

QUEM É O VENCEDOR ENTRE O FUTEBOL E O DESEMPENHO FINANCEIRO?

Anderson Ribeiro Duarte¹, Josino José Barbosa¹, Helgem de Souza Ribeiro Martins¹
 Gabriel Vieira de Amorim¹

RESUMO

A cada dia, cifras financeiras mais vultosas são envolvidas nos esportes de alto rendimento. Particularmente, no futebol, os valores são realmente surpreendentes. Diante desse contexto, uma pergunta impactante vem à tona, “quem de fato vence as competições esportivas, o desempenho puramente esportivo, ou é o desempenho financeiro?” Termos como doping financeiro, marketing da bola na rede, entre outros, surgem pelo mundo. Aliado com isso, todos tentam observar o esporte como competição apenas entre atletas, mas cada vez mais os departamentos financeiros das equipes tem papel decisivo nos campeonatos. As ferramentas estatísticas podem ser de grande valia para tentar elucidar este dilema. Esta investigação busca atender este objetivo através de técnicas associadas com valores outliers. Os valores outliers são elementos usualmente incomuns ao conjunto de dados, valores excessivamente extremos quanto a ordem de grandeza das variáveis em estudo. Os procedimentos de detecção de valores outliers já são bastante difundidos, mas os estudos para valores outliers multivariados ainda não são plenamente estabelecidos. Técnicas inovadoras para este propósito são apresentadas na literatura, em particular, a metodologia Data-driven Cluster Analysis Method (DDCAM). Este estudo utiliza a metodologia para investigar a forte associação entre desempenho financeiro e resultado desportivo em clubes do futebol brasileiro de alta performance.

Palavras-chave: Futebol. Arrecadação financeira. Desempenho esportivo. Outliers multivariados.

ABSTRACT

Who is the winner between football and financial performance?

Nowadays, larger financial values are involved in high-performance sports. Particularly in football, the values are astonishing. Given this context, an impactful question arises, who wins the sports competitions? Is the performance purely sports? Or is it financial performance? Terms like financial doping, and marketing of goal, among others, appear around the world. Allied with this, everyone tries to see the sport as competition only between athletes, but increasingly the financial departments of teams play a decisive role in championships. Statistical tools can be of great help in trying to elucidate this quandary. This investigation seeks to meet this objective through techniques associated with outliers values. The outlier values are usually distinctive elements in the datasets, values that are excessively extreme in terms of the order of magnitude of the variables under study. The procedures for detecting outliers values are already quite widespread, but studies for multivariate outliers values are not yet fully established. Innovative techniques for this purpose are presented in the literature, in particular, the Data-driven Cluster Analysis Method (DDCAM). This present study uses this methodology to investigate the strong association between financial performance and sports results in high-performance Brazilian football.

Key words: Football. Financial performance. Sports performance. Multivariate outliers.

1 - Universidade Federal de Ouro Preto, Campus Morro do Cruzeiro, Departamento de Estatística, Ouro Preto-MG, Brasil.

E-mail dos autores:
 anderson.duarte@ufop.edu.br
 josino@ufop.edu.br
 helgem@ufop.edu.br
 gabriel.amorim@aluno.ufop.edu.br

INTRODUÇÃO

A prática de esportes de alto rendimento ocupa um papel de destaque na nossa cultura atual.

Particularmente, o Futebol tem enorme destaque neste cenário. O volume de praticantes profissionais, praticantes amadores e aficionados acompanhantes desse esporte somam números grandiosos (Duarte, Martins e Silva, 2018).

Os veículos de mídia dedicam enorme espaço às informações acerca do futebol. Estas considerações entre inúmeras outras servem como motivação para investigação de temas ligados ao Futebol.

Em decorrência de todo este interesse, os valores financeiros envolvidos na prática profissional do Futebol alcançam patamares grandiosos.

É razoável supor, sem informação científica prévia, que as equipes de maior poderio financeiro alcancem os melhores resultados desportivos.

Porém, a literatura científica não apresenta uma quantidade significativa de estudos que confirmem tal tese.

Dada a premissa inicial da especulação sobre uma associação direta entre o desempenho financeiro e o desempenho esportivo das equipes.

Uma avaliação com subsídio estatístico das variáveis de desempenho e planejamento financeiro poderia prever os maiores favoritos ao melhor desempenho esportivo.

Neste estudo, a aplicação de um método de detecção de valores outliers nas variáveis do desempenho financeiro é utilizada para prever outliers de valores máximos que são os mais fortes candidatos para alcançar os melhores desempenhos esportivos no que tange à conquista de campeonatos.

Dentre os diversos problemas cotidianos que envolvem conjuntos de dados, é possível que vez por outra, dados extremamente discrepantes em relação aos demais se revelem.

Usualmente, dados nessa situação são chamados de valores outliers. De acordo com Barbosa, Duarte e Martins (2020), um valor outlier é um valor que escapa do padrão dos demais elementos do conjunto de dados.

A presença de outliers é capaz de afetar análise futura acerca dos dados. Para

conjuntos de dados com existência de valores outliers, eles tendem a ser encontrados mais facilmente para base de dados menores.

Porém, o crescimento dos conjuntos de dados, tanto em número de observações quanto em número de variáveis, tende a ofuscar a detecção rápida e fácil desses valores.

A análise de dados conduz diversas tomadas de decisões muito importantes.

Contudo, a presença de dados espúrios no conjunto de dados analisados pode levar a decisão tomada para um caminho inadequado, ou pelo menos para um caminho que não seja a melhor opção disponível.

Segundo Aggarwal (2017), um procedimento para detecção da presença de valores outliers pode tornar a análise da base de dados mais confiável. Este estudo tem como motivações, a apresentação e utilização de uma técnica específica para detecção de equipes outliers em variáveis do aspecto financeiro entre equipes profissionais de futebol.

Diversos autores apresentaram metodologias e técnicas ligas ao procedimento de detecção de valores outliers em conjuntos de dados.

Diante disso, este estudo apresenta uma revisão com métodos e aplicações mais atualizados acerca deste tema.

Wahid e Rao (2019) realizaram um estudo sobre detecção de valores outliers baseado em distância.

O procedimento é executado através da utilização do clássico algoritmo Particle swarm optimization (PSO).

O algoritmo atribui um grau de distância a cada ponto dos dados usando a soma das distâncias entre os pontos e seu conjunto de vizinhos mais próximos. Com isso, o PSO é utilizado para detectar subespaços nos quais podem existir valores atípicos entre as vizinhanças.

Lejeune, Mothe e Soubki (2020) apresentaram um método de detecção de outliers baseado em dados funcionais multivariados.

