



**Universidade de Brasília
Departamento de Estatística**

**Análise Multivariada por fatores determinantes do desempenho de jogadores
do Brasileirão Série A - 2022**

Francisco Iago dos Reis Ferreira

Projeto apresentado para o Departamento
de Estatística da Universidade de Brasília
como parte dos requisitos necessários para
obtenção do grau de Bacharel em Es-
tatística.

**Brasília
2023**

Francisco Iago dos Reis Ferreira

**Análise Multivariada por fatores determinantes do desempenho de jogadores
do Brasileirão Série A - 2022**

Orientador: Prof. Gladston Luiz da Silva

Projeto apresentado para o Departamento
de Estatística da Universidade de Brasília
como parte dos requisitos necessários para
obtenção do grau de Bacharel em Es-
tatística.

**Brasília
2023**

Agradecimentos

Foi uma longa jornada até aqui, agradeço a todas as pessoas que contribuíram para a realização desse trabalho.

À minha mãe, a Sra. Francislene Maria dos Reis, por ter batalhado para que minha irmã e eu pudessemos ter a melhor educação possível. Todo o seu esforço não foi em vão.

À minha companheira, Evellyn Cristina Lopes Guerbe, cujo o amor e apoio incondicional foram cruciais para superar os momentos de indecisão.

Aos vários amigos que fiz durante toda a graduação, seja no curso de estatística, seja na Casa do Estudante, tenho muita consideração por cada um de vocês.

À meu orientados, Prof. Gladston Luis da Silva, pela empolgação em orientar esse trabalho, sua orientação e conselhos foram fundamentais para aprimorar minha pesquisa e direcionar meus esforços na direção certa.

E por último, à todos os professores do Departamento de Estatística e à Universidade de Brasília, por proporcionarem o ambiente de aprendizado e os recursos necessários para a concretização deste projeto.

Resumo

O Campeonato Brasileiro, também conhecido como Brasileirão, é a principal competição de futebol profissional no Brasil. Ele é disputado todos os anos por clubes que representam diversos estados de todo o país. Adotado em 2003, o formato de pontos corridos é disputado por 20 clubes que se enfrentam em jogos de ida e volta durante 38 rodadas, a equipe que fizer a maior pontuação se sagra como a campeã daquela edição. A competição disputada em 2022 nos forneceu uma vasta quantidade de dados acerca dos jogadores que dela participaram, e tais dados serão base para uma Análise Multivariada detalhada. Esta análise nos permitirá explorar as complexas interações entre as variáveis de desempenho.

Palavras-chaves: Brasileirão, Desempenho, Correlação, Análise Fatorial, Análise de Componentes Principais.

Abstract

The Campeonato Brasileiro, also known as the Brasileirão, is the main professional soccer competition in Brazil. It is contested every year by clubs representing various states across the country. Adopted in 2003, the straight points format is contested by 20 clubs who face each other in back-to-back matches over 38 rounds, with the team that scores the most points becoming the champion of that edition. The 2022 competition provided us with a vast amount of data about the players who took part in it, and this data will be the basis for a detailed Multivariate Analysis. This analysis will allow us to explore the complex interactions between the performance variables.

Keywords: Brasileirão, Performance, Correlation, Factor Analysis, Principal Component Analysis.

Lista de Tabelas

1	Descrição das variáveis dos goleiros	15
2	Descrição das variáveis de Chute	16
3	Descrição das variáveis de Passe	17
4	Descrição das variáveis de Criação de finalizações e gols	18
5	Descrição das variáveis de Ações defensivas	18
6	Descrição das variáveis de Posse	19
7	Descrição das variáveis Tempo de jogo	20
8	Descrição das variáveis Diversas	21

Lista de Figuras

Sumário

1 Introdução	8
2 Objetivos	10
2.1 Objetivo Geral	10
2.2 Objetivos Específicos	10
3 Referencial Teórico	11
3.1 Análise Estatística Descritiva	11
3.2 Análise Multivariada	11
4 Metodologia	12
5 Resultados	13
6 Conclusão	14
7 Anexo	15
7.1 Variáveis específicas dos goleiros	15
7.2 Variáveis para o jogadores de linha	16
7.2.1 Chute	16
7.2.2 Passe	17
7.2.3 Criação de finalizações e gols	18
7.2.4 Ações defensivas	18
7.2.5 Posse	19
7.2.6 Tempo de jogo	20
7.2.7 Diversas	21

1 Introdução

No ano de 2002, Billy Beane, diretor geral da equipe de beisebol *Oakland Athletics*, desafiou tudo que se acreditava sobre o esporte competitivo até então, ao adotar técnicas estatísticas avançadas para avaliar o desempenho de jogadores. Contrapondo-se ao critérios que eram utilizados pelos olheiros à época, como capacidade de correr, de arremessar, receber a bola, de rebater e rebater com potência. Billy e Paul dePodesta (até então um jovem executivo do clube), foram capazes de concluir que nem todas as características importantes num jogador de beisebol tem a mesma importância. Que velocidade na corrida, habilidade na defesa e força bruta tendiam a ser características demasiadamente superestimadas.

