



**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(a): 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**

## Sumário

<b>1 Introdução</b>	4
<b>2 Objetivos</b>	5
2.1 Objetivo Geral	5
2.2 Objetivos Específicos	5
<b>3 Metodologia</b>	6
3.1 CRISP-DM	6
<b>4 Referencial Teórico</b>	7
4.1 Análise Estatística Descritiva	7
4.2 Análise Multivariada	7
4.3 Análise de Componentes Principais	7
4.3.1 Componentes Principais Populacionais	8
4.4 Análise Fatorial	11
4.4.1 Modelo Fatorial	11
4.4.2 Estrutura da Covariância	12
4.4.3 Metodo de Estimação	13
<b>5 Banco de Dados</b>	14
5.1 Descrição do Banco	14
5.2 Posições	14
<b>6 Cronograma</b>	16

# 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 futebol, o avanço no que tange à informação que os dados têm a dizer não poderia ser negligenciada. Segundo Anderson e Sally (2013), são os números que vão contrapor os conceitos pré definidos e destruir crenças antigas. São os números que irão nos permitir ter um entendimento do jogo como nunca visto antes.

No Brasil, o Brasileirão Série A é uma competição que movimenta milhões de 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 Flamengo (2018) e Palmeiras (2019) 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 jo-

gadores 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.

## 2 Objetivos

### 2.1 Objetivo Geral

Utilizar técnicas de Análise Multivariada 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 por meio de testes;
- Interpretar os resultados obtidos para descobrir *insights* que possam ser úteis para os analistas.

## 3 Metodologia

### 3.1 CRISP-DM

Nessa fase, serão definidas todos os métodos para a elaboração do projeto, que serão baseadas na metodologia **CRISP-DM**, esse modelo irá permitir que a análise seja padronizada em etapas, de modo que o mesmo fique mais coerente, estruturado e eficiente.

- **Fase 1 - Entendimento do problema:** Na etapa inicial, será feita uma análise preliminar dos dados para entender o problema. Também será feita uma revisão de literatura sobre análise multivariada e a seleção do método mais adequado para o estudo dos dados.
- **Fase 2 - Entendimento dos dados:** Nesta etapa, será realizada uma análise exploratória dos dados para entender o seu formato e a qualidade. Serão realizadas técnicas exploratórias para avaliar a distribuição, identificar possíveis valores faltantes e analisar a correlação entre as variáveis.
- **Fase 3 - Preparação dos dados:** Após a etapa anterior, os dados serão limpos e preparados para a análise em si, isso inclui a seleção das variáveis mais relevantes, a transformação de variáveis, se necessário, a criação de novas variáveis e a exclusão de dados inconsistentes.
- **Fase 4 - Modelagem:** Será realizada a análise multivariada dos dados para entender o desempenho dos jogadores. Serão utilizados métodos como análise fatorial, análise de cluster, análise discriminante entre outros. Além de serem realizados testes de hipóteses para avaliar os resultados obtidos.
- **Fase 5 - Avaliação:** Os resultados obtidos serão avaliados em relação aos objetivos do trabalho e às hipóteses formuladas. Caso necessário, serão realizadas análises adicionais para melhorar os resultados já obtidos e obter conclusões mais pertinentes.
- **Parte 6 - Implementação:** Finalmente, os resultados obtidos durante toda a análise serão apresentados.

## 4 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.

### 4.1 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).

### 4.2 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. Um dos métodos comuns de simplificação da estrutura de variabilidade é a redução de dimensionalidade, que visa diminuir o número de variáveis originais mantendo as informações essenciais para a análise.

### 4.3 Análise de Componentes Principais

Segundo (JOLLIFFE, 2002), a essência da técnica de Análise de Componentes Principais é a redução da complexidade de um conjunto de dados que contém inúmeras variáveis( $p$ ). Essa redução busca preservar ao máximo a variabilidade dos dados originais. Esse tipo de simplificação pode ser alcançada por meio da transformação dos dados em um novo conjunto de variáveis chamadas de componentes principais( $k$ ). Tais componentes não possuem uma correlação entre si e são organizados de modo que as primeiras componentes possam reter a maior parte da variação presente no conjunto original dos dados.

O desenvolvimento matemático expresso na sessão 4.3.1 foi extraído da obra de (JOHNSON; WICHERN et al., 2002).

#### 4.3.1 Componentes Principais Populacionais

Esse método tem por objetivo principal explicar a estrutura da variância/covariância dos dados através de combinações lineares das variáveis, onde as  $p$  variáveis são retidas em  $k$  componentes. Podemos escrever esses componentes como combinações lineares das  $p$  variáveis  $X_1, X_2, \dots, X_p$ . Geometricamente, essas combinações representam a seleção de um novo sistema de coordenadas obtido por meio da rotação do sistema original com  $X_1, X_2, \dots, X_p$  como o eixo de coordenadas, tais eixos representam as direções com a variabilidade máxima e nos fornecem uma descrição mais parcimoniosa da estrutura de covariância.