O método consiste na identificação de outliers por meio de funções de mapeamento de geometria diferencial. São funções que captam diferentes características periféricas dos dados.

Um estudo experimental em dados reais comprovou a eficácia do método, quando combinado com algoritmos mais modernos de detecção de outliers, e ainda, capacidade de superar métodos baseados em profundidade funcional.

Barbosa, Martins e Oliveira (2018) apresentaram uma alternativa para a detecção de outliers baseada em análise de agrupamentos, em casos multivariados.

São apresentadas algumas desvantagens de se usar métodos baseados na distância de Mahalanobis.

A comparação foi feita por meio de um procedimento de simulação.

Foram comparadas técnicas baseadas em distância de Mahalanobis através de estimação robusta, com os estimadores minimum covariance determinant (MCD), o minimum volume ellipsoid (MVE), além de uma técnica baseada em análise de agrupamentos.

Barbosa, Duarte e Martins (2020) realizaram outro estudo comparativo entre metodologias de detecção de outliers multivariados baseadas em distância de Mahalanobis através de estimação robusta, com os estimadores minimum covariance determinant (MCD), o minimum volume ellipsoid (MVE), e a técnica baseada em análise de agrupamentos.

O objetivo era verificar possíveis deficiências e capacidade de evolução para a técnica de análise de agrupamentos.

As métricas definidas para medir a qualidade dos métodos foram sensibilidade, especificidade e a precisão de cada método, além do tempo computacional para realização dos procedimentos.

Como conclusão, a análise de agrupamentos se mostrou superior tanto nas medidas de qualidade, quanto no tempo de execução do processo, desde que o número de agrupamentos ótimo possa ser conhecido.

A literatura não apresenta de forma enfática, trabalhos que associem a busca por valores outliers em dados associados à prática do futebol.

Como mencionado na introdução, este estudo busca apresentar uma associação entre resultados esportivos e financeiros do futebol por meio de aplicação de um procedimento de detecção de outliers.

Alguns estudos correlatos observados serão mencionados, mas nenhum apresenta

alguma aplicação com o exato propósito deste estudo.

O estudo de Breunig, Kriegel, Ng e Sander (2000) consideram valores extremos como propriedades binárias e afirma que é mais significativo atribuir a cada objeto um grau sobre a possibilidade de ser um outlier, o fator de outlier local (LOF) de um objeto.

O método é aplicado para informações técnicas de atletas na Bundesliga e conclui de fato que alguns atletas tidos como outliers pelas variáveis são de fato atletas históricos da liga como o brasileiro Giovanni Elber.

Dantas, Silvia e Boente (2011) realizam uma análise de outliers nas demonstrações contábeis do Sport Club Corinthians Paulista, nos anos de 2008 a 2010.

O estudo foi realizado com a utilização do teste de Grubbs, com intuito de analisar o impacto dos ativos intangíveis sobre as contas do clube, por meio da detecção dos outliers nos anos de estudo.

Para cada um dos anos de estudos, diferentes variáveis foram classificadas como outliers e motivos como a chegada do atacante Ronaldo e o rebaixamento do clube em 2008 ajudou a explicar o motivo destas conclusões.

Nascimento, Dantas, Azevedo (2019) realizaram um estudo sobre quais variáveis financeiras e esportivas mais influenciam no valor dos clubes de futebol brasileiros mais valiosos. Para a análise, foram utilizados dados da BDO RCS (2015) (Daniel, 2015), que classifica os clubes mais valiosos do Brasil, e as demonstrações contábeis dos clubes entre os anos de 2011 e 2014.

Como conclusão, as variáveis ranking CBF, grau de endividamento e custo/receita se apresentaram mais significativas no modelo. Novamente aqui não existe uma aplicação direta de detecção de outliers, mas novamente os dados multivariados são discutidos de forma correlacionados com o desempenho financeiro.

Ferreira, Marques e Macedo (2018) realizaram outro estudo para verificar a relação entre o desempenho esportivo e financeiro dos clubes brasileiros de futebol.

Foram investigados todos os clubes que disputaram a Série A ou a Série B do Campeonato Brasileiro pelo menos uma vez entre os anos de 2013 e 2016.

Para medir o desempenho esportivo, foi levado em conta a posição do clube no Ranking Oficial de Clubes da CBF e a

pontuação em um ranking elaborado pelos próprios autores. Já na vertente financeira, foram consideradas algumas variáveis retiradas dos balanços financeiros dos clubes.

Os principais resultados foram as relações positivas e significativas das variáveis receita bruta, despesa com salários e endividamento com o desempenho esportivo.

De uma maneira mais abrangente, o objetivo deste estudo é apresentar uma metodologia para processo de detecção de outliers e propor uma aplicação efetiva do método para buscar explicar a associação entre o poderio financeiro das equipes de futebol e os resultados desportivos alcançados por estas equipes.

A técnica Data-driven Cluster Analysis Method (DDCAM) proposta por Duarte e colaboradores. (2022) será apresentada neste estudo e aplicada em dados reais do Futebol brasileiro.

De acordo com a especulação da existência de uma associação direta entre o desempenho financeiro e o desempenho esportivo das equipes, a avaliação das variáveis de desempenho e planejamento financeiro podem predizer os maiores favoritos ao melhor desempenho esportivo.

A aplicação de detecção de outliers nas variáveis do desempenho financeiro tende a prever outliers de máximo que são os mais fortes candidatos para alcançar os melhores desempenhos esportivos no que tange à conquista de campeonatos.

MATERIAIS E MÉTODOS

Este estudo está direcionado para a aplicação de uma técnica inovadora para o problema de detecção de valores outliers.

Além da discussão correlata da sua aplicabilidade em diversos problemas de interesse prático. O método em questão é o DDCAM de Duarte e colaboradores (2022).

O método DDCAM parte de premissas iniciais associadas à aplicação de técnica predecessora baseada em análise de agrupamentos Cluster Analysis Method (CAM) proposto por Barbosa, Pereira e Oliveira (2018).

E acrescenta um fator adaptativo para metodologia, que se ajusta de acordo com informações inerentes aos dados, daí a nomenclatura Data-driven.

A aplicação do método CAM, usa o procedimento de análise de clusters, para a construção de k grupos por meio do procedimento de análise de agrupamentos k-médias.

O método k-médias parte da seleção aleatória de k centroides, cada um destes fica associado com um dos agrupamentos gerados.

Obviamente, a escolha do valor k tem impacto decisivo no procedimento. A metodologia prévia CAM escolhe este valor de forma ad-hoc, ou seja, de forma completamente arbitrária.