Eles foram capazes de mostrar que, características como porcentagem de rebatidas, porcentagem de base alcançadas, habilidade de controle da base de strike e etc. Essa abordagem revelou informações valiosas sobre a eficácia de um jogador, permitindo com que talentos muitas vezes subvalorizados por outras equipes fossem identificados.

Com um orçamento limitado, sendo o time com a terceira menor folha salarial da liga na época, mas com uma abordagem estatística autêntica, o time do Oakland Athletics quebrou o recorde de vitórias consecutivas, de uma temporada (LEWIS, 2004). O impacto dessa abordagem abriu caminho para uma nova era no esporte, extrapolou o beisebol e inspirou diversas modalidades. A utilização de técnicas estatísticas promoveu uma mudança na maneira com que o desempenho dos jogadores seriam avaliados e compreendidos.

No âmbito do futebol, é imperativo considerar o avanço relativo à informação contida nos dados. De acordo com Anderson e Sally (2013), os números têm o poder de desafiar conceitos pré-estabelecidos e dismantelar crenças antigas no esporte. Eles oferecem um entendimento do jogo como nunca visto antes, possibilitando uma visão contraposta aos pressupostos tradicionais.

Um exemplo de como os dados podem influenciar na tomada de decisão, é a renovação de contrato feita pelo jogador do Manchester City, Kevin de Bruyne¹. Na época o jogador conseguiu uma extensão salarial na casa dos 80 milhões de euros, utilizando apenas análise de dados ao invés de um agente².

No Brasil, o Brasileirão Série A é uma competição que movimenta milhões de

¹<https://www.transfermarkt.com.br/kevin-de-bruyne/profil/spieler/88755>

²<https://www.uol.com.br/esporte/futebol/ultimas-noticias/2021/04/08/big-data-e-sem-agente-como-de-bruyne-renovou-com-o-city.htm>

reais, instiga torcedores, investidores e apostadores todos os anos. Com o crescente avanço da tecnologia, o uso de técnicas estatísticas tem sido de imensa importância para se avaliar e entender a performance dos atletas durante a competição. No cenário nacional, por exemplo, times como o Flamengo³ e Palmeiras⁴ utilizam análise de desempenho para tomada de decisões administrativas e esportivas.

O presente trabalho, visa analisar, por meio de técnicas de Análise Multivariada, o desempenho de atletas da primeira divisão do Campeonato Brasileiro Série A 2022, pois permite uma avaliação mais precisa e completa, além de identificar padrões e tendências que não seriam encontradas por meio de análises mais simples. Por meio dessa análise, espera-se identificar as características que mais contribuíram para o desempenho dos jogadores durante o campeonato.

Tal abordagem aplicada ao contexto do campeonato brasileiro pode ser útil para os treinadores, apostadores, analistas de desempenho, investidores e patrocinadores, além de entusiastas por futebol, que poderão tomar melhores decisões no âmbito esportivo utilizando-se de informações munidas de análise estatística.

³<https://www.flamengo.com.br/noticias/futebol/flamengo-e-o-primeiro-clube-da-america-do-sul-a-implementar-c>

⁴<https://www.palmeiras.com.br/noticias/palmeiras-renova-parceria-com-empresa-de-software-de-analise-de-desempenho>

2 Objetivos

2.1 Objetivo Geral

Utilizar técnicas de Análise mMultivariada para identificar os principais fatores que influenciam o desempenho de atletas de futebol que jogaram o Campeonato Brasileiro da série A em 2022, considerando as posições em que esses jogadores atuaram.

2.2 Objetivos Específicos

Para alcançar o objetivo geral, será necessário realizar os seguintes passos:

- Coletar e preparar os dados de desempenho dos jogadores que participaram do campeonato;
- Realizar uma análise exploratória dos dados para identificar possíveis valores ausentes e inconsistências;
- Aplicar técnicas de redução de análise de dimensionalidade para selecionar as variáveis mais relevantes para a análise;
- Utilizar as técnicas de Análise Multivariada, para identificar padrões de desempenho entre os jogadores;
- Validar os resultados obtidos;
- Interpretar os resultados obtidos.

