

## Análise de Cooperativas de Crédito Paranaenses por meio da Análise das Componentes Principais e Método de Agrupamento Hierárquico

Bruno Samways dos Santos (Universidade Tecnológica do Paraná, UTFPR, Câmpus Londrina; Pontifícia Universidade Católica do Paraná, PUCPR) [brunosantos@utfpr.edu.br](mailto:brunosantos@utfpr.edu.br)  
Maria Teresinha Arns Steiner (Pontifícia Universidade Católica do Paraná, PUCPR) [maria.steiner@pucpr.br](mailto:maria.steiner@pucpr.br)  
Pedro José Steiner Neto (Universidade Positivo) [pedrosteiner@up.edu.br](mailto:pedrosteiner@up.edu.br)  
Luís Guilherme Ribeiro Martins (Banco do Brasil) [lgrmlgm@hotmail.com](mailto:lgrmlgm@hotmail.com)

### Resumo:

As cooperativas de crédito buscam levar o acesso ao crédito às diferentes regiões do Brasil, visando melhor atender aos cooperados e, conseqüentemente, a sociedade. As instituições financeiras têm procurado cada vez mais orientar suas decisões para técnicas matemáticas e estatísticas em conjunto com o aprendizado de máquina, pelo fato de serem objetivas e precisas. Neste contexto, o presente estudo tem como objetivo mostrar a aplicação de técnicas estatísticas e de aprendizado não-supervisionado em cooperativas de crédito paranaenses. A amostra inicial consistia em 45 unidades (singulares) de cooperativas pertencentes a três diferentes sistemas: Sicoob-Unicoob; Uniprime e Sicredi. Com o auxílio do *software* SPSS, realizou-se a eliminação de *outliers* e a análise dos componentes principais, reduzindo assim o tamanho da amostra para 39 unidades e cinco componentes. Na sequência, foi realizado o agrupamento hierárquico e então os testes de Kruskal-Wallis e U de Mann-Whitney para posteriores análises dos grupos (*clusters*) formados. Os resultados mostraram a formação dos grupos, apontando os aspectos fortes e fracos das cooperativas pertencentes a um mesmo grupo.

**Palavras chave:** Cooperativa de crédito, Agrupamentos, Análise dos Componentes Principais.

## Analysis of Parana's Credit Union through Principal Component Analysis and Hierarchical Clustering Method

### Abstract

The Credit union seek to allow credit access to the different regions of Brazil, in order to better serve the Credit union and, consequently, the society. Financial institutions have increasingly sought to guide their decisions to mathematical and statistical techniques with machine learning, because they are direct and accurate. In this context, the present study aims to show the application of statistical techniques and unsupervised learning in Parana's Credit union. The initial sample consisted of 45 units of Credit union belonging to three different systems: Sicoob-Unicoob; Uniprime and Sicredi. Using SPSS software, the outliers detection and analysis of the main components was carried out, reducing the sample size to 39 units and five components. After, the hierarchical clustering and then the Kruskal-Wallis and Mann-Whitney U tests were performed for further analyzes of the clusters formed. The results showed the clusters composition, emphasizing the credit union strengths and the weakness aspects belonging to a same cluster.

**Key-words:** Credit union, Clustering, Principal Component Analysis.

### 1. Introdução

A crescente complexidade e competitividade nas organizações deixa evidente a necessidade de se buscar soluções a partir da interação entre julgamentos (humanos) e métodos

(computacionais) para que as mesmas consigam se manter ativas e atualizadas face às rápidas mudanças do mercado.

Jacques e Gonçalves (2016) destacam que o crédito é instrumento de crescimento regional e nacional e todas as pessoas deveriam ter acesso a este, entretanto, por inviabilidade econômica ou outros, vários bancos optam por não instalar agências bancárias em vários municípios brasileiros. Neste contexto, surge como alternativa a constituição de cooperativas de crédito.

Para Bittencourt et al. (2017), as cooperativas de crédito buscam melhorar o acesso ao crédito para diversas classes sociais, trabalhando em prol do crescimento econômico, mas principalmente em benefício dos cooperados e assim ganham relevância no desenvolvimento regional. De acordo com o Banco Central do Brasil (2015), apesar do grande potencial inclusivo, as cooperativas de crédito ainda tem espaços significativos para ocupar (em termos regionais, principalmente) e também estimulando competitividade e reduzindo custos do Sistema Financeiro Nacional.

As decisões das cooperativas são tomadas em assembleias, não por acionistas, mas sim por cooperados. As cooperativas também são instituições financeiras, trabalham de forma similar aos bancos e caixas econômicas, desenvolvendo serviços como operações de pagamento e recebimento, empréstimos, aplicações em fundos de investimenots, entre outros (PINHEIRO, 2008).

