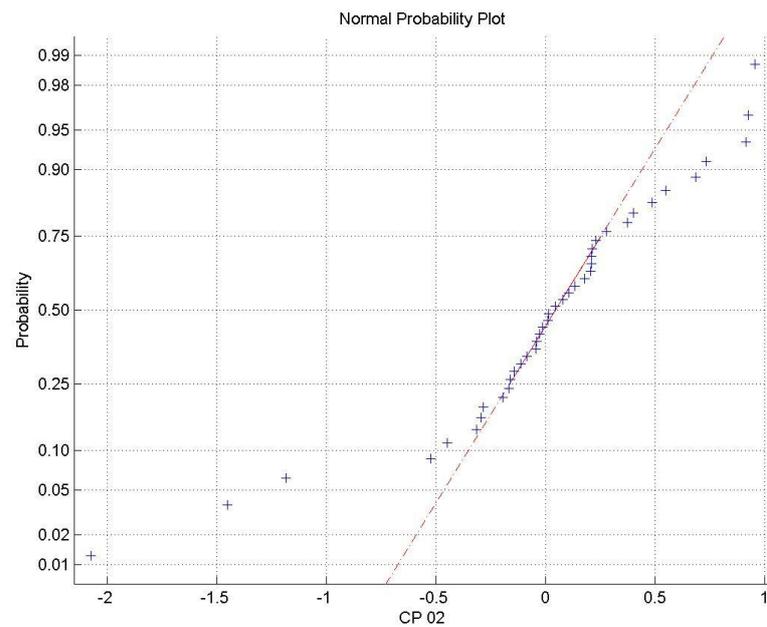


Figura 45. Avaliação de normalidade CP02.

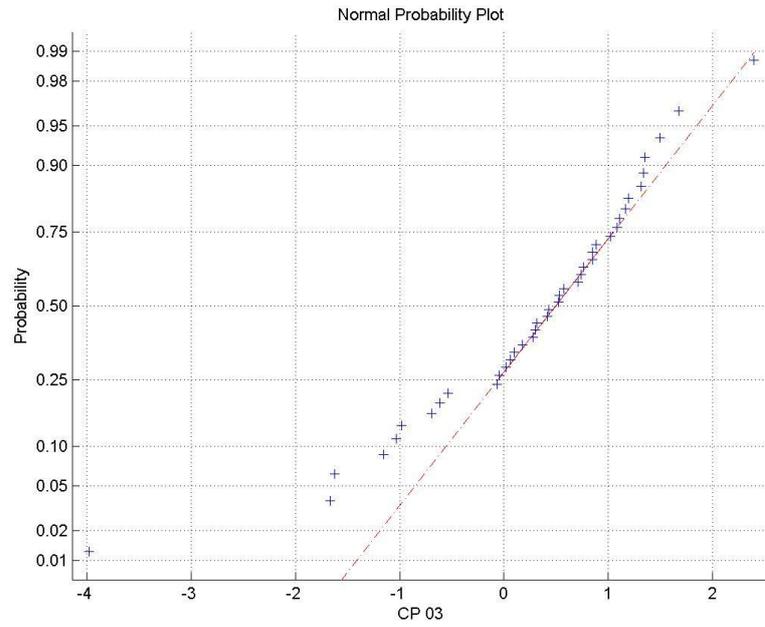


Figura 46. Avaliação de normalidade CP03.

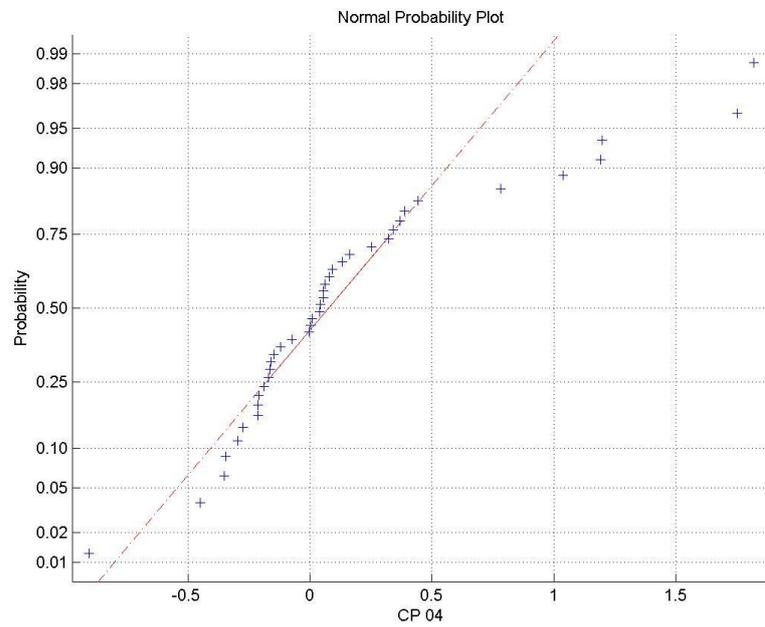


Figura 47. Avaliação de normalidade CP04.