A nova abordagem DDCAM apresenta um mecanismo adaptativo aos dados para a escolha de um valor k mais adequado para aquele conjunto de dados, para então partir para a construção dos agrupamentos através do clássico método de agrupamento de Ward para a geração dos centroides iniciais do método k-médias. Isso assegura que, para o mesmo conjunto de dados, em duas realizações distintas, sempre será obtida a mesma partição.

Após a construção da partição dos dados, a metodologia DDCAM busca definir uma distância entre os centroides dos agrupamentos e a mediana referente ao conjunto de dados tal que seja possível buscar agrupamentos heterodoxos no que tange à distância.

O objetivo é verificar a possível existência de agrupamentos suficientemente distantes da mediana com respeito à norma euclidiana. Um centroide de um determinado agrupamento que esteja muito distante em relação à mediana é potencialmente um centroide que determina um agrupamento de elementos que se comportem como valores outliers.

O desvio padrão (s_c) entre os centroides de agrupamentos, bem como a mediana \bar{X} do conjunto completo de dados são utilizados pelo método CAM para o procedimento de detecção dos valores extremos multivariados. Se a norma euclidiana que determina a distância entre o centroide e a mediana ultrapassar a cota de $2,5s_c$, o agrupamento associado à esse centroide é considerado ser composto por valores outliers.

A metodologia DDCAM é uma melhoria imposta ao método CAM, a estratégia é conduzida em 4 etapas, denominadas:

- estimação do valor δ ;

- processo de refinamento I;
- processo de refinamento II;
- busca pelo valor adequado k.

Essas etapas funcionam como a reparametrização do método CAM. Em seguida, o procedimento CAM é executado, agora com a parametrização estabelecida pelas quatro etapas prévias.

Estimação do Valor δ

Uma discussão não tão recorrente acerca dos procedimentos de detecção de valores outliers reside na determinação de uma quantidade máxima de valores aparentemente admissível para existência dos valores extremos.

Visto por outro prisma, não parece atender a razoabilidade partir da admissão de que seja factível uma proporção muito elevada de elementos definidos como outliers.

Proporções muito elevadas retratariam na prática que são valores não raros, ou seja, contrapondo a definição específica de valores extremos.

A quantidade δ é definida pelo método DDCAM como um limiar admissível para essa proporção de possíveis valores outliers.

Processo de Refinamento I

Dado que o método propõe vasculhar possíveis valores k e escolher um mais adequado aos dados, um valor máximo para o número k de agrupamentos precisa ser estabelecido. O DDCAM avalia diversos valores, até um valor máximo denominado k_{max} determinado pela razão entre o total de dados e o logaritmo dessa quantidade. O número máximo de agrupamentos analisados pelo método é limitado pelo valor atribuído para k_{max} . Diante disso, o método parte da premissa de que o valor adequado de k pertença ao conjunto $\mathcal{K}=\{2,3,\dots, k_{max}\}$.

A investigação é iniciada para o agrupamento com $k=2$ e segue pela verificação se o menor agrupamento gerado (em volume de elementos) tem no máximo $\hat{\delta} \times n$ valores. Na suposição dessa condição ser atendida, o agrupamento com $k=2$ é considerado ser um agrupamento válido. Por outro lado, se o agrupamento com $k=2$ não atende tal condição,

O DDCAM propõe uma estratégia específica para estimar δ através das informações do próprio conjunto de dados.

Essa primeira etapa, tem por foco central estimar δ , ou seja, determinar a quantidade de outliers que parece ser razoável para ser admitida para a base de dados em investigação.

O processo de estimação proposto pelo DDCAM parte de proposições univariadas acerca dos dados em estudo.

Considere a média amostral (\bar{X}_i) e o desvio-padrão amostral (s_i) para o vetor composto por n observações ($X_{1i}, X_{2i}, \dots, X_{ni}$) respectivo à i-ésima variável em estudo.

O estimador $\hat{\xi}_i$ é determinado pela proporção de valores no vetor ($X_{1i}, X_{2i}, \dots, X_{ni}$) que extrapolam a distância de η desvios padrão amostrais s_i da média \bar{X}_i . O valor η é definido pelo usuário de acordo com seu conhecimento acerca das variáveis em estudo. O estimador $\hat{\delta}$ (veja Equação (1)) é o máximo entre as estimativas univariadas $\hat{\xi}_i$, dentre as variáveis unidimensionais em estudo.

$$\hat{\delta} = \max_{1 \leq i \leq p, i \in \mathbb{N}} (\hat{\xi}_i) \quad (1)$$

é dito um agrupamento inválido para a investigação. Esse procedimento é repetido para todos os demais valores k no conjunto \mathcal{K} . Esse procedimento determina um conjunto \mathcal{K} refinado, composto pelos valores k válidos, denominado \mathcal{K}_1 , tal que $\mathcal{K}_1 \subseteq \mathcal{K}$.

Processo de Refinamento II

Um segundo refinamento também é realizado. Dentre os valores $k \in \mathcal{K}_1$, é verificada a existência de agrupamentos cujos centroides estejam suficiente afastados da mediana (pela cota $\phi \times s_c$, com o valor ϕ pré-fixado pelo usuário do método) e que simultaneamente possuam quantidade de elementos limitada por no máximo $\hat{\delta} \times n$ elementos. O não atendimento dessa condição também exclui esse valor k da investigação. Dessa forma, um segundo conjunto de valores k ainda mais refinado é produzido, o conjunto \mathcal{K}_2 , com $\mathcal{K}_2 \subseteq \mathcal{K}_1$.

É importante ressaltar que após os procedimentos de refinamento, todos os possíveis valores de k podem ser excluídos. Na eventualidade de $\mathcal{K}_1 = \emptyset$, o método conclui que não existem valores outliers. Da mesma forma, se $\mathcal{K}_2 = \emptyset$, mesmo que $\mathcal{K}_1 \neq \emptyset$ então o método também conclui que não existem valores outliers nos dados em estudo.

Busca pelo Valor Adequado k

A admissibilidade da possível existência de valores outliers decorre da ocorrência de $\mathcal{K}_2 \neq \emptyset$. Mas este fato, a menos que \mathcal{K}_2 seja um conjunto unitário, demanda a escolha pelo valor $k \in \mathcal{K}_2$ mais adequado para o procedimento de detecção de outliers. Portanto, o método deve apresentar uma estratégia de procura para a escolha do valor k que seja capaz de fornecer os melhores

$$\text{BIC}(k) = \text{RSS} + \log(n) \times k \times p \quad (2)$$

Dessa forma, dada à admissibilidade da existência de valores extremos no conjunto em estudo, ou seja, $\mathcal{K}_2 \neq \emptyset$, o valor k utilizado será o valor k^* , aquele que minimiza o critério

$$k^* = \underset{k \in \mathcal{K}_2}{\text{argmax}} \text{BIC}(k) \quad (3)$$

Uma vez bem estabelecida a metodologia de detecção e valores outliers, o banco de dados associado a este estudo precisa ser detalhado. Este estudo discute resultados associados aos clubes disputantes do Campeonato Brasileiro de futebol. Inicialmente um detalhamento do processo de captação de dados será descrito.

