



**UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"**

Campus Presidente Prudente

PEDRO MARQUES GUARNIERI

**UM RANKING ESTATÍSTICO DO DESEMPENHO DOS ATLETAS DA NBA NAS
TEMPORADAS 2019-20 A 2023-24**

Revisado pelo Orientador

Documento assinado digitalmente
 **KLAUS SCHLUNZEN JUNIOR**
Data: 17/12/2025 07:31:57-0300
Verifique em <https://validar.it.gov.br>

Assinatura do Orientador

Data ____ / ____ / 2025

Presidente Prudente

2025

PEDRO MARQUES GUARNIERI

**UM RANKING ESTATÍSTICO DO DESEMPENHO DOS ATLETAS DA NBA NAS
TEMPORADAS 2019-20 A 2023-24**

Relatório Final para Trabalho de Conclusão de
Curso apresentado ao Curso de Graduação em
Estatística da FCT/Unesp para aproveitamento na
disciplina Trabalho de Conclusão de Curso.

Orientador: Prof. Dr. Klaus Schlünzen Junior

PRESIDENTE PRUDENTE

2025

G916r

Guarnieri, Pedro Marques

Um ranking estatístico do desempenho dos atletas da nba nas temporadas 2019-20 a 2023-24 / Pedro Marques Guarnieri. --

Presidente Prudente, 2025

157 f. : il., tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Estatística)

Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Klaus Schlunzen Junior

1. Análise de componentes principais. 2. Análise por agrupamento.
3. Análise fatorial. 4. Basquetebol. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

TERMO DE APROVAÇÃO

Pedro Marques Guarnieri

UM RANKING ESTATÍSTICO DO DESEMPENHO DOS ATLETAS DA NBA NAS TEMPORADAS 2019-20 A 2023-24

Relatório de Final de Trabalho de Conclusão de Curso aprovado como requisito para obtenção de créditos na disciplina Trabalho de Conclusão do curso de graduação em Estatística da Faculdade de Ciências e Tecnologia da Unesp, pela seguinte banca examinadora:

Documento assinado digitalmente
 **KLAUS SCHLUNZEN JUNIOR**
Data: 17/12/2025 07:30:43-0300
Verifique em <https://validar.iti.gov.br>

Orientador: _____

Prof. Dr. Klaus Schlunzen Junior

Departamento de Estatística

Documento assinado digitalmente
 **MANOEL IVANILDO SILVESTRE BEZERRA**
Data: 17/12/2025 13:17:01-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Manoel Ivanildo Silvestre Bezerra

Departamento de Estatística

Presidente Prudente, 12 de dezembro de 2025.

RESUMO

Este trabalho desenvolve um ranking estatístico objetivo do desempenho dos atletas da NBA nas temporadas 2019-20 a 2023-24, utilizando técnicas multivariadas para superar avaliações subjetivas tradicionais. Foram coletados dados via web scraping do Basketball Reference, abrangendo métricas como pontos, assistências, rebotes, eficácias de arremessos e indicadores defensivos, com tratamento para consolidação de registros duplicados e segmentação por posição (PG, SG, SF, PF, C), considerando apenas jogadores com ≥ 20 jogos e 300 minutos. A metodologia incluiu análises exploratórias (estatísticas descritivas, boxplots, histogramas, correlogramas), Análise de Componentes Principais (PCA, com 75-89% de variância explicada pelos 3 primeiros componentes), clusterização K-médias ($k=2-4$ clusters por posição, validada por silhueta e cotovelo), e Análise Fatorial Exploratória (AFE) por posição, com extração máxima verossimilhança, rotação oblimin, testes KMO/Bartlett e escores fatoriais padronizados para rankings normalizados (0-100). Os resultados revelam perfis distintos por posição — armadores priorizando AST/PTS/X3P., pivôs ORB/DRB/BLK/X2P. —, com tops como Joel Embiid (1º geral 2022-23), Nikola Jokić (1º C 2020-21) e Shai Gilgeous-Alexander (3º geral 2022-23), comparados a MVPs para validar coerência. Conclui-se que a abordagem promove transparência quantitativa no basquete, alinhada aos ODS 9 (inovação) e 3 (saúde), fornecendo rankings robustos e insights sobre perfis via clusters.