O Sistema OCEPAR – Sindicato e Organizações das Cooperativas do Estado do Paraná - OCEPAR (2017), foi a primeira organização estadual a buscar representação sindical patronal no Paraná. A mesma afirma que “representa as cooperativas nas negociações e desenvolvimento de atividades que visam aprimorar a relação capital e trabalho em seus diferentes ramos do cooperativismo” (OCEPAR, 2017). O Sistema Ocepar possui três sistemas de cooperativas de crédito: Sistema Sicredi, Sistema Sicoob-Unicoobi e Sistema Uniprime.

O Sistema Sicredi (SICREDI, 2017) é uma instituição financeira cooperativa com 117 cooperativas de crédito filiadas com uma rede de operação que supera 1500 agências ao todo. O método participativo abrange mais de 3,5 milhões de associados que tomam as decisões da cooperativa de crédito. Também possui 21,8 mil colaboradores, 70,1 bilhões de ativos, 11,4 bilhões em patrimônio líquido e 37,8 bilhões de saldo carteira de crédito.

O Sistema Sicoob-Unicoob (SICOOB-UNICOOB, 2017) está presente em 140 municípios do Paraná, Santa Catarina, Pará e Amapá. O sistema surgiu em 2014 após incorporações de cooperativas fora do Paraná e possui mais de 3 mil colaboradores, 20 cooperativas financeiras, 237 pontos de atendimento e 240 mil associados. O sistema também é composto por mais cinco empresas coligadas que oferecem suporte às cooperativas e cooperados: Instituto Sicoob, Fundo Garantidor de Liquidez, Sancor Seguros do Brasil, Unicoob Administradora de Consórcios e Unicoob Corretora de Seguros.

O Sistema Uniprime (UNIPRIME, 2017) possui nove cooperativas singulares (1º grau) que atuam nos estados do Paraná, Mato Grosso do Sul e São Paulo que correspondem à cooperativa de crédito Uniprime Central (2º grau). A Uniprime Central foi fundada em 1998 por um grupo de profissionais da área médica e posteriormente estendeu seus serviços para diversos ramos de atividade. O sistema possui 56 postos de atendimentos cooperativos, abrange 1123 municípios, 1 bilhão em operações de crédito, 141 milhões em sobras, 2,1 bilhões em depósitos totais, 41281 associados, 704 milhões em patrimônio líquido e 2,9 bilhões em ativos (UNIPRIME, 2017).

A necessidade por tomadas de decisões estão sempre muito presentes em cooperativas de crédito e, por este motivo, ferramentas e métodos quantitativos têm sido cada vez mais necessários para auxiliar os gestores, assim como acontece em instituições bancárias. Com o avanço da tecnologia da informação e seu processamento, decisores têm buscado analistas e especialistas para conseguir extrair o máximo de conhecimento em meio à competitividade e incertezas.

O objetivo do presente trabalho é verificar os principais constructos para as cooperativas de crédito paranaenses analisando os pontos fortes e fracos das pertencentes a um mesmo grupo (*cluster*). Para tanto, fez-se uso do método de Análise de Componentes Principais (Principal Component Analysis, PCA ou *Principal Component Analysis*) e do Agrupamento Hierárquico (*Hierarchical Clustering*). Também foram utilizados os testes de Kruskal-Wallis e Man-Whitney para discussões complementares. Tais análises foram obtidas por meio do software Statistical Package for Social Sciences (SPSS).

O presente artigo está organizado da seguinte forma: além desta seção introdutória, têm-se na seção 2, alguns trabalhos correlatos ao tema aqui tratado; na seção 3 é descrito brevemente o problema das cooperativas paranaenses. A seção 4 traz a metodologia aqui utilizada; a seção 5, a obtenção e discussão dos resultados e, finalmente, na seção 6 são apresentadas as conclusões e sugestões para futuros trabalhos.

## 2. Trabalhos Correlatos

A ideia de agrupar (*clustering*) é antiga e bem conhecida na literatura, contendo variadas aplicações e variações do método para comparações. A seguir são citados alguns trabalhos encontrados na base *Science Direct*, utilizando-se termos sobre “*clustering*” e “*financial institutions*” e similares.

Costea (2014) aplicou a lógica *fuzzy*, com o algoritmo *Fuzzy C-means* e obteve grupos para comparar o desempenho em instituições financeiras não bancárias na Romênia. O conjunto de indicadores de desempenho consistem em definição da adequação de capital, qualidade dos ativos e rentabilidade. Os padrões (indicadores de desempenho) foram enquadrados em três classes de desempenho: ruim, médio ou bom. Posteriormente, os autores aplicaram redes neurais artificiais com mudanças de parâmetros para encontrar a melhor arquitetura. Finalmente, foi utilizado algoritmo genético para treinar a rede com a arquitetura já definida. Os dados utilizados foram de períodos quadrimestrais entre 2007 e 2012 com 990 observações, descartando-se 5 por falta de informações ou incertezas. A taxa de acerto do conjunto de treinamento foi de 92,32%, enquanto que para o de validação foi de 21,34%. O conjunto total de treinamento obteve 92,11% de acuracidade e o conjunto de teste obteve 89,55%.