A coleta de dados foi realizada a partir de consultas aos balanços financeiros dos 20 clubes participantes do Campeonato Brasileiro Série A do ano de 2022.

Foram consultados os balanços dos anos de 2016 até 2020. A legislação vigente no Brasil obriga os clubes a declarar sua situação financeira ao final de cada ano. Particularmente, era objetivo avaliar variáveis diretamente ligadas à arrecadação financeira das equipes.

Existem diversas fontes de receita, mas usualmente em todo o mundo, existem algumas fontes de receitas dominantes.

resultados no procedimento de detecção. A escolha do valor mais adequado para k apresentada no método DDCAM é determinada através do critério de informação Bayesiano (BIC).

O critério de informação Bayesiano (BIC) é uma métrica bastante utilizada em procedimentos de seleção de modelos. Trata-se de uma medida penalizadora para modelos mais pesados, com quantidade excessiva de parâmetros. O intuito é evitar os problemas de sobreajuste. O critério utiliza a função de verossimilhança e quanto menor o valor de BIC, maior a adequabilidade do modelo. Para dados previamente padronizados, a expressão $\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - c_j)^2 (x_{ij} - c_j)$ representa a soma de quadrados residual (RSS) e o valor do critério de informação Bayesiano pode ser reescrito pela equação (2):

BIC restrito aos valores k pertencente ao conjunto \mathcal{K}_2 , como definido pela equação (3):

Cinco fontes específicas de receita foram definidas como as variáveis de interesse: as verbas associadas à transferência de atletas, o matchday (arrecadação associada ao dia efetivo de jogo), as cotas de participações em competições e em transmissões televisivas e radiofônicas, as receitas relacionadas ao marketing e ao ambiente comercial, e em um último grupo, as demais receitas em geral.

Apesar da legislação vigente, algumas agremiações não publicam os balanços anuais.

Além disso, alguns balanços não se encontram adequadamente expostos, seja por páginas de internet inacessíveis para consulta ou outras dificuldades de acesso aos dados.

Diante disso, a captação de dados produziu um conjunto com a ocorrência de dados faltantes.

Os valores foram tabulados separados por temporadas para execução de procedimento de análise descritiva.

RESULTADOS

pelos clubes separados por temporadas em estudo.

A Figura 1 apresenta a distribuição entre os diversos tipos de receitas auferidas

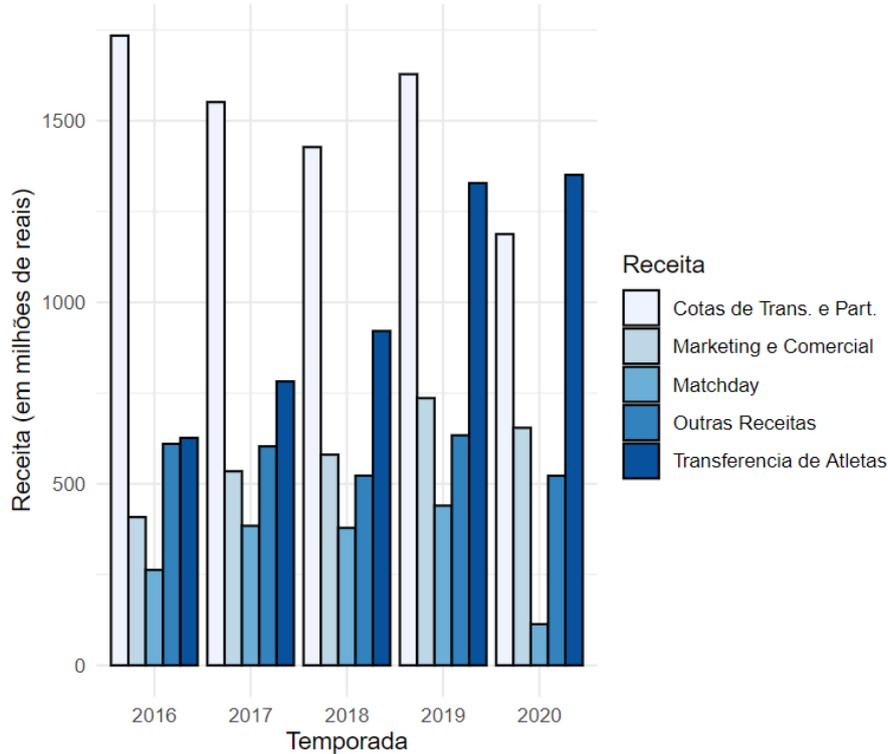


Figura 1 - Distribuição das receitas por temporada.

Nota-se como maior fonte de renda as cotas referentes à transmissão e participação. Em seguida, as transferências de atletas.

Por fim, o matchday é o tipo de receita que menos agrega valor aos clubes dentre as analisadas. As receitas distintas (com declarações variadas por diferentes clubes)

foram agregadas em outras receitas e apresentam patamares bastante semelhantes às receitas de marketing e comerciais.

A Figura 2 mostra a distribuição entre os tipos de receitas agora agregados com todas as temporadas em estudo.

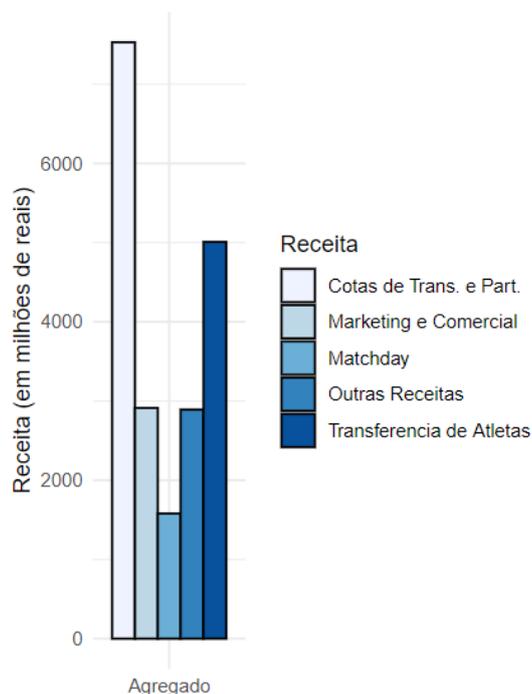


Figura 2 - Distribuição das receitas agregadas.