3 Referencial Teórico

Neste seção, serão descritos os métodos de Análise Estatística que serão empregados na pesquisa. Tais métodos foram selecionados devido à sua relevância dentro do contexto dos dados coletados e contribuirão para alcançar os objetivos descritos na seção anterior.

3.1 Análise Estatística Descritiva

É uma técnica muito utilizada para descrever as características essenciais de um conjunto de dados. Nos permite examinar os dados de forma quantitativa, como média, mediana, quartis, gráficos e etc. A intenção de utilizar essa análise é fornecer uma visão resumida das variáveis presentes, com o intuito de obter um entendimento dos dados, estabelecendo uma base sólida para a etapa subsequente (MORETTIN; BUSSAB, 2017).

3.2 Análise Multivariada

A Análise Multivariada é um conjunto de métodos estatísticos usados quando múltiplas variáveis são medidas simultaneamente em cada elemento amostral. Geralmente, são variáveis que estão correlacionadas entre si, e quanto maior o número, mais complexa se torna a análise. O objetivo é simplificar a interpretação do fenômeno estudado. Ela pode ser dividida em duas categorias principais: técnicas exploratórias que buscam simplificar a estrutura de variabilidade dos dados, sintetizando as variáveis, e técnicas de inferência.

4 Metodologia

5 Resultados

6 Conclusão

7 Anexo

7.1 Variáveis específicas dos goleiros

Tabela 1: Descrição das variáveis dos goleiros

Codificação	Resumo
G01	Número de partidas que o goleiro realizou no campeonato
G02	Minutos em que o goleiro esteve em campo
G03	Soma dos gols sofridos durante a competição
G04	Média de gols por partida
G05	Chutes no alvo
G06	Chutes defendidos
G07	Defesas realizadas, em porcentagem
G08	Jogos sem sofrer gols
G09	Gols de pênalti sofridos
G10	Pênaltis defendidos
G11	Passes longos certos (São considerados os passes acima de 36 metros)
G12	Tentativas de passe longo
G13	Total de passes tentados (inclui todo tipo de passe)
G14	Lançamentos feitos (Considerados aqueles feitos com as mãos)
G15	Aproveitamento de passe, em porcentagem
G16	Distância média, em metros, dos passes tentados (Não incluem os tiros de meta)
G17	Número de lançamentos adversários para a pequena área
G18	Interceptações de lançamentos a pequena área
G19	Número de ações defensivas fora da pequena área
G20	Distância média, em metros, das ações defensivas fora da pequena área

7.2 Variáveis para o jogadores de linha

7.2.1 Chute

Tabela 2: Descrição das variáveis de Chute

Codificação	Resumo
A01	Número de gols marcados
A02	Total de finalizações
A03	Total de finalizações ao alvo
A04	Total de finalizações ao alvo, em porcentagem
A05	Média de finalizações por jogo
A06	Média de finalizações certas por jogo
A07	Média de gols por finalização
A08	Média de gols por finalizações certas
A09	Distância média, em metros, de todas as finalizações (Não incluem os chutes de pênalti)
A10	Gols de Falta
A11	Pênaltis cobrados
A12	Gols de pênalti

7.2.2 Passe

Tabela 3: Descrição das variáveis de Passe

Codificação	Resumo
P01	Total de passes certos
P02	Total de passes tentados
P03	Aproveitamento dos passes, em porcentagem
P04	Distância total, em metros, de todos os passes realizados
P05	Distância progressiva dos passes, em metros (Considerados os passes em direção ao gol adversário)
P06	Passes curtos certos (Entre 4,5m e 13,7m)
P07	Tentativas de passes curtos
P08	Aproveitamento nos passes curtos, em porcentagem
P09	Passes médios certos (Entre 13,7m e 27,4m)
P10	Tentativas de passes médios
P11	Aproveitamento nos passes médios, em porcentagem
P12	Passes longos certos (Acima de 27,4m)
P13	Tentativas de passes longos
P14	Aproveitamento nos passes longos, em porcentagem
P15	Assistências para gol
P16	Passes que geraram um chute ao gol
P17	Passes no último terço do campo (Terço defensivo do time adversário)
P18	Passes completos para a grande área
P19	Cruzamentos completos para a grande área
P20	Passes que movem a bola em qualquer direção a linha de gol à pelo menos 9,1m dos seus ou qualquer passe concluído para a área adversária