Seja o vetor aleatório  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  com matrix de covariância  $\Sigma$  com autovetores próprios  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = 0$

Considere as combinações lineares:

$$Y_1 = a'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (4.3.1)$$

$$Y_2 = a'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (4.3.2)$$

$$\vdots \quad (4.3.3)$$

$$Y_p = a'_p X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \quad (4.3.4)$$

que pode ser escrita como

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}_{p \times 1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}_{p \times p} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}_{p \times 1}$$

Onde

$$E(Y_i) = E(a'_i \underline{X}) \quad (4.3.5)$$

$$E(Y_i) = a'_i E(\underline{X}) \quad (4.3.6)$$

$$E(Y_i) = a'_i \underline{\mu} \quad (4.3.7)$$



$$Var(Y_i) = V(a'_i \underline{X}) \quad (4.3.8)$$

$$Var(Y_i) = a'_i V(\underline{X}) a_i \quad (4.3.9)$$

$$Var(Y_i) = a'_i \Sigma a_i, \quad i = 1, 2, \dots, p \quad (4.3.10)$$

e

$$Cov(Y_i, Y_k) = V(a'_i \underline{X}, a'_k \underline{X}) \quad (4.3.11)$$

$$Cov(Y_i, Y_k) = a'_i \Sigma a_k, \quad i, k = 1, 2, \dots, p \quad (4.3.12)$$

São combinações lineares não correlacionadas de  $Y_1, Y_2, \dots, Y_p$  no qual a variância é máxima. Onde, podemos definir a partir disso:

- A 1ª componente principal será a combinação linear de  $a'_1 X$  que maximizará a  $V(a'_1 X)$  sujeita à restrição  $a'_1 a_1 = 1$
- A 2ª componente principal será a combinação linear de  $a'_2 X$  que maximizará a  $V(a'_2 X)$  sujeita à restrição  $a'_2 a_2 = 1$  e  $Cov(a'_1 X, a'_2 X) = 0$

⋮

- A i-ésimaª componente principal será a combinação linear de  $a'_i X$  que maximizará a  $V(a'_i X)$  sujeita à restrição  $a'_i a_k = 1$  e  $Cov(a'_i X, a'_k X) = 0$  para  $k < i$

Seja  $\Sigma$  a matriz de covariância associada ao vetor aleatório  $\mathbf{X}' = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]$ , com pares de autovalores-autovetores  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ , onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , então a i-ésimaª componente principal é dada por

$$Y_i \approx e'_i X = e_{i1} X + e_{i2} X + \dots + e_{ip} X \quad i = 1, 2, \dots, p \quad (4.3.13)$$

Com

$$Var(Y_i) = e'_i \Sigma e_i = \lambda_i \quad i = 1, 2, \dots, p \quad (4.3.14)$$

$$Cov(Y_i, Y_k) = e'_i \Sigma e_k = 0, \quad i \neq k \quad (4.3.15)$$

E seja,  $\mathbf{X}' = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]$  com matriz de covariância  $\Sigma$ , e com pares de autovalores-autovetores  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ , onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , tomemos  $Y_i \approx e_i' X = e_i 1X + e_i 2X + \dots + e_i pX \quad i = 1, 2, \dots, p$ , como os componentes principais. Então

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (4.3.16)$$

Se  $Y_1 = e_1' X = Y_2 = e_2' X, \dots, Y_p = e_p' X$  são os componentes principais obtidos através da matriz de covariância  $\Sigma$ , então

$$\rho_{Y_i X_k} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p \quad (4.3.17)$$

são os coeficientes de correlação entre as componentes  $(Y_i)$  e as variáveis  $(X_k)$ . E  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$  são os pares de autovalores-autovetores para a matriz  $\Sigma$ . Alternativamente, também é possível obter a proporção total da  $k$ -ésima<sup>a</sup> componente principal por meio da relação:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad i, k = 1, 2, \dots, p \quad (4.3.18)$$

## 4.4 Análise Fatorial

A Análise Fatorial, segundo (HAIR et al., 2009), é uma técnica que fornece as ferramentas necessárias para examinar a estrutura das relações entre as variáveis, identificando o conjunto que apresentam correlações fortes, conhecidas como fatores.

Em outras palavras, essa técnica é utilizada para explorar a estrutura subjacente de um conjunto de dados, com a missão de identificar fatores latentes que explicam a correlação entre as variáveis observadas.

O desenvolvimento matemático expresso na sessão 4.4.1, 4.4.2 foi extraído da obra de (JOHNSON; WICHERN et al., 2002).