Verbas Associadas à Transferência de Atletas

A transferência de um atleta é um processo comercial entre dois clubes de futebol.

No Brasil, essa transação precisa respeitar as diretrizes previstas na Lei Pelé e no Regulamento do Estatuto de Transferência de Jogadores da FIFA, além de outras leis nacionais que devem incidir sobre os clubes brasileiros.

Em geral, os processos da transferência de jogadores são realizados por meio de intermediadores. De acordo com a legislação, o intermediário é o profissional que atua como representante de jogadores de

futebol, técnicos e/ou clubes, com interesse em negociar, celebrar, alterar ou renovar contratos de trabalho e transferência de jogadores.

Usualmente, em situações cuja proposta de transferências é feita para um atleta antes do término do seu atual contrato com o time em que atua, o clube interessado na transferência de jogador deve pagar uma compensação ao clube que possui o contrato em vigor.

Este valor é conhecido como taxa de transferência. Elevadas cifras são envolvidas nestes processos para os clubes de maior destaque no cenário brasileiro.

A Figura 3 apresenta os dados discriminados por temporada e o agregado das cinco temporadas sob investigação.

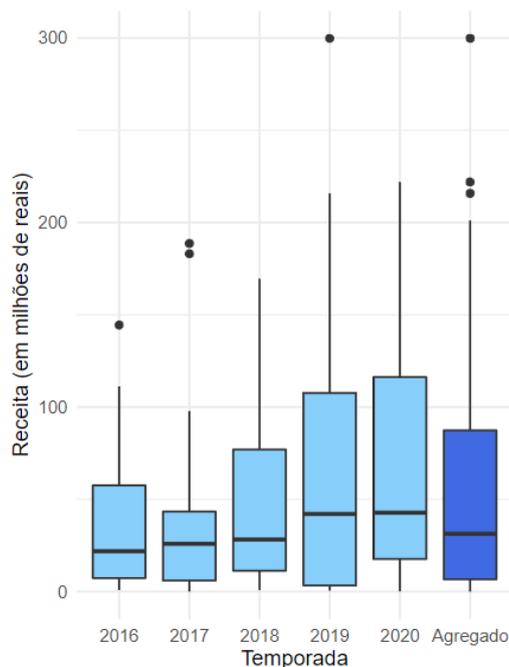


Figura 3 - Verbas de transferência de atletas.

A análise anterior ilustra um aumento de variabilidade na receita produzida através de vendas de atletas ao longo das temporadas analisadas.

Ocorreu também um aumento progressivo da mediana, porém este aumento parece bem sutil. Não ocorreu a presença de outliers univariados em todas as temporadas, apenas nas temporadas 2016, 2017 e 2019.

Em todas as situações os outliers univariados foram verificados para extremos máximos. Destacam-se o Sport Club Corinthians Paulista em 2016, o Clube de Regatas do Flamengo e o São Paulo Futebol Clube em 2017, e, novamente, o Clube de Regatas do Flamengo em 2019.

Quando as cinco temporadas são analisados de forma agregada, os outliers univariados que se apresentam são o Clube de Regatas do Flamengo (duas vezes) e o Club Athletico Paranaense.

Verbas Associadas ao matchday

As receitas relacionadas ao matchday são aquelas relacionadas aos dias de jogos dos clubes. Bilheteria e comércio são algumas das

principais formas de renda dos clubes nos dias que jogam como mandante. Além disso, alguns clubes apresentam gastos extras nestes dias de jogos, por não possuírem estádios próprios.

A Figura 4 apresenta os dados das verbas associadas à matchday por temporada e o agregado das temporadas.

Os gráficos demonstram aumento da variabilidade com o passar das temporadas. A mediana também exibe uma variação bem leve, indicando um ligeiro aumento.

Como exceção, se destaca a temporada do ano de 2020. Essa temporada apresenta uma variabilidade e uma mediana bem menores quando comparadas com as temporadas anteriores.

Como explicação, é importante citar a pandemia causada pelo Coronavírus (COVID-19), que acarretou fechamento dos estádios para a presença dos torcedores. Desta forma, as receitas relacionadas ao matchday sofreram uma queda brusca neste ano.

No que se diz respeito aos outliers, é nítida a grande quantidade de pontos destacados. Esse comportamento é previsível, já que os clubes não possuem as torcidas uniformemente distribuídas.

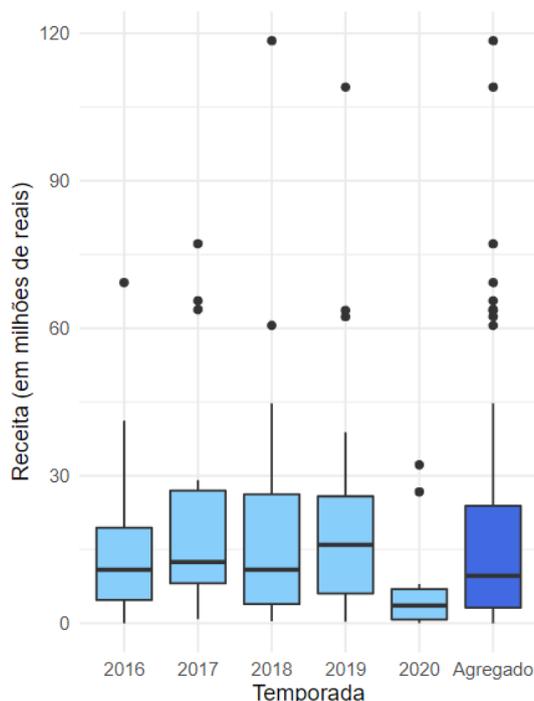


Figura 4 - Verbas de matchday.

Alguns clubes possuem uma quantidade mais elevada de torcedores. Consequentemente, estes clubes atraem uma quantidade maior de torcedores aos estádios.

Desta forma, a receita proveniente desta fonte é mais elevada para estes clubes. Outrossim, as disparidades econômicas brasileiras fazem com que o preço efetivo dos itens disponíveis sejam bastante heterogêneos por regiões.

A Sociedade Esportiva Palmeiras em 2016, o Sport Club Corinthians Paulista, o Clube de Regatas do Flamengo e a Sociedade Esportiva Palmeiras em 2017 e 2019, o Sport Club Corinthians Paulista e a Sociedade Esportiva Palmeiras em 2018 e o Clube de Regatas do Flamengo e o Santos Futebol Clube na temporada de 2020 foram os clubes detectados como outliers ao longo das temporadas.

Com as temporadas agregadas, o número de outliers é elevado. Os outliers são os mesmos, com exceção do Santos Futebol Clube, da temporada de 2020, que não foi considerado outlier com os dados agregados.