Palavras-chave: análise de agrupamento; análise de componentes principais; K-médias; análise fatorial exploratória; NBA.

ABSTRACT

This study develops an objective statistical ranking of NBA players' performance across the 2019-20 to 2023-24 seasons, employing multivariate techniques to overcome traditional subjective evaluations. Data were collected via web scraping from Basketball Reference, encompassing metrics such as points, assists, rebounds, shooting efficiencies, and defensive indicators, with processing for duplicate record consolidation and segmentation by position (PG, SG, SF, PF, C), considering only players with ≥ 20 games and 300 minutes. The methodology included exploratory analyses (descriptive statistics, boxplots, histograms, correlograms), Principal Component Analysis (PCA, with 75-89% variance explained by the first 3 components), K-means clustering ($k=2-4$ clusters per position, validated by silhouette and elbow methods), and Exploratory Factor Analysis (EFA) by position, using maximum likelihood extraction, oblimin rotation, KMO/Bartlett tests, and standardized factorial scores for normalized rankings (0-100). Results reveal distinct profiles by position—point guards prioritizing AST/PTS/X3P., centers ORB/DRB/BLK/X2P.—with top performers like Joel Embiid (1st overall 2022-23), Nikola Jokić (1st C 2020-21), and Shai Gilgeous-Alexander (3rd overall 2022-23), compared to MVPs for validation. The approach promotes quantitative transparency in basketball, aligned with SDGs 9 (innovation) and 3 (health), providing robust rankings and cluster-based performance insights.

Keywords: cluster analysis; principal component analysis; K-means; exploratory factor analysis; NBA.

SUMÁRIO

1 INTRODUÇÃO.....	5
2 REFERENCIAL TEÓRICO.....	5
3 METODOLOGIA.....	8
3.1 DADOS	8
3.1.1 Coleta e composição da base de dados	8
3.1.2 Tratamento e consolidação dos dados	8
3.1.3 Segmentação por posição	9
3.1.4 Considerações iniciais	9
4 ANÁLISE DE COMPONENTES PRINCIPAIS	10
5 ANÁLISE DE AGRUPAMENTO	12
6 K-MÉDIAS.....	13
6.1 MÉTODO DA SILHUETA.....	14
7 ANÁLISE FATORIAL EXPLORATÓRIA.....	15
8 RESULTADOS	17
8.1 APRESENTAÇÃO DOS DADOS.....	17
8.2 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2019-2020.....	19
8.2.1 Estatísticas das variáveis da temporada 2019-2020	19
8.2.2 Análise gráfica das variáveis da temporada 2019-2020	22
8.2.3 Análise de componentes principais(PCA) 2019-2020.....	28
8.2.4 Clusterização por K-médias 2019-2020	35
8.2.5 Análise fatorial exploratória (AFE) e construção dos índices por posição 2019-2020	36
8.2.6 Rankings por posição 2019-2020	43
8.2.7 Ranking geral 2019-2020	46
8.3 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2020-2021.....	48
8.3.1 Estatísticas das variáveis dos dados da temporada 2020-2021.....	49
8.3.2 Análise gráfica das variáveis da temporada 2020-2021	51
8.3.3 Análise de componentes principais 2020-2021	57
8.3.4 Clusterização por k-médias 2020-2021	62
8.3.5 Análise fatorial exploratória por posição 2020-2021	63
8.3.6 Construção de rankings por posição 2020-2021.....	69
8.3.7 Ranking geral 2020-2021	73
8.4 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2021-2022.....	76
8.4.1 Estatísticas das variáveis da temporada 2021-2022	76
8.4.2 Análise gráfica das variáveis da temporada 2021-2022	79
8.4.3 Análise de componentes principais 2021-2022	85
8.4.4 Clusterização por k-médias 2021-2022	90
8.4.5 Análise fatorial exploratória por posição 2021-2022	91
8.4.6 Construção dos rankings por posição 2021-2022.....	97