### 4.4.1 Modelo Fatorial

O vetor aleatório  $\mathbf{X}$ , com  $p$  componentes, tem média  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Sigma}$ . No modelo fatorial  $\mathbf{X}$ , é linearmente dependente em relação à algumas variáveis aleatórias não observáveis  $F_1, F_2, \dots, F_m$ , chamadas de fatores comuns, e  $p$  fontes de adicionais de variação  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , chamados de erros ou de fatores específicos. Em particular o modelo pode ser descrito como:

$$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \quad (4.4.1)$$

$$X_2 - \mu_2 = \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \quad (4.4.2)$$

$$\vdots \quad (4.4.3)$$

$$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \quad (4.4.4)$$

Que pode ser expressa em notação matricial

$$(\mathbf{X} - \boldsymbol{\mu})_{p \times 1} = \mathbf{L}_{p \times m} \mathbf{F}_{m \times 1} + \boldsymbol{\varepsilon}_{p \times 1} \quad (4.4.5)$$

- $\mu$  média da variável  $i$
- $\ell_{ij}$  são as cargas fatoriais da  $i$ -ésima variável no  $j$ -ésimo fator;
- $\mathbf{L}$  é a matriz das cargas fatoriais;
- $F_j$  são as variáveis latentes; e
- $\varepsilon_i$  são os erros não observáveis

Ainda do modelo geral, assume-se que:

- $E(F) = 0_{m \times 1}$ ;
- $Cov(F) = E(FF') = I_{m \times m}$
- $E(\varepsilon) = 0_{p \times 1}$
- $Cov(\varepsilon) = \Psi_{p \times p}$

Onde, as seguintes condições são satisfeitas:

- $\mathbf{F}$  and  $\epsilon$  são independentes
- $E(\mathbf{F}) = 0$ ,  $Cov(\mathbf{F}) = \mathbf{I}$
- $E(\epsilon) = 0$ ,  $Cov(\epsilon) = \Psi$ , tal que  $\Psi$  é a matriz ortogonal

Ainda no modelo fatorial ortogonal, a covariância pode ser representada como:

$$\begin{aligned} (X - \mu)(X - \mu)' &= (LF + \epsilon)(LF + \epsilon)' \\ &= (LF + \epsilon)((LF)' + \epsilon') \\ &= LF(LF)' + \epsilon(LF)' + LF\epsilon' + \epsilon\epsilon' \end{aligned}$$

Então, aplicando a esperança na expressão, temos

$$\begin{aligned} E(X - \mu)(X - \mu)' &= \\ &= LE(FF')L' + E(\epsilon\epsilon')L' + LE(F\epsilon') + E(\epsilon\epsilon') \\ &= LL' + \Psi \end{aligned}$$

$$\Sigma = LL' + \Psi \tag{4.4.6}$$

#### 4.4.2 Estrutura da Covariância

$$1. Cov(\mathbf{X}) = \mathbf{LL}' + \mathbf{\Psi}$$

ou

$$\begin{aligned} Var(X_i) &= \ell_{i1}^2 + \dots + \ell_{im}^2 + \Psi_i \\ Cov(X_i, X_k) &= \ell_{i1}\ell_{k1} + \dots + \ell_{im}\ell_{km} \end{aligned}$$

$$2. Cov(\mathbf{X}, \mathbf{F}) = \mathbf{L}$$

ou

$$Cov(X_i, F_j) = \ell_{ij}$$

A parte da variância da  $i$ -ésima variável contribuída pelos  $m$  fatores comuns é chamada de comunalidade. A parte da  $Var(X_i) = \sigma_{ii}$  devida ao fator específico é chamada de singularidade ou variância específica. Denotando-se a  $i$ -ésima comunalidade por  $h_i^2$ :

$$\sigma_{ii} = \ell_{i1}^2 + \dots + \ell_{im}^2 + \Psi_i$$

Onde

$$\begin{aligned} h_i^2 &= \ell_{i1}^2 + \dots + \ell_{im}^2 \\ \sigma_{ii} &= h_i^2 + \Psi_i \quad i = 1, 2, \dots, p \end{aligned}$$

Em outras palavras, a comunalidade medirá a proporção da variância total de uma variável observada que pode ou não ser explicada pelos fatores identificados no modelo. Quanto mais próximo de 1, a variável será completamente explicada pelos fatores, e quanto mais próxima de 0, a variável é única e não é explicada por nenhum dos fatores.

#### 4.4.3 Metodo de Estimação

Descrever o que for utilizar na análise...