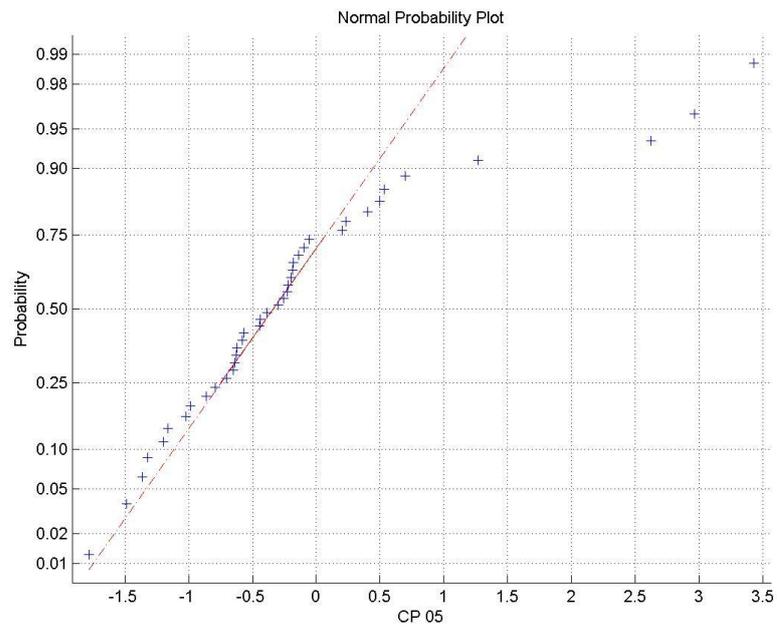


Figura 48. Avaliação de normalidade CP05.

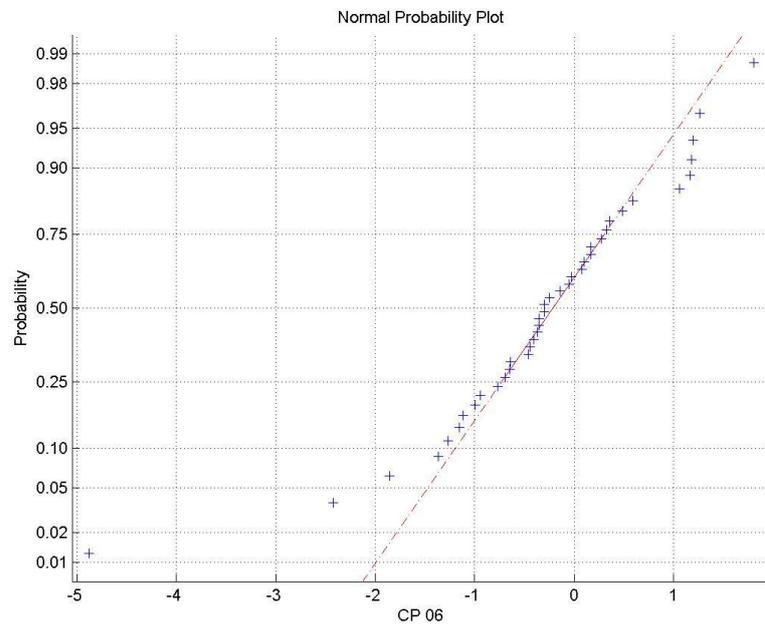


Figura 49. Avaliação de normalidade CP06.

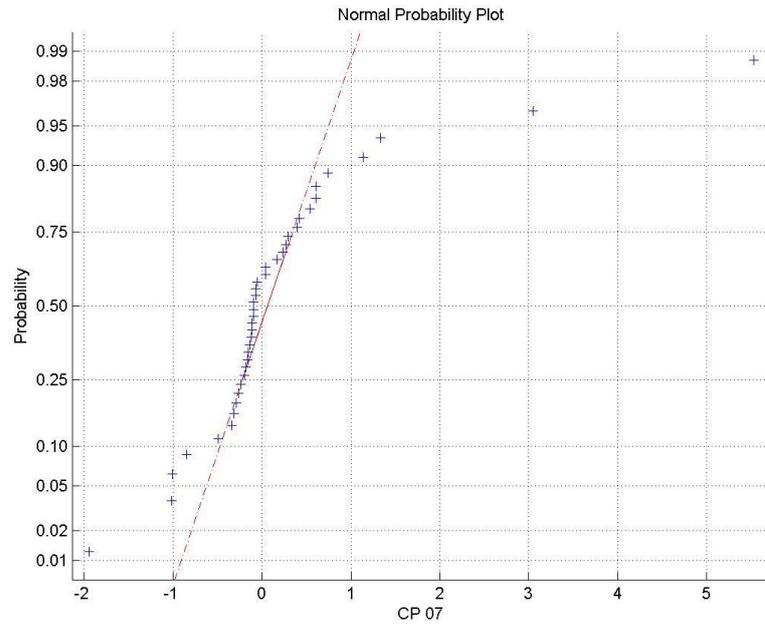


Figura 50. Avaliação de normalidade CP07.