8.4.7 Ranking geral 2021-2022	100
8.5 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2022-2023.....	102
8.5.1 Estatísticas das variáveis da temporada 2022-2023	102
8.5.2 Análise gráfica das variáveis da temporada 2022-2023	104
8.5.3 Análise de componentes principais 2022-2023	110
8.5.4 Clusterização por k-médias 2022-2023	115
8.5.5 Análise fatorial exploratória 2022-2023	116
8.5.6 Construção dos ranking por posição 2022-2023	122
8.5.7 Ranking geral 2022-2023	125
8.6 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2023-2024.....	127
8.6.1 Estatísticas das variáveis da temporada 2023-2024	127
8.6.2 Análise gráfica das variáveis da temporada 2023-2024	130
8.6.3 Análise de componentes principais 2023-2024	135
8.6.4 Clusterização por k-médias 2023-2024	140
8.6.5 Análise fatorial exploratória por posição 2023-2024	141
8.6.6 Construção de rankings por posição 2023-2024.....	147
8.6.7 Ranking geral 2023-2024	150
9 CONCLUSÕES	153
REFERÊNCIAS	155

1 INTRODUÇÃO

O basquetebol da NBA representa um ambiente altamente competitivo, onde detalhes mínimos podem influenciar significativamente os resultados. Com o avanço do esporte, a aplicação de métodos estatísticos avançados tornou-se essencial para avaliar o desempenho dos atletas de forma objetiva e estratégica. Este trabalho propõe o uso de técnicas multivariadas e modelos estatísticos para analisar, entre as temporadas de 2019-20 a 2023-24, os desempenhos dos jogadores da NBA, considerando variáveis como pontuação, assistências, rebotes e eficiências ofensiva e defensiva.

O objetivo central é desenvolver um sistema de ranking que classifique os atletas por posição (armadores, alas-armadores, alas, alas-pivôs e pivôs), com base exclusivamente em métricas quantitativas. Além disso, os resultados obtidos serão comparados com as premiações de MVP (Most Valuable Player), permitindo avaliar a coerência entre as análises estatísticas e os critérios subjetivos dessas premiações.

A proposta dialoga com os Objetivos de Desenvolvimento Sustentável (ODS), especialmente o ODS 9, ao aplicar inovação e infraestrutura analítica no contexto esportivo, e, de forma indireta, com o ODS 3, ao fomentar práticas mais saudáveis e eficientes no treinamento e na gestão do desempenho atlético. Ao fornecer uma abordagem quantitativa para reconhecimento de talentos, este estudo contribui para a modernização das avaliações esportivas, promovendo maior transparência, justiça e precisão na mensuração do mérito individual no basquete profissional.

2 REFERENCIAL TEÓRICO

A análise estatística no esporte é uma área crescente que utiliza dados quantitativos para monitorar e melhorar o desempenho de atletas e equipes. O uso de estatísticas avançadas no esporte permite uma avaliação mais objetiva do desempenho, substituindo percepções subjetivas por análises concretas e contribuindo para o desenvolvimento de estratégias embasadas em dados (Cabral, 2022; Tiquinho, 2024). Na NBA, métricas como pontos, assistências, rebotes e eficiência de arremessos são amplamente utilizadas para mensurar o desempenho dos jogadores (Alves, 2023). Essas estatísticas básicas, em conjunto com métricas avançadas, formam um panorama abrangente das habilidades e do impacto dos jogadores em suas equipes (Almas, 2015; Canan; Mendes; Silva, 2015).

Para que uma análise estatística seja efetiva, é necessário selecionar métricas que representem adequadamente as habilidades dos jogadores em suas diversas posições. A teoria de análise estatística no esporte propõe a seleção de métricas que abranjam múltiplas dimensões do desempenho, desde pontuação até aspectos defensivos (ALMAS, 2015; JUNIOR, 2002). Cada posição no basquete exige um conjunto específico de habilidades e responsabilidades, conforme discutido por Junior (2002), que também analisou diferenças de desempenho com base em categorias de jogos e mando de quadra, o que torna fundamental a escolha de métricas ajustadas a essas necessidades (Chua, 2023; Hair et al., 2009).