He e Chen (2016) extraíram conhecimento de um banco de dados de empréstimos de plataformas chinesas de financiamento local, investigando a topologia e taxonomia das relações de crédito com dados empíricos de 2013, contendo 301 plataformas de financiamento com 221 bancos e 29 regiões abrangidas. Os métodos utilizados foram a árvore de expansão mínima e árvores hierárquicas de modo a construir uma rede complexa de similaridades entre instituições acerca da similaridade quanto ao crédito que providencia meios de reconhecer instituições importantes e suas regiões de modo sistemático, detectando as instituições que são “muito grande para falir” e “muito central para falir”.

Scitovski e Sarlija (2014) aplicaram o agrupamento adaptado de Mahalanobis para melhorar a avaliação de crédito para clientes do varejo em um conjunto de dados reais de clientes de um banco da Croácia (3165 clientes). Os fatores utilizadas foram separadas em “demografia”,

“comportamento”, “pessoal”, “financeiro” e “empréstimo”. Os autores aplicaram o algoritmo DIRECT para buscar o número ótimo de grupos e posteriormente o agrupamento em si. Os resultados mostraram que foram utilizados três grupos e houve diferenças significativas entre eles, destacando a necessidade de diferentes estratégias de risco de crédito para cada um dos grupos.

Tsai e Chen (2010) aplicaram quatro tipos diferentes de modelos de aprendizado de máquina para comparação: “Classificação + Classificação”, “Classificação + Agrupamento”, “Agrupamento + Classificação” e “Agrupamento + Agrupamento”. O estudo teve a intenção de utilizar primeiramente um classificador ou agrupador para reduzir dados ou filtrar aqueles relevantes para utilizar como entrada para um outro classificador ou agrupador, melhorando seus resultados. O objetivo foi desenvolver técnicas para concessão de crédito e os dados utilizados pertenciam a um banco em Taiwan. Os resultados mostraram que o modelo híbrido “Classificação + Classificação” baseado na regressão logística e redes neurais mostraram uma acuracidade melhor para classificação e maximização de lucro.

Tsai (2014) desenvolveu um novo modelo híbrido em um problema financeiro (prever falência) que combina técnica de agrupamento e classificadores. Foram comparados métodos clássicos de classificação, híbridos e “ensemble”. Em particular, duas técnicas de agrupamento foram utilizadas: *Self-Organizing Maps* e *k-means*; e três técnicas de classificação: regressão logística, redes neurais do tipo *multilayer-perceptron* e árvores de decisão. Como resultado, 21 modelos diferentes são comparados no que diz respeito à previsão média e erros de Tipo I e Tipo II. Erro Tipo I mostra a taxa de erros de previsão de um modelo que classifica incorretamente um caso de crédito bom como um crédito ruim. Tipo II é exatamente o contrário, sendo assim mais crítico que o primeiro, pois contém maiores riscos para a instituição financeira. Utilizando bases de dados australianas (690 casos), alemã (1000 casos), japonesa (653 casos), *bankruptcy dataset* (241 casos) e *UC competition* (106.777 casos), o modelo híbrido *Self-Organizing Maps* com redes neurais *multilayer perceptron ensemble* foi o que apresentou o melhor resultado, ou seja, a melhor acurácia de previsão e erros menores de Tipo I e Tipo II.

### 3. Descrição do problema das cooperativas

Para o desenvolvimento do trabalho aqui apresentado foi utilizada uma base de dados fornecida pelo Sistema OCEPAR que contém as características das cooperativas de crédito do Paraná. Esta base é composta por um total de 494 indicadores divididos em 57 categorias, coletados no decorrer de 2009 a 2015, totalizando 28 períodos (trimestres) distintos.

Para os estudos serão utilizadas 45 unidades individuais de cooperativas, com a seguinte composição:

- 16 unidades pertencentes ao Sistema Sicoob-Unicoob;
- 24 unidades pertencentes ao Sistema Sicredi, e;
- 5 unidades pertencentes ao sistema Uniprime.

Após entrevista com especialista, o total de 494 indicadores foi reduzido para 20 indicadores (atributos) distribuídos em oito categorias.

| Categoria                            | Indicadores                     | Unidade    |
|--------------------------------------|---------------------------------|------------|
| Tesouraria                           | Grau de alavancagem financeira. | Não possui |
| Avaliação dos resultados<br>(mensal) | Margem financeira bruta;        | R\$        |
|                                      | Despesas administrativas;       | %          |

|                                           |                                                |            |
|-------------------------------------------|------------------------------------------------|------------|
| Avaliação da eficiência social e política | Despesa pessoal;                               | %          |
|                                           | Margem operacional;                            | %          |
|                                           | Margem líquida.                                | %          |
|                                           | Capital social por associado;                  | R\$        |
| Classificação da carteira de crédito      | Quantidade de associados;                      | Pessoas    |
|                                           | Quantidade de funcionários.                    | Pessoas    |
|                                           | Operações de risco nível H.                    | Não possui |
| Contas auxiliares                         | Patrimônio líquido;                            | R\$        |
|                                           | Resultado antes destinações e depois ajustes;  | R\$        |
|                                           | Provisão para créditos de liquidação duvidosa; | R\$        |
|                                           | Sobras antes das destinações;                  | R\$        |
| Rentabilidade                             | Rentabilidade da carteira de crédito;          | %          |
|                                           | Rentabilidade dos ativos.                      | %          |
| Segregação de ingressos/receitas          | Rendas e operações de crédito.                 | R\$        |
| Índices de desempenho                     | Custo médio de captações;                      | %          |
|                                           | Despesas pessoal x Receita total;              | %          |
|                                           | Spread.                                        | %          |