Verbas Associadas às Cotas de Transmissão e Participação

Outra importante forma de receita dos clubes são as cotas firmadas em direitos de transmissão e participações em campeonatos. Os direitos de transmissões são geralmente pactuados com alguma emissora televisiva ou radiofônica.

Desta forma, os direitos de imagem e de transmissão se tornam exclusivos para essa emissora, de forma que nem mesmo o próprio clube pode realizar transmissões de maneira independente.

Além disso, os clubes recebem uma receita relacionada às competições que disputam. A quantia destas receitas aumenta de acordo com o rendimento do clube no referido campeonato.

Por exemplo, a Copa do Brasil de Futebol é uma competição nacional composta por 92 clubes espalhados Brasil afora. A competição é no formato de partidas eliminatórias em que os clubes vencedores avançam de fase, popularmente conhecido como “mata-mata”.

Nestes tipos de competição, quanto mais o clube avança as fases, maior é o montante recebido. A maior recompensa que

pode ser conquistada é aquela adquirida quando se ganha a final e se sagra campeão do torneio.

A Figura 5 ilustra o comportamento destes tipos de verba. Verifica-se uma grande variabilidade na temporada de 2016, que, com o passar das temporadas, apresenta tendência de queda. O mesmo comportamento é notado

na mediana, que também sofre uma queda com o passar dos anos.

Outro ponto de destaque é a escassez de outliers, apenas a temporada 2020 apresentou essa situação, a Sociedade Esportiva Palmeiras. Quando as temporadas se apresentam de forma agregada, somente um outlier também é observado, neste caso, o Clube de Regatas do Flamengo.

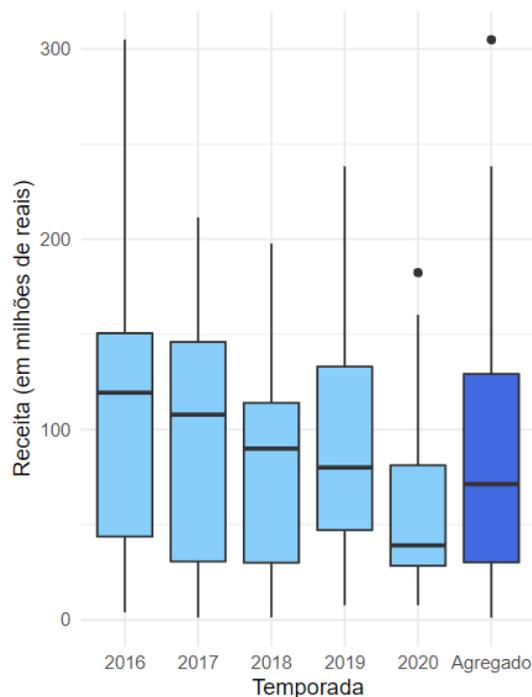


Figura 5 - Verbas de cotas de transmissão e participação.

Verbas Associadas ao Marketing e ao Comercial

Talvez uma das maneiras mais efetivas de geração de receitas dos clubes, os valores associados ao marketing e às ações comerciais dos clubes podem apresentar valores altíssimos.

A principal fonte neste tipo de receita são os patrocínios dos clubes, geralmente exibidos em seus uniformes e estádios.

Estes acordos de patrocínios são firmados em troca da exibição da marca da empresa contratante.

Os clubes em sua maioria apresentam mais de um patrocínio, exibindo assim várias marcas, principalmente em seus uniformes.

Outra forma de receita são as ações comerciais realizadas pelos clubes.

Como exemplo, neste mês de outubro de 2022, a fornecedora de materiais Adidas, juntamente com os clubes que são patrocinados pela mesma no Brasil, realizaram uma campanha de conscientização à prevenção do câncer de mama, a já conhecida campanha Outubro Rosa.

Desta forma, cada clube lançou uma camisa na cor rosa, como forma de recado ao cuidado à prevenção.

A Figura 6 apresenta a receita por temporada e seu agregado. Como padrão, nota-se a constância desta variável com o passar das temporadas.

Observa-se uma sutil modificação na variabilidade com o passar dos anos, com variações entre aumento e declínio. O mesmo comportamento é verificado para a mediana.

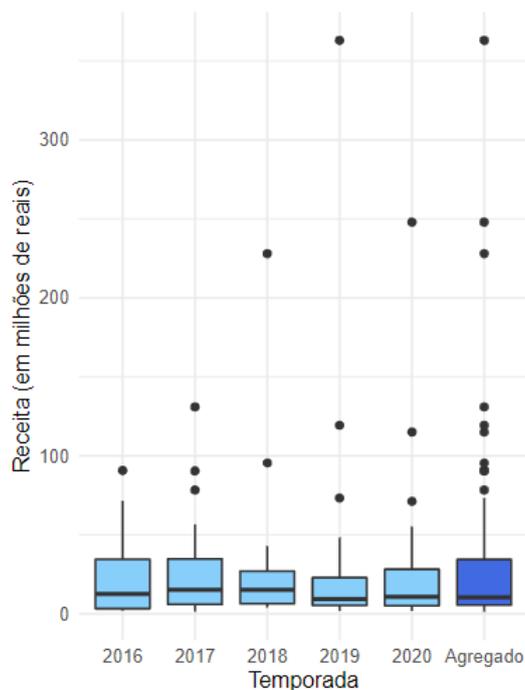


Figura 6 - Verbas associadas ao marketing e ao comercial.

Assim como na variável matchday, aqui também surpreende a elevada quantidade de outliers univariados observados.

Em 2016, a Sociedade Esportiva Palmeiras, em 2017, 2019 e 2020, o Sport Club Corinthians Paulista, o Clube de Regatas do Flamengo e a Sociedade Esportiva Palmeiras, por último, em 2018, o Clube de Regatas do Flamengo e a Sociedade Esportiva Palmeiras.

De maneira agregada, esta renda também apresenta diversos outliers.

Para a análise agregada das temporadas, todos os mesmos outliers se repetem nessa análise.

Verbas Associadas às Outras Receitas

Todas as receitas que não foram classificadas em alguma das quatro variáveis anteriores foram alocadas nesta última. Alguns clubes não detalham este tipo de receita em seus balanços.

Porém, outros detalham as formas de receita, como aluguel de imóveis, clubes de lazer e mensalidade de escolas de futebol de categorias iniciantes.

A Figura 7 demonstra a receita desta última variável.

Também neste tipo de renda ocorre um destaque para a grande variabilidade, que decai com o passar das temporadas. Igualmente a mediana apresenta um declínio suave ao longo das temporadas.