Os métodos multivariados são fundamentais para análises que envolvem várias variáveis inter-relacionadas, como a avaliação de desempenho de jogadores, que exige a consideração simultânea de múltiplos fatores. Técnicas como Análise de Componentes Principais (ACP) e Análise de Agrupamento são comumente aplicadas para organizar os dados de forma que padrões e relações possam ser mais facilmente identificados. Segundo Johnson e Wichern (2007), a ACP permite sintetizar múltiplas variáveis em componentes principais, destacando os fatores mais influentes. A Análise de Agrupamento, por sua vez, é útil para identificar grupos de jogadores com habilidades semelhantes, permitindo uma segmentação por padrões de desempenho (Bussab; Morettin, 2003).

A modelagem estatística é uma etapa essencial para transformar dados em conhecimento acionável. Conforme exposto por Agresti (2013), modelos estatísticos como regressão múltipla, análise discriminante e redes neurais permitem prever variáveis de interesse a partir de múltiplos fatores explicativos. Neste estudo, a modelagem será empregada para desenvolver um modelo de ponderação das habilidades dos jogadores da NBA, ajustando as variáveis conforme as características específicas de cada posição (Jovanovic, 2019).

O uso de rankings é comum na avaliação de desempenho esportivo, pois facilita a comparação direta entre atletas e pode ser ajustado para refletir as habilidades exigidas por cada posição. No basquete, por exemplo, armadores e pivôs têm responsabilidades e habilidades distintas, de modo que o modelo de ranking deve considerar essas diferenças para ser justo e representativo (Alves, 2023).

As premiações de MVP (Most Valuable Player) representam um reconhecimento subjetivo do valor de um jogador, e estudos como os de Oliver (2004) mostram que esses prêmios nem sempre refletem objetivamente o desempenho (Chua, 2023). Comparar os

rankings estatísticos desenvolvidos neste trabalho com os prêmios de MVP permitirá analisar a correspondência (ou falta dela) entre as avaliações subjetivas e as análises baseadas em dados (Almas, 2015; Chua ,2023).

3 METODOLOGIA

3.1 DADOS

Nesta seção, apresenta-se a caracterização da base de dados utilizada no estudo, incluindo informações sobre sua origem, tratamento, estrutura e segmentação. Esta etapa é fundamental para compreender a composição do banco de dados e embasar as análises multivariadas subsequentes.

3.1.1 Coleta e composição da base de dados

Os dados utilizados neste trabalho foram obtidos por meio de técnicas de extração automatizada de informações da internet, denominadas web scraping (Mitchell, 2018). A coleta foi realizada por meio da linguagem de programação R, utilizando a biblioteca “rvest”, que permite a raspagem estruturada de páginas HTML (Lawson, 2015).

As informações foram extraídas do site Basketball Reference (<https://www.basketball-reference.com>), uma das principais plataformas públicas de estatísticas da National Basketball Association (NBA), amplamente utilizada para análises de desempenho no basquete (Alves, 2023; Chua, 2023).

As bases de dados extraídas possuem um total de 31 colunas, que abrangem variáveis categóricas e numéricas. Dentre essas variáveis, destacam-se: nome do atleta, idade, posição em quadra, nome da equipe, pontos totais, média de pontos por jogo, número de partidas disputadas, porcentagem de aproveitamento nos arremessos, assistências, rebotes, além de indicadores de premiações individuais na temporada, como presença em seleções da liga ou títulos de MVP (Most Valuable Player).

Essas informações representam múltiplas dimensões do desempenho técnico, físico e tático dos atletas ao longo da temporada regular da NBA, fornecendo uma base robusta para a análise estatística e para a construção dos rankings de desempenho propostos neste estudo.