Fonte: Adaptado de OCEPAR (2016)

Tabela 1 - Categorias, indicadores selecionados e suas respectivas unidades

Vale ressaltar que, no presente estudo, uma cooperativa singular é identificada por números de três dígitos (por exemplo, singular 489) e os dados utilizados pertencem ao último trimestre de 2014.

#### 4. Metodologia

A metodologia para o análise das cooperativas paranaenses constou, basicamente, da Análise de Componentes Principais (PCA) e do Agrupamento Hierárquico.

O PCA é uma técnica de redução de dimensionalidade multivariada que cria uma nova plataforma de componentes ótimos que correspondem a mais importante parte da informação dos dados (BITAR; MADIÈS; TAMARASCO, 2017). A técnica identifica as principais variáveis que contribuem em maior parte com a variância total dos dados com a menor perda de informação possível. O PCA é uma técnica que pode obter padrões e similaridades entre os dados de forma mais rápida se comparado à clássica matriz de correlação, a partir da rotação dos eixos ao longo da dispersão medida e alinhando dois ou três eixos de uma elipsóide que representa a dispersão do fenômeno estudado (POMPELLA; DICANIO, 2016).

Para a aplicação do PCA é necessário verificar dois aspectos importantes sobre os dados: teste de esfericidade de Bartlett e a medida de adequação amostral de Kayser-Meyer-Olkin (KMO). O primeiro busca um valor abaixo do nível de significância (0,05 para este estudo) (FIELD, 2009) e o segundo um valor acima de 0,7 para que o PCA seja considerado viável (HUTCHESON; SOFRONIOU, 1999).

Já a análise de agrupamento ou agrupamento é um método de classificação não-supervisionado que busca agrupar elementos com características similares, fornecendo informações relevantes sobre a amostra e pode ser aplicado em diferentes tipos de problemas (PARADI; ZHU; EDELSTEIN, 2012). O agrupamento hierárquico se difere quanto ao não-hierárquico pois pode ter o seu resultado representado em formato de árvore (conhecido como dendrograma) que mostra a sequência de agrupamento ao longo da árvore, apontando vários resultados para diferentes números de grupos. Um dos métodos mais comuns de junção dos elementos durante o agrupamento, e também utilizado no presente estudo, é chamado de Ward linkage (ligação de Ward) que, a cada passo do agrupamento, encontra o par de grupos

cujas junções aumentam o mínimo possível a soma total dos quadrados das distâncias internas do grupo (WARD, 1963; JAYO, 2012). Todos os resultados foram obtidos no *software* SPSS.

## 5. Resultados e Discussão

Para a limpeza inicial dos dados, foram excluídos os casos (singulares) que tinham dois ou mais valores zerados (*missing*) em seus respectivos atributos, que foram as singulares 448, 210, 453 e 233 restando, portanto, 41 singulares após este processo. As singulares em que havia apenas um valor *missing*, os mesmos foram substituídos pela média. Posteriormente, foram excluídas as singulares em que os indicadores apareciam com maior frequência como *outliers* (via *boxplot*). Como o número de singulares é considerado baixo, então foram excluídos apenas aqueles que tiveram valores 3xIQR (*Interquartile range*), neste caso, as 450 e 257.

A Tabela 2 corresponde aos dados obtidos pela estatística descritiva, incluindo as variáveis, número de casos válidos, média e desvio-padrão.

| Variáveis                                           | N  | Média        | Desvio-padrão  |
|-----------------------------------------------------|----|--------------|----------------|
| Grau de alavancagem financeira                      | 39 | 15,5382      | 10,06449       |
| Margem financeira bruta (R\$)                       | 39 | 3,29834389e7 | 2,7516644848e7 |
| Despesas administrativas (%)                        | 39 | ,069131      | ,0240833       |
| Despesas pessoal (%)                                | 39 | ,063910      | ,0147351       |
| Margem operacional (%)                              | 39 | ,065572      | ,0330108       |
| Margem líquida (%)                                  | 39 | ,057654      | ,0336770       |
| Capital social por associado (R\$)                  | 39 | 1738,1921    | 1625,15633     |
| Quantidade de associados                            | 39 | 16965,59     | 15315,077      |
| Quantidade de funcionários                          | 37 | 133,27       | 103,148        |
| Operações de risco nível H                          | 39 | ,021795      | ,0119463       |
| Patrimônio líquido (R\$)                            | 39 | 3,58031482e7 | 3,1075861256e7 |
| Resultado antes destinações e depois ajustes (R\$)  | 39 | 2,6742911e6  | 2,6797520495e6 |
| Provisão para créditos de liquidação duvidosa (R\$) | 39 | 6,27615420e6 | 5,2213165724e6 |
| Sobras antes das destinações (R\$)                  | 39 | 5,89844397e6 | 5,6666879124e6 |
| Rentabilidade da carteira de crédito (%)            | 39 | ,045838      | ,0254646       |
| Rentabilidade dos ativos (%)                        | 39 | ,011177      | ,0063624       |
| Rendas de operações de crédito (R\$)                | 39 | 4,07745215e7 | 3,3163717683e7 |
| Custo médio de captações (%)                        | 39 | ,046215      | ,0072200       |
| Despesas pessoal vs receita total (%)               | 39 | ,149836      | ,0318003       |
| Spread (%)                                          | 39 | ,150267      | ,0479599       |
| N válido                                            | 37 |              |                |