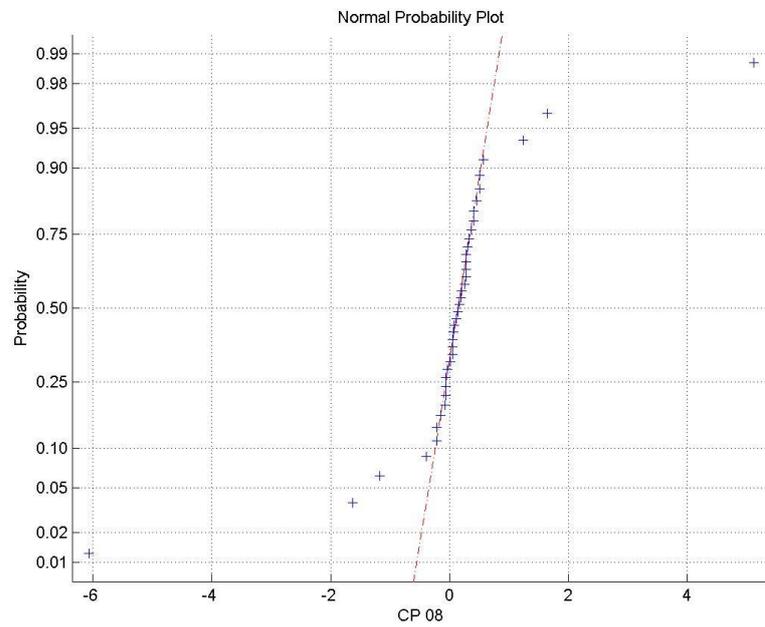


Figura 51. Avaliação de normalidade CP08.

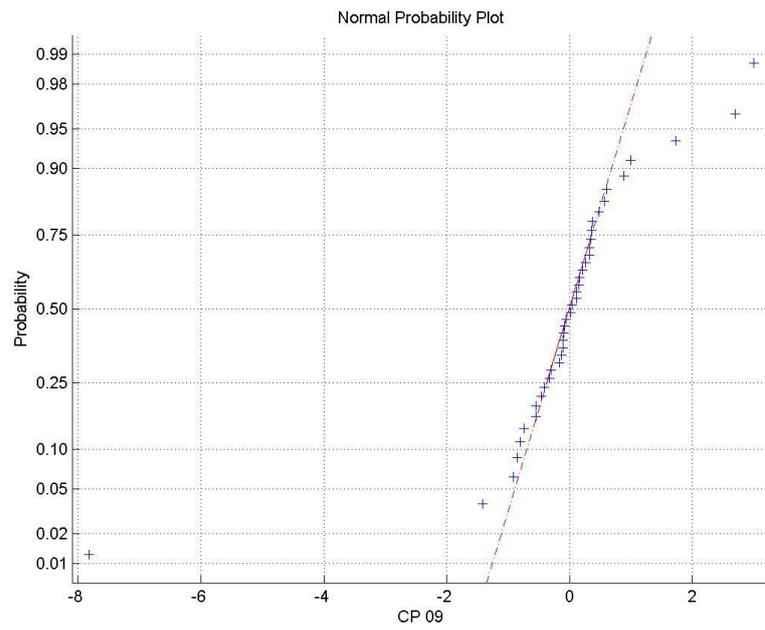


Figura 52. Avaliação de normalidade CP09.

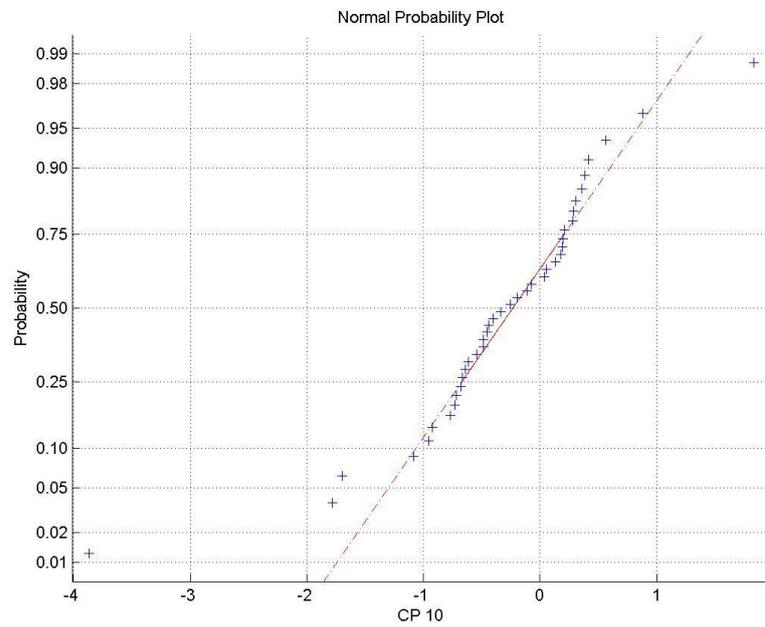


Figura 53. Avaliação de normalidade CP10.