3.1.2 Tratamento e consolidação dos dados

Durante a unificação das bases, observou-se a presença de múltiplos registros para determinados jogadores em uma mesma temporada, situação causada por transferências de equipe ao longo do campeonato. Nesses casos, o site original apresenta os dados de cada passagem por equipe em linhas separadas.

Para evitar duplicações e garantir a consistência analítica, os dados foram tratados conforme os seguintes procedimentos: identificação de jogadores com registros duplicados, consolidação estatística por jogador (utilizando, quando necessário, médias ponderadas pelo tempo em quadra), e posterior exclusão das observações redundantes. Como resultado, o número total de linhas por temporada foi reduzido de aproximadamente 700–800 observações para cerca de 500 registros únicos, cada um representando um jogador por temporada.

3.1.3 Segmentação por posição

Após a consolidação da base geral, os jogadores foram classificados conforme suas posições predominantes em quadra, conforme disponibilizado pelo *Basketball Reference*. Foram consideradas as cinco posições tradicionais do basquete profissional: armador (*point guard* – PG), ala-armador (*shooting guard* – SG), ala (*small forward* – SF), ala-pivô (*power forward* – PF) e pivô (*center* – C). Para cada posição, foi gerado um subconjunto da base contendo exclusivamente os atletas identificados naquela função.

Tal segmentação é essencial para a condução das análises de desempenho por posição, considerando as especificidades táticas e estatísticas esperadas de cada papel desempenhado no jogo.

3.1.4 Considerações iniciais

A etapa de coleta e preparação dos dados mostrou-se essencial para garantir a integridade e a confiabilidade das análises estatísticas posteriores. A utilização de *web scraping* com a linguagem R e o pacote *rvest* permitiu a extração eficiente de dados atualizados, enquanto os procedimentos de limpeza, padronização e segmentação asseguraram a coerência da base. Dessa forma, estabeleceu-se uma estrutura de dados robusta, adequada para a aplicação de técnicas multivariadas e para a construção de rankings de desempenho dos atletas da NBA.

4 ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (PCA – Principal Component Analysis) é uma técnica estatística multivariada amplamente utilizada para a redução de dimensionalidade de conjuntos de dados que apresentam um grande número de variáveis inter-relacionadas. Seu principal objetivo é transformar um conjunto de variáveis originais em um novo conjunto, denominado componentes principais, os quais são combinações lineares das variáveis originais, construídas de forma que retenham a maior parte da variabilidade presente nos dados.

Ao aplicar a PCA, é possível identificar quais combinações de variáveis explicam melhor as variações observadas no conjunto de dados, permitindo a síntese da informação de forma mais compacta e interpretável. Essa redução de dimensionalidade facilita a visualização dos dados, a detecção de padrões e a posterior aplicação de outras técnicas estatísticas, como análise de agrupamento (clusterização) e modelagem.

No contexto deste trabalho, a PCA será aplicada com o intuito de: (i) eliminar possíveis redundâncias entre métricas de desempenho dos jogadores, (ii) reduzir a complexidade dos dados mantendo o máximo de variância explicada, e (iii) preparar os dados para etapas subsequentes como a clusterização e a criação de modelos de pontuação. A utilização da PCA é especialmente relevante, dado que muitas das métricas escolhidas para avaliar os jogadores apresentam correlação entre si.

A matriz de correlação R simétrica de ordem “ $p \times p$ ”, é dada por:

$$R = \begin{pmatrix} 1 & r(x_1x_2) & r(x_1x_3) & \cdots & r(x_1x_p) \\ r(x_2x_1) & 1 & r(x_2x_3) & \cdots & r(x_2x_p) \\ r(x_3x_1) & r(x_3x_2) & 1 & \cdots & r(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(x_px_1) & r(x_px_2) & r(x_px_3) & \cdots & 1 \end{pmatrix} \quad (1)$$

A determinação dos componentes principais é feita por meio da equação característica da matriz de correlação:

$$\det(R - \lambda I_p) = P_A(\lambda) \quad (2)$$

A partir da equação característica, calculam-se os autovalores e autovetores da matriz R , onde o “det” representa o determinante da matriz R , I_p é a matriz identidade de ordem p , e os valores de λ que anulam o determinante correspondem aos autovalores associados a R .