7.2.3 Criação de finalizações e gols

Tabela 4: Descrição das variáveis de Criação de finalizações e gols

Codificação	Resumo
GC01	Total de ações que geraram finalizações durante o campeonato
GC02	Total de ações médias que geraram finalizações (durante 90 minutos)
GC03	Total de ações que geraram gols durante o campeonato
GC04	Total de ações médias que geraram gols (durante 90 minutos)

7.2.4 Ações defensivas

Tabela 5: Descrição das variáveis de Ações defensivas

Codificação	Resumo
DEF01	Número de divididas
DEF02	Número de vezes em que o time ganhou a posse de bola após uma dividida
DEF03	Divididas no terço defensivo do campo
DEF04	Divididas no terço central do campo
DEF05	Divididas no terço de ataque do campo
DEF06	Duelos individuais ganhos (Considera-se as vezes em que o jogador frustrou um drible adversário)
DEF07	Total de duelos individuais
DEF08	Porcentagem de duelos ganhos
DEF09	Duelos individuais perdidos
DEF10	Total de vezes em que bloqueou a bola
DEF11	Número de vezes em que bloqueou um chute
DEF12	Número de vezes em que bloqueou um passe
DEF13	Número de interceptações (Vezes que, conscientemente, o jogador impediu uma ação ofensiva)
DEF14	Número de divididas mais número de interceptações
DEF15	Erros individuais que levaram a uma finalização do time adversário

7.2.5 Posse

Tabela 6: Descrição das variáveis de Posse

Codificação	Resumo
PO01	Número de vezes que o jogador tocou na bola
PO02	Toques na grande área de defesa
PO03	Toques no terço defensivo
PO04	Toques no terço central
PO05	Toques no terço de ataque
PO06	Toques na grande área de ataque
PO07	Toques com a bola em jogo
PO08	Número de tentativas de drible
PO09	Tentativas de drible bem sucedidas
PO10	Porcentagem de dribles bem sucedidos
PO11	***
PO12	***
PO13	Número de vezes que o jogador dominou a bola
PO14	Distância total, em metros, que a bola foi movida após o domínio (Até o toque seguinte na bola)
PO15	Distância progressiva que a bola foi movida após o domínio (Até o toque seguinte na bola)
PO16	Vezes que o jogador se moveu na direção do gol adversário (Pelo menos 9 metros do ponto inicial)
PO17	Domínios no terço ofensivo
PO18	Domínios na área adversária
PO19	Vezes que o jogador não conseguiu dominar a bola
PO20	Número de vezes que o jogador foi desarmado após o domínio da bola
PO21	Número de vezes que o jogador recebeu um passe
PO22	Passes que moveram a bola em direção ao gol adversário

7.2.6 Tempo de jogo

Tabela 7: Descrição das variáveis Tempo de jogo

Codificação	Resumo
T01	Jogos disputados
T02	Minutos que esteve em campo (Total)
T03	Média de minutos por partida
T04	Porcentagem de minutos jogados *
T05	Jogos como titular
T06	Média de minutos como titular
T07	Partidas completas (O jogador não foi substituído)
T08	Jogos como reserva
T09	Minutos médios em campo após sair do banco de reservas
T10	Jogos que o jogador esteve no banco de reservas mas não atuou

7.2.7 Diversas

Tabela 8: Descrição das variáveis Diversas

Codificação	Resumo
M01	Cartões Amarelos
M02	Cartões Vermelhos
M03	2ª Cartão amarelo (No mesmo jogo)
M04	Faltas cometidas
M05	Faltas sofridas
M06	Impedimentos
M07	Cruzamentos*
M08	Cortes
M09	Divididas em que o próprio jogador ficou com a bola
M10	Pênaltis convertidos*
M11	Pênaltis concedidos*
M12	Gols contra
M13	Bolas recuperadas
M14	Disputas aéreas ganhas
M15	Disputas aéreas perdidas
M16	Porcentagem de disputas aéreas ganhas

Referências

ANDERSON, C.; SALLY, D. *The numbers game: Why everything you know about soccer is wrong*. [S.l.]: Penguin, 2013.

LEWIS, M. *Moneyball: The art of winning an unfair game*. [S.l.]: WW Norton & Company, 2004.

MORETTIN, P. A.; BUSSAB, W. O. *Estatística Básica*. [S.l.]: Saraiva Educação SA, 2017.