Fonte: Autores.

Tabela 2 – Estatística descritiva dos dados.

De acordo com a Tabela 2, observa-se as médias que substituíram os valores faltantes nos dados *missing* de acordo com o critério citado anteriormente. Dois foram os casos necessários para substituição destes valores, ambos do indicador “Quantidade de funcionários”.

Como resultado da limpeza dos dados, obteve-se 39 singulares a serem analisadas com 20 indicadores. Neste contexto, aplicou-se o PCA para a extração dos principais fatores e obtenção dos novos valores para cada singular.

Primeiramente, foi aplicado o teste de esfericidade de Bartlett e a medida de adequação Kaiser-Meyer-Olkin, obtendo-se os resultados mostrados na Tabela 3.

Medida de adequação de amostra de Kaiser-Meyer-Olkin.

,737

|                                   |                         |          |
|-----------------------------------|-------------------------|----------|
| Teste de Esfericidade de Bartlett | Chi-quadrado aproximado | 1349,151 |
|                                   | Graus de liberdade      | 190      |
|                                   | Valor <i>p</i>          | ,000     |

Fonte: Autores.

Tabela 3 – Teste de esfericidade de Bartlett e medida de adequação KMO

O valor para medida de adequação foi acima de 0,6 (0,737), enquanto que o teste de esfericidade encontrou um valor significativo (abaixo de 0,05). Desta forma, pode-se prosseguir com a redução da dimensionalidade.

Para o caso estudado, após aplicação do método de extração PCA no SPSS sobre indicadores, foram encontrados 5 componentes (fatores). A Tabela 4 mostra os autovalores e a porcentagem da variância para cada componente.

| Componentes | Autovetores Iniciais |                |             |
|-------------|----------------------|----------------|-------------|
|             | Total                | % de Variância | % Acumulada |
| 1           | 7,835                | 39,175         | 39,175      |
| 2           | 4,558                | 22,788         | 61,963      |
| 3           | 2,676                | 13,378         | 75,342      |
| 4           | 2,134                | 10,672         | 86,014      |
| 5           | 1,034                | 5,172          | 91,186      |

Fonte: Autores

Tabela 4 – Variância total explicada

Nesta Tabela 4, na coluna “Total” tem-se os autovalores que variam de acordo com a componente e, pelo critério de escolha de componentes com valores acima de “1”, então são 5 fatores que, de acordo com a coluna “% Acumulada”, juntas explicam aproximadamente 92% dos dados. Também pode ser verificado no *Scree Plot* da Figura 1 que são cinco componentes que explicam a maior variância dos dados.

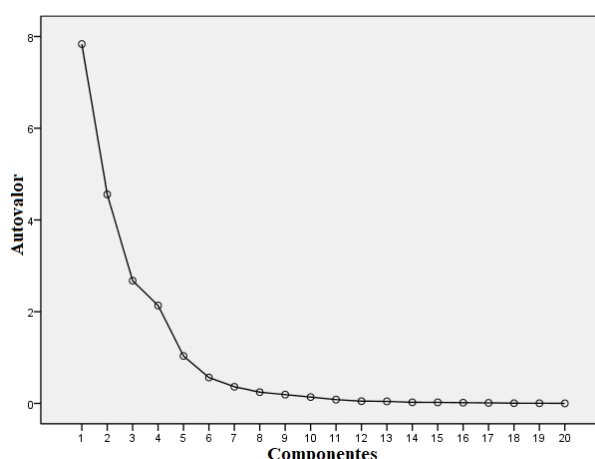


Figura 1 – Autovalores dos componentes

Fonte: Autores

Com o método de rotação *varimax* com normalização de Kaiser, obteve-se a alocação das 20 variáveis (indicadores) para as cinco componentes como mostrado na Tabela 5.