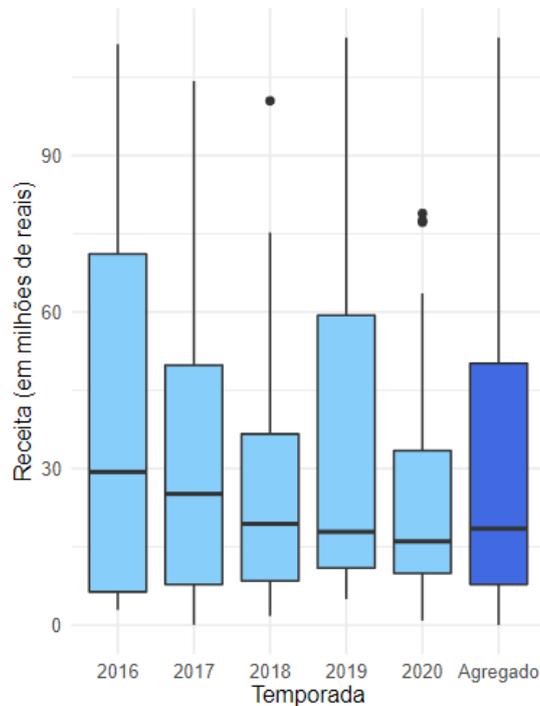


Figura 7 - Verbas de outras receitas.

Novamente, outro ponto de destaque é a ocorrência de poucos clubes classificados como outliers univariados.

Em 2018, a Sociedade Esportiva Palmeiras, em 2020 o Sport Club Internacional e, novamente, a Sociedade Esportiva Palmeiras são os únicos outliers destacados.

De maneira agregada, a variabilidade é muito expressiva e não ocorre a presença de outliers.

DISCUSSÃO

A literatura tende a afirmar que o resultado das partidas não está ligado exclusivamente ao aporte financeiro das equipes.

Por outro lado, ao investigar resultados de campeonatos e não apenas de partidas isoladas, o fator econômico tende a ter impactos decisivos.

Essa investigação, como dito outrora, busca verificar por meio de uma técnica inovadora a possível existência de alguma associação aparente entre as equipes outliers no que tange ao fator econômico com as equipes com melhor desempenho esportivo nos campeonatos.

Os dados analisados se referem ao futebol brasileiro nas temporadas de 2016 até 2020. Obviamente, o aspecto financeiro não ultrapassa somente um campeonato, mas permeia toda a temporada de um clube de futebol. Ao avaliar os resultados desportivos, a métrica será o Campeonato Brasileiro de Futebol da primeira divisão de cada uma das temporadas em estudo. Os torneios eliminatórios não serão considerados. Uma vez que este estudo propõe por tese que este efeito financeiro é mais contundente para os torneios de longo prazo como os campeonatos disputados com confrontos entre todas as equipes.

Nas temporadas em estudo, será verificada a possível presença de outliers multivariados para todas as variáveis em estudo: verbas de transferência de atletas, verbas de matchday, verbas de transmissão e participação, verbas de marketing e comerciais, e outras receitas. A busca por equipes outliers por meio do banco de dados multivariado será executada por meio da técnica DDCAM.

A Tabela 1 apresenta os resultados dos primeiros colocados dos campeonatos brasileiros de futebol das temporadas em estudo.

Tabela 1 - Desempenho de equipes nos campeonatos Brasileiros de Futebol.

	Campeão	Vice-campeão	3º colocado
2016	Palmeiras-SP	Santos-SP	Flamengo-RJ
2017	Corinthians-SP	Palmeiras-SP	Santos-SP
2018	Palmeiras-SP	Flamengo-RJ	Internacional-RS
2019	Flamengo-RJ	Santos-SP	Palmeiras-SP
2020	Flamengo-RJ	Internacional-RS	Atlético-MG

O método DDCAM identificou as equipes outliers multivariadas apresentadas na Tabela 2.

A busca foi executada para cada temporada e para dados agregados das cinco temporadas.

Tabela 2 - Equipes outliers verificadas por temporada e agregado.

2016	Corinthians-SP
2017	Palmeiras-SP
2018	Palmeiras-SP
2019	Flamengo-RJ
2020	Flamengo-RJ

Para a temporada de 2016, o Sport Club Corinthians Paulista foi detectado como equipe outlier em arrecadação.

O resultado desportivo não acompanhou, a equipe ficou em sétimo lugar no certame.

Porém, no ano seguinte o título ocorreu, talvez em decorrência do bom desempenho financeiro anterior. Apesar do título do Sport Club Corinthians Paulista em 2017, nessa temporada, a equipe outlier em arrecadação foi a Sociedade Esportiva Palmeiras, que também não alcançou o título, porém terminou o certame em segundo lugar.

Já nas temporadas seguintes dessa análise, sempre a equipe outlier multivariada verificada coincidiu com a equipe campeã do Campeonato Brasileiro de Futebol. Em 2018 a Sociedade Esportiva Palmeiras, em 2019 e 2020 o Clube de Regatas do Flamengo. J

á para análise de dados agregados, a presença da Sociedade Esportiva Palmeiras e do Clube de Regatas do Flamengo confirmam a constatação, trata-se das equipes que mais figuraram nas três primeiras posições ao longo das temporadas investigadas.

Todo este cenário, faz crer que uma arrecadação elevada, aliada a um superávit pode impactar no ano corrente, mas também no ano seguinte de disputas esportivas.

Diante disso, duas constatações específicas vêm à tona. Primeiro, o método DDCAM é hábil no processo de detecção de outliers multivariados.

Em sequência, um segundo fato, as equipes que destoam financeiramente em arrecadação são preponderantemente favoritas em campeonatos de longo prazo.

Claro que esta é uma investigação ainda incipiente, porém extremamente motivadora. Isso faz com que estudos futuros mais profundos sejam instigantes e necessários para subsidiar estes resultados.

CONCLUSÃO

Informações de séries históricas de dados tem alcançado grande utilização para os mais variados setores.

A capacidade computacional foi ampliada substancialmente nos últimos anos, isso permite o tratamento de conjuntos de dados cada vez maiores e mais complexos. Esta situação tem impulsionado a utilização de diversas técnicas estatísticas em análise de dados.

Como norte nessa linha de conduta, este estudo apresenta uma aplicação inovadora.

Esta investigação usou por base uma técnica específica para análise de dados e

estabeleceu uma percepção singular acerca da informação contida nos dados para direcionar tomadas de decisão.

Neste caso particular, o achado de pesquisa conduziu para decisões da área esportiva através de dados do futebol.