Para cada autovalor λ_i existe um autovetor $\tilde{a}_i$:

$$\tilde{a}_i^T = (a_{i1} \ a_{i2} \ \dots \ a_{ip}) \quad (3)$$

Os autovetores indicam a contribuição de cada variável para a formação dos componentes principais, enquanto os autovalores representam a quantidade de variância explicada por cada componente.

A aplicação da PCA impactará diretamente a estrutura analítica do estudo, pois os componentes principais extraídos substituirão (parcial ou totalmente) as variáveis originais em algumas etapas da análise. Isso garantirá maior robustez estatística e evitará problemas como multicolinearidade, que podem distorcer os resultados em modelos estatísticos.

Como próximos passos, será realizada a padronização das variáveis (transformação para z-score), condição essencial para que a PCA seja aplicada de forma adequada. Em seguida, serão estimados os componentes principais com base nas variáveis selecionadas para cada posição. Será avaliado o percentual de variância explicada por cada componente, utilizando gráficos de autovalores (scree plot), e os componentes com maior relevância serão selecionados para uso nas análises subsequentes.

5 ANÁLISE DE AGRUPAMENTO

A análise de agrupamento, também conhecida como clusterização, é uma técnica estatística multivariada que tem como objetivo principal identificar padrões ou estruturas subjacentes em conjuntos de dados, organizando as observações em grupos homogêneos com base em suas similaridades. Esses grupos, denominados clusters, são formados de modo que as observações dentro de um mesmo grupo sejam mais semelhantes entre si do que em relação a observações de outros grupos (Hair et al., 2009; Johnson; Wichern, 2007).

No contexto deste trabalho, a clusterização será utilizada para agrupar jogadores da NBA com base nas métricas estatísticas selecionadas para cada posição. A ideia é identificar perfis de jogadores que compartilham características de desempenho semelhantes, mesmo que não ocupem o mesmo nível de destaque dentro da liga. Isso permite explorar diferentes tipos de contribuição dentro de uma mesma posição, auxiliando na identificação de grupos de atletas com estilos de jogo ou eficiências comparáveis (Chua, 2023; Alves, 2023).

Aplicar métodos de agrupamento neste estudo tem como principais objetivos: (i) complementar a análise descritiva e a PCA com uma visão exploratória baseada em padrões de semelhança, (ii) apoiar a construção de rankings com maior contexto e nuance, permitindo destacar jogadores dentro de seus respectivos clusters, e (iii) oferecer interpretações mais ricas sobre os diferentes perfis estatísticos existentes em cada posição (Jovanovic, 2019).

A clusterização será aplicada sobre os componentes principais gerados pela PCA como a utilizada por Galerani (2024) em seu estudo sobre montagem de times ideais na NBA com base em agrupamentos e PCA, a clusterização neste trabalho será aplicada, o que garante a ausência de multicolinearidade e reduz a dimensão do espaço de análise, favorecendo a estabilidade dos agrupamentos formados. Será utilizado o método K-means como abordagem principal, com avaliação do número ótimo de clusters por meio de métricas como a soma dos quadrados intra-grupos (inertia) e o método do cotovelo (elbow method). Alternativamente, também poderão ser testados métodos hierárquicos ou baseados em densidade, caso os dados apresentem estruturas não esféricas.

Como próximos passos, será realizada a aplicação da clusterização individualmente para cada posição, com base nos componentes extraídos da PCA. Cada jogador será classificado em um grupo, e a interpretação de cada cluster permitirá compreender diferentes tipos de

atuação dentro da mesma função, além de embasar análises posteriores sobre o ranqueamento e comparação entre atletas.