| Atributos                            | Componente |   |   |   |   |
|--------------------------------------|------------|---|---|---|---|
|                                      | 1          | 2 | 3 | 4 | 5 |
| RENDAS DE OPERAÇÕES DE CRÉDITO (R\$) | ,987       |   |   |   |   |

|                                                     |      |       |
|-----------------------------------------------------|------|-------|
| PATRIMÔNIO LÍQUIDO (R\$)                            | ,974 |       |
| MARGEM FINANCEIRA BRUTA                             | ,970 |       |
| QUANTIDADE DE FUNCIONÁRIOS                          | ,960 |       |
| QUANTIDADE DE ASSOCIADOS                            | ,951 |       |
| PROVISÃO PARA CRÉDITOS DE LIQUIDAÇÃO DUVIDOSA (R\$) | ,945 |       |
| SOBRAS ANTES DAS DESTINAÇÕES (R\$)                  | ,934 |       |
| RESULTADO ANTES DESTINAÇÕES E DEPOIS AJUSTES        | ,883 |       |
| RENTABILIDADE DOS ATIVOS (%)                        | ,953 |       |
| MARGEM OPERACIONAL (%)                              | ,922 |       |
| MARGEM LÍQUIDA (%)                                  | ,916 |       |
| RENTABILIDADE DA CARTEIRA DE CRÉDITO (%)            | ,829 |       |
| SPREAD (%)                                          |      | ,898  |
| OPERAÇÕES DE RISCO NÍVEL H                          |      | ,777  |
| CUSTO MÉDIO DE CAPTAÇÕES (%)                        |      | -,761 |
| CAPITAL SOCIAL POR ASSOCIADO (R\$)                  |      | -,689 |
| DESPESAS PESSOAL (%)                                |      | ,973  |
| DESPESAS PESSOAL X RECEITA TOTAL (%)                |      | ,949  |
| DESPESAS ADMINISTRATIVAS (%)                        |      | ,719  |
| GRAU DE ALAVANCAGEM FINANCEIRA                      |      | ,949  |

Fonte: Autores

Tabela 5 – Matriz Rotacionada de Componentes

Esta Tabela 5 apresenta os cinco componentes que podem ser definidas por diferentes constructos e que têm baixa interdependência entre si. Os constructos gerados podem adquirir as seguintes denominações:

- Constructo 1: Porte;
- Constructo 2: Desempenho;
- Constructo 3: Qualidade da Carteira;
- Constructo 4: Custos Administrativos;
- Constructo 5: Alavancagem.

Após os resultados gerados, também obteve-se por meio da regressão linear os novos valores de cada singular da amostra para constructo. Estes novos valores foram utilizados para a aplicação do método de agrupamento hierárquico (*Hierarchical Cluster*).

O primeiro passo foi aplicar o método de agrupamento apenas para a identificação do número de grupos. Como não se tinha inicialmente nenhuma ideia do número de grupos seria interessante analisar, executou-se o algoritmo de classificação no SPSS para diferentes cenários, variando o número de grupos de dois a nove, para então observar suas características de frequência para cada execução.

Para a escolha do número de grupos, adotou-se quatro como referência, pois cada uma dos quatro grupos apresentou um número integrante de singulares similares. Entretanto, o grupo 4 (apenas quatro singulares como pode ser observado na Tabela 6 mais adiante) é o que menos tem singulares e, independentemente do número de grupos adotado para inicialização, ele não fez ligação com nenhum outro ao longo do método de ligação de Ward. Isto indica que as características deste grupo são significativamente distintas de todos os outros. A Tabela 6 mostra a frequência e divisão dos agrupamentos considerando-se quatro grupos.

| Grupo | Sicoob-Unicoob | Uniprime | Sicredi | Total |
|-------|----------------|----------|---------|-------|
| 1     | 5              | 0        | 6       | 11    |

|       |    |   |    |    |
|-------|----|---|----|----|
| 2     | 9  | 0 | 2  | 11 |
| 3     | 1  | 0 | 12 | 13 |
| 4     | 0  | 3 | 1  | 4  |
| Total | 15 | 3 | 21 | 39 |

Fonte: Autores

Tabela 6 – Frequência e divisão para agrupamento com 4 grupos

A Tabela 6 mostra que os grupos em geral possui homogeneidade interna no que diz respeito aos sistemas de cooperativas à que as singulares pertencem, exceto pelo primeiro grupo (45% das singulares são do sistema Sicoob-Unicoob e 55% pertencem ao sistema Sicredi). O grupo 2 possui aproximadamente 80% de suas singulares pertencentes ao sistema Sicoob-Unicoob, o grupo 3 possui próximo de 92% das suas singulares correspondentes ao sistema Sicredi e o grupo 4 tem 75% de suas singulares pertencentes ao sistema Uniprime.