## 5 Banco de Dados

### 5.1 Descrição do Banco

Os dados que serão analisados foram coletados manualmente e organizados em tabelas Excel. Eles foram obtidos da seguinte fonte: **FBref**<sup>1</sup>. O conjunto consiste em registros dos jogadores referentes aos 20 clubes que disputaram o campeonato brasileiro do ano de 2022. Serão considerados dados de 752 jogadores dos clubes que participaram das 38 rodadas da competição, cujas análises irão considerar a área de atuação dos atletas em campo - defesa, meio de campo e ataque -, ao invés da posição propriamente dita do jogador.

### 5.2 Posições

No futebol, as posições desempenham um papel importante no funcionamento tático de uma equipe. Cada jogador é atribuído para uma posição específica dentro do campo. O objetivo principal é que ele ocupe aquele espaço de maneira estratégica para que contribua com o sucesso coletivo do time.

- **Goleiros** - Uma das posições mais importantes dentro do esporte certamente é a de goleiro. Sua principal função dentro de campo é impedir que o time adversário marque gols. Além de serem ágeis, eles também precisam ter boa coordenação motora para serem capazes de se movimentar rapidamente entre as traves.
- **Defensores** - Atuam na linha de defesa e podem ser divididos em zagueiros e laterais. Os zagueiros são responsáveis por marcar os atacantes adversários, desarmando-os e evitando que eles cheguem ao gol. Os laterais são um pouco mais versáteis, pois além de atuar na defesa auxiliando os zagueiros, eles também apoiam o ataque ajudando a criar jogadas ofensivas.
- **Meio Campistas** - São os jogadores que atuam entre as linha de defesa e ataque, sendo responsáveis por recuperar a posse de bola, distribuir passes, controlar o ritmo de jogo e armar jogadas ofensivas. Podem ser divididos em volantes, que têm uma função mais defensiva, ajudando na marcação e no desarme do time adversário e os meio-ofensivos que ajudam na criação de jogadas, dando assistências e marcando gols.
- **Atacantes** - São os jogadores responsáveis por marcar gols. Atuam na linha de

---

<sup>1</sup><https://fbref.com/en/comps/24/2022/2022-Serie-A-Stats>

frente tendo como principal função finalizar as jogadas. Podem ser divididos em centroavantes, que são os jogadores que atuam mais centralizados, buscam sempre se posicionar bem para finalizar as jogadas, e os pontas, que atuam principalmente nas laterias do campo, geralmente são jogadores rápidos e habilidosos, sendo responsáveis por driblar os adversário e criar situações de gols.

Os dados para os jogadores de linha consistem em variáveis que se encaixam em algum tipo de aspecto do jogo. São eles: **Informações gerais, Chute, Passe, Tipo de passe, Ações defensivas, Tempo de jogo e Status diversos**. Já os goleiros possuem estatísticas próprias que serão analisadas separadamente.

## 6 Cronograma

As etapas e a forma em que foram organizadas podem ser visualizadas a partir da tabela abaixo.

Tabela 1: Cronograma

	1/2023				2/2023				
	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									

- |  |   |
|--|---|
| 1. Escolha do tema a ser abordado.         | 10. Entrega do relatório parcial ao Prof. Orientador. |
| 2. Desenvolvimento da proposta de projeto. | 11. Correção do relatório parcial.                    |
| 3. Entrega da proposta de projeto.         | 12. Entrega do relatório parcial para a banca.        |
| 4. Revisão de literatura.                  | 13. Desenvolvimento do modelo.                        |
| 5. Elaboração da apresentação da proposta. | 14. Elaboração do relatório final.                    |
| 6. Apresentação oral da proposta.          | 15. Entrega do relatório final ao Prof. Orientador.   |
| 7. Manipulação do banco de dados.          | 16. Correção do relatório final.                      |
| 8. Análise exploratória do banco de dados. | 17. Entrega do relatório final para a banca.          |
| 9. Elaboração do relatório parcial.        |   |



## Referências

- ANDERSON, C.; SALLY, D. *The numbers game: Why everything you know about soccer is wrong*. [S.l.]: Penguin, 2013.
- FLAMENGO, C. *Flamengo é o primeiro clube da América do Sul a implementar o SAP Sports One*. 2018. <<https://www.flamengo.com.br/noticias/futebol/flamengo-e-o-primeiro-clube-da-america-do-sul-a-implementar-o-sap-sports-one>>. Acesso em 12 de maio de 2023.
- HAIR, J. F. et al. *Análise multivariada de dados*. [S.l.]: Bookman editora, 2009.
- JOHNSON, R. A.; WICHERN, D. W. et al. *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002.
- JOLLIFFE, I. T. *Principal component analysis for special types of data*. [S.l.]: Springer, 2002.
- 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.
- PALMEIRAS, D. de C. *Palmeiras renova parceria com empresa de software de análise de desempenho*. 2019. <<https://www.palmeiras.com.br/noticias/palmeiras-renova-parceria-com-empresa-de-software-de-analise-de-desempenho/>>. Acesso em 12 de maio de 2023.