Um procedimento de análise de detecção de valores outliers foi à ferramenta utilizada neste estudo. A investigação e tratamento de elementos outliers é um tema de grande apelo científico e com uma vasta gama de aplicações. Diversas técnicas específicas, em diferentes cenários, podem ser observadas na literatura.

Este estudo foi conduzido para um conjunto importante de objetivos, alvos que foram efetivamente alcançados. O método DDCAM, um método inovador, foi apresentado de forma detalhada.

Além disso, o estudo apresentou um banco de dados específico, com informações de área esportiva (dados de clubes de futebol profissional no Brasil).

Uma análise descritiva abrangente foi realizada para o referido banco de dados. Por fim, a contribuição inédita deste estudo, a busca pela associação entre os resultados desportivos do futebol com os dados de arrecadação financeira das equipes disputantes, isso através de métodos de detecção de equipes outliers para variáveis financeiras de arrecadação.

O desempenho financeiro nem sempre é refletido no desempenho esportivo das equipes de futebol.

Porém, para análises de longo prazo, é possível verificar que os clubes melhores estruturados financeiramente são aqueles que possuem resultados esportivos dentro de campo mais efetivos.

Como consequência, acabam muitas vezes por se tornarem campeões dos torneios disputados. Os resultados das análises apontam para essa mesma conclusão. Em campeonatos mais longos compostos por muitas partidas, os clubes mais bem estruturados financeiramente acabam por mostrar alguma vantagem sobre os demais.

Geralmente, estes clubes possuem um melhor plantel de atletas, uma melhor estrutura física, e uma maior estrutura organizacional.

Esse conjunto de fatores, muitas vezes acaba por gerar uma grande diferença no final. Esse mesmo favoritismo, nem sempre é

confirmado quando levado em conta os campeonatos mais curtos, em geral compostos por partidas eliminatórias, como são os casos das copas nacionais, como por exemplo, a Copa do Brasil de Futebol. Para campeonatos desse tipo, é possível ver clubes com um poderio financeiro menor, alcançando sucesso contra clubes mais bem estruturados, revelando resultados inesperados.

Particularmente para os dados analisados, as equipes outliers em níveis de arrecadação financeira nas diversas variáveis de fontes de receita quase apresentam coincidência exata nos anos analisados com as equipes que alcançaram o título do campeonato brasileiro de futebol da série A.

Existe uma tendência de construção de ciclos virtuosos em clubes que alcançam receitas de faturamento elevadas. Os clubes com uma melhor organização financeira tendem a serem campeões, os clubes campeões, tendem a ter uma fonte de receita ainda maior nas temporadas seguintes.

Com essa fonte de renda mais elevada, estes clubes tendem a continuar no topo do desempenho esportivo, e, assim, o ciclo se forma. Os clubes que já se organizaram financeiramente tendem a dominar o esporte nos próximos anos, desde que mantenham esse nível de organização.

É o resultado visto com a análise dos dados agregados, que apontou a Sociedade Esportiva Palmeiras e o Clube de Regatas do Flamengo como equipes outliers de arrecadação financeira. Essas foram as equipes que mais acumularam títulos ao longo desse período no Brasil.

Este estudo apresenta possibilidades reais de propostas futuras de continuidade. O método DDCAM pode ser aplicado para dados detalhados internos de cada equipe, ou até mesmo para avaliações de desempenho de atletas.

Por exemplo, a possibilidade de detectar atletas com valências específicas que aumentam a probabilidade de resultados positivos para suas equipes. E assim, aproveitar esta informação em prol de toda a equipe.

Outra possibilidade de estudo próximo seria a aplicação deste mesmo tipo de estudo para dados do futebol em outros centros, como os países europeus com valores de arrecadação financeiros bastante vultosos.

REFERÊNCIAS

- 1-Aggarwal, C. C. editor. *Outlier Analysis*. New York. Springer International Publishing. 2017. p. 422.
- 2-Barbosa, J.J.; Duarte, A.R.; Martins, H.S.R. A Performance Evaluation in Multivariate Outliers Identification Methods. *Ciência & Natura*. Vol. 42. Num. e16. 2020. p. 1-14.
- 3-Barbosa, J.J.; Pereira, T.M.; Oliveira, F.L.P. Uma proposta para identificação de outliers multivariados. *Ciência & Natura*. Vol. 40. Num. E40. 2018. p. 1-8.
- 4-Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: identifying density-based local outliers. In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000. p. 93-104.
- 5-Daniel, P. 8º Valor das marcas dos clubes brasileiros: finanças dos clubes. In: *BDO RCS Auditores Independentes*. São Paulo. 2015.
- 6-Dantas, M.G.S.; Silva, J.A.; Boente, D.R. Detection of outliers: The financial performance of sport club corinthians paulista between 2008 and 2010. *Revista Ambiente Contabil*. Vol. 3. Num. 2. 2011. p. 17.
- 7-Duarte, A.R.; Babosa, J.J.; Martins, H.S.R.; Oliveira, F.L.P. Data-driven cluster analysis method: a novel outliers detection method in multivariate data. (Under review). 2022.
- 8-Duarte A.R.; Martins, H.S.R.; Silva, B.F. Correção de empates para modelagem dinâmica de partidas de Futebol. *Revista Brasileira de Futsal e Futebol*. São Paulo. Vol. 10. Num. 41. 2018. p. 774-784.
- 9-Ferreira, H.L.; Marques, J.A.V.C.; Macedo, M.A.S. Desempenho econômico-financeiro e desempenho esportivo: uma análise com clubes de futebol do Brasil. *Contextus - Revista Contemporânea de Economia e Gestão*. Vol. 16. Num. 3. 2018. p. 124-150.
- 10-Lejeune, C.; Mothe, J.; Soubki, A.; Teste, O. Shape-based outlier detection in multivariate functional data. *Knowledge-Based Systems*. Vol. 198. 2020. p. 105960.
- 11-Nascimento, C.L.D.; Dantas, M.G.S.; Azevedo, Y.G.P. A Influência dos Fatores Financeiros e Esportivos Sobre o Valor dos Clubes de Futebol Brasileiros. *Revista Evidenciação Contábil & Finanças*. Vol. 7. Num. 1. 2019. p. 94-111.
- 12-Wahid, A.; Rao, A.C.S. A distance-based outlier detection using particle swarm optimization technique. In: *Information and Communication Technology for Competitive Strategies*. Springer Singapore. 2019.

Autor correspondente:

Anderson Ribeiro Duarte.

anderson.duarte@ufop.edu.br

Universidade Federal de Ouro Preto.

Campus Morro do Cruzeiro.

Departamento de Estatística.

Ouro Preto-MG, Brasil.

CEP: 35400-000.

Recebido para publicação em 22/11/2022

Aceito em 18/01/2022