À medida em que o algoritmo avançava e as ligações pelo método de Ward eram feitas, as singulares 465, 419, 449 e 464 sempre permaneceram juntas e em nenhum momento se juntaram a outros grupos. Observa-se então que estas cooperativas têm uma ou mais características bem distintas das de outros grupos. A Tabela 7 mostra o resultado do agrupamento.

| Grupo |     |     |     |
|-------|-----|-----|-----|
| 1     | 2   | 3   | 4   |
| 479   | 289 | 358 | 465 |
| 462   | 346 | 304 | 419 |
| 517   | 252 | 318 | 449 |
| 507   | 357 | 315 | 464 |
| 485   | 493 | 431 |     |
| 492   | 484 | 290 |     |
| 490   | 546 | 237 |     |
| 416   | 452 | 319 |     |
| 478   | 406 | 250 |     |
| 473   | 489 | 260 |     |
| 469   | 299 | 301 |     |
|       |     | 251 |     |
|       |     | 498 |     |

Fonte: Autores

Tabela 7 – Singulares pertencentes aos diferentes grupos

Após a definição do número de grupos, aplicou-se o teste de Kruskal-Wallis (não-paramétrico, pois as amostras são baixas para cada grupo), para analisar e identificar as diferentes características entre os grupos. A comparação entre as médias de cada grupo se torna interessante para uma avaliação mais detalhada, indicando características significativas ou não. A Tabela 8 mostra o teste para cada constructo, comparando os grupos.

|                    | Porte  | Desempenho | Qualidade da carteira | Custos administrativos | Alavancagem |
|--------------------|--------|------------|-----------------------|------------------------|-------------|
| Chi-quadrado       | 12,712 | 7,551      | 20,000                | 18,426                 | 16,792      |
| Graus de liberdade | 3      | 3          | 3                     | 3                      | 3           |
| Valor <i>p</i>     | ,005   | ,056       | ,000                  | ,000                   | ,001        |

Fonte: Autores

Tabela 8 – Estatísticas do teste de Kruskal-Wallis

De acordo com a Tabela 8, o teste mostrou que existe diferença significativa para todos os constructos, sendo que apenas o constructo “Desempenho” obteve seu *p-value* acima de 0,05, mostrando que os grupos não diferenciam significativamente para este constructo. A Tabela 9 identifica se há diferença significativa entre os grupos para cada constructo.

| Grupo        | N  | Porte | Desempenho | Ranks                 |                        |             |
|--------------|----|-------|------------|-----------------------|------------------------|-------------|
|              |    |       |            | Qualidade da Carteira | Custos Administrativos | Alavancagem |
| 1            | 11 | 15,55 | 25,45      | 26,45                 | 17,55                  | 8,73        |
| 2            | 11 | 16,45 | 14,36      | 12,91                 | 31,73                  | 21,91       |
| 3            | 13 | 29,08 | 17,77      | 25,77                 | 15,92                  | 27,54       |
| 4            | 4  | 12,50 | 27,75      | 3,00                  | 7,75                   | 21,25       |
| <b>Total</b> | 39 |       |            |                       |                        |             |

Fonte: Autores

Tabela 9 – Ranks para o teste de Kruskal-Wallis

A Tabela 9 indica, também, que o grupo 3 possui uma média maior no que se refere aos indicadores aos constructos de “Porte” e “Alavancagem”. Quanto à “Alavancagem”, o grupo 1 apresenta característica inferior quando comparado aos outros grupos. O grupo 4 tem média significativamente menor tanto para “Qualidade da Carteira” quanto para “Custos Administrativos”.

Aplicando o teste U de Mann-Whitney para o constructo “Custos administrativos”, conseguiu-se perceber uma diferença significativa do grupo 2 para os outros grupos, enquanto que o grupo 4 possui diferença significativa para os grupos 1 e 2, porém não para 3.

Fazendo um cruzamento de informações sobre as diferenças das Tabelas 6 e 9, observa-se que o grupo 1 é o mais heterogêneo quanto aos sistemas participantes e tem uma alavancagem inferior aos demais grupos. O grupo 2 é predominantemente composto por singulares do sistema Sicoob-Unicoob e estes tem custos administrativos significativamente mais elevados que os outros grupos, principalmente se comparado ao grupo 4. O grupo 3 tem, do total de 13 singulares, 12 pertencentes ao sistema Sicredi e possui o constructo “Porte” mais elevado comparado aos outros. Já o grupo 4 possui uma “Qualidade de Carteira” inferior comparado aos outros grupos, principalmente a 1 e 3. Pode-se afirmar estatisticamente que todos os grupos possuem características similares quanto ao parâmetro “Desempenho”.

## 6. Conclusões

A análise de agrupamentos é bastante conhecida e aplicada em diversas áreas, como saúde, finanças, engenharia, gestão, meio ambiente, entre outros. A classificação não-supervisionada ganha destaque por não necessitar de nenhum *output* inicial e assim pode ser aplicada em variadas realidades. No presente estudo, analisou-se as similaridades de diferentes grupos de cooperativas paranaenses por meio da redução de dimensionalidade com a técnica do PCA e agrupamento hierárquico. Testes estatísticos de comparação de grupos foram realizados para a diferenciação das características de cada constructo. Os resultados se mostraram satisfatórios ao diferenciar características em que diferentes singulares se apresentavam no último trimestre do ano de 2014, conseguindo observar diferenças em quatro constructos, porém excluindo “desempenho”, que não teve diferença estatisticamente significativa. O Quadro 1 mostra de forma resumida os pontos fortes e fracos dos grupos analisados.

| Grupo | Ponto(s) forte(s)     | Ponto(s) fraco(s)            |
|-------|-----------------------|------------------------------|
| 1     | Qualidade da carteira | Alavancagem                  |
| 2     | Alavancagem           | Custos administrativos       |
| 3     | Porte e Alavancagem   | Não tem nenhum significativo |
| 4     | Desempenho            | Qualidade da carteira        |

Fonte: Autores

Quadro 1 – Pontos fortes e fracos das cooperativas

As técnicas apresentaram resultados significativos e encontraram diferenças entre grupos para diferentes constructos, o que não impede de futuramente outras técnicas serem aplicadas para comparar seus resultados, dentre as quais pode-se citar, por exemplo, *k-means*, *fuzzy c-means*, bem conhecidas na literatura.

### Agradecimentos

Os autores agradecem à OCEPAR pelo fornecimento dos dados para o desenvolvimento do presente trabalho.

### Referências

**BITAR, M.; MADIÈS, P.; TARAMASCO, O.** *What makes Islamic banks different? A multivariate approach.* Economic Systems, v. 41, n. 2, p. 215-235, 2017.

**BITTENCOURT, W. R.; BRESSAN, V. G. F.; GOULART, C. P.; BRESSAN, A. A., COSTA, D. R. de M.; LAMOUNIER, W. M.** *Rentabilidade em Bancos Múltiplos e Cooperativas de Crédito Brasileiros.* Revista de Administração Contemporânea, v. 21, p. 22-40, 2017.

**COSTEA, A.** *Applying fuzzy logic and machine learning techniques in financial performance predictions.* Procedia: Economics and Finance, v. 10, p. 4-9, 2014.

**FIELD, A.** *Discovering Statistics Using SPSS.* 3 ed. Londres: Sage, 2009.

**HE, F. CHEN, X.** *Credit networks and systemic risk of Chinese local financing platforms: Too central or too big to fail? –based on different credit correlations using hierarchical methods.* Physica A: Statistical Methods and its Applications, v. 461, p. 158-170, 2016.

**HUTCHESON, G.; SOFRONIOU, N.** *The multivariate social scientist.* Londres: Sage, 1999.

**JACQUES, E. R.; GONÇALVES, F. de O.** *Cooperativas de crédito no Brasil: evolução e impacto sobre a renda dos municípios brasileiros.* Economia e Sociedade, v. 25, n. 2, p. 489-509, 2016.

**JAYO, M.; DINIZ, E. H.; ZAMBALDI, F.; CHRISTOPOULOS, T. P.** *Groups of services delivered by Brazilian branchless banking and respective network integration models.* Electronic Commerce Research, and Applications, v. 11, n. 5, p. 504-517, 2012.

**PARADI, J. C.; ZHU, H.; EDELSTEIN, B.** *Identifying managerial groups in a large Canadian bank branch network with a DEA approach.* European Journal of Operational Research, v. 2019, n. 1, p. 178-187, 2012.

**PINHEIRO, M. A. H.** *Cooperativas de crédito: história da evolução normativa no Brasil.* 6 ed., Brasília: Banco Central do Brasil, 2008.

**POMPELLA, M.; DICANIO, A.** *Rating based Inference and Credit Risk: Detecting likely-to-fail Banks with the PC-Mahalanobis Method.* Economic Modelling. In Press, 2016.

**RELATÓRIO DE INCLUSÃO FINANCEIRA.** Banco Central do Brasil. Brasília, 2015.

**SCITOVSKI, S.; SARLIJA, N.** *Cluster analysis in retail segmentation for credit scoring.* Croatian Operational Research Review, v. 5, p. 235-245, 2014.

**SICOOB-UNICOOB.** Disponível em: < <http://www.sicoobunicoob.com.br/sicoob-unicoob/>>. Acesso em: jun 2017.

**SICREDI: SOBRE O SICREDI.** Disponível em: < <https://www.sicredi.com.br/html/conheca-o-sicredi/quem-somos/>>. Acesso em: jun 2017.

**SISTEMA OCEPAR: MISSÃO E VISÃO.** Disponível em: <  
<http://www.paranacooperativo.coop.br/ppc/index.php/2011-12-05-11-28-07/2011-12-05-11-31-38>>. Acesso em:  
mai 2017.

**TSAL, C-F.** *Combining cluster analysis with classifier ensembles to predict financial distress.* Information Fusion, v. 16, p. 46-58, 2014.

**TSAL, C-F; CHEN, M-L.** *Credit rating by hybrid machine learning techniques.* Applied Soft Computing, v. 10, p. 374-380, 2010.

**UNIPRIME CENTRAL: CONHEÇA A COOPERATIVA.** Disponível em: <  
<http://www.uniprimecentral.com.br/cooperativa/quemsomos>>. Acesso em: jun 2017.

**WARD, J.** *Hierarchical grouping to optimize an objective function.* Journal of the American Statistical Association, v. 58, n. 301, p. 236-244, 1963.