6 K-MÉDIAS

O método de clusterização K-médias (ou K-means) é uma técnica estatística de aprendizado não supervisionado amplamente utilizada para particionar um conjunto de observações em grupos (clusters) baseados em suas similaridades. Seu funcionamento baseia-se na minimização da variabilidade intra-cluster e na maximização da variabilidade entre clusters, atribuindo cada observação ao grupo cuja média (centroide) está mais próxima (Hair et al., 2009; Johnson; Wichern, 2007).

O algoritmo inicia com a escolha de um número pré-definido de clusters (k) e uma seleção inicial de centroides. A cada iteração, as observações são atribuídas ao cluster cujo centroide está mais próximo, com base na distância euclidiana (ou outra métrica definida). Em seguida, os centroides são recalculados como a média dos pontos atribuídos a cada grupo, e o processo é repetido até que os centroides converjam ou não haja mais mudanças significativas nas atribuições (Chua, 2023).

Neste trabalho, o método K-médias será utilizado para agrupar os jogadores da NBA com base nas componentes principais obtidas por meio da PCA. A utilização dessa abordagem permite explorar padrões estatísticos de desempenho de forma não supervisionada, formando grupos de atletas com características semelhantes. Isso possibilita compreender melhor os diferentes perfis de atuação em cada posição, enriquecendo a análise dos rankings com uma perspectiva de segmentação baseada em dados (Alves, 2023; Almas, 2015).

distância euclidiana:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^p (x_{ij} - c_{kj})^2} \quad i = 1, \dots, n; k = 1, \dots, K; i \neq k \quad (4)$$

Em que x_{ij} representa o valor da j-ésima variável (ou componente principal) para o i-ésimo jogador, c_{kj} é a coordenada do centroide do k-ésimo cluster nessa mesma variável, p é o número total de variáveis consideradas. A escolha do número ótimo de clusters será feita por meio do método do cotovelo (elbow method), que analisa a variação da soma dos quadrados

intra-cluster conforme diferentes valores de k . O ponto onde a redução da variância começa a se estabilizar é interpretado como a melhor escolha para o número de grupos. Complementarmente, poderão ser usados índices estatísticos de validação interna, como o índice de silhueta.

A aplicação do K-médias neste estudo será realizada separadamente para cada posição de jogador, utilizando os componentes principais padronizados como base. Ao final do processo, cada jogador será alocado em um grupo, e esses clusters serão interpretados à luz das métricas originais para caracterizar os diferentes perfis de desempenho observados dentro de cada posição.

6.1 MÉTODO DA SILHUETA

O índice de silhueta é uma métrica estatística utilizada para avaliar a qualidade de uma partição obtida por métodos de clusterização. Essa medida fornece uma estimativa de quão bem cada observação está posicionada em seu cluster em comparação com outros clusters, permitindo julgar a coerência interna de cada grupo formado (Hair et al., 2009; Johnson; Wichern, 2007).

Para cada observação, o índice de silhueta é calculado com base em duas medidas: (i) a distância média entre essa observação e todas as outras do mesmo cluster (a); e (ii) a distância média entre essa observação e os pontos do cluster mais próximo (b). A silhueta é então definida como:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad i = 1, 2, \dots, n \quad (5)$$

Em que $s(i)$ é o índice de silhueta da observação i , com i representando cada jogador da amostra; $a(i)$ é a distância média entre o jogador i e os demais jogadores pertencentes ao mesmo cluster (distância média intra-cluster); e $b(i)$ é a menor distância média entre o jogador i e os jogadores dos demais clusters, isto é, a distância média ao cluster vizinho mais próximo.

Os valores do índice de silhueta variam entre -1 e 1. Valores próximos de 1 indicam que a observação está bem agrupada e distante de outros clusters. Valores próximos de 0 indicam que a observação está na fronteira entre dois grupos, e valores negativos sugerem que a observação pode ter sido alocada no cluster incorreto (Chua, 2023).

Neste trabalho, o índice de silhueta será utilizado como critério complementar ao método do cotovelo para determinar o número ideal de clusters (k) a ser adotado na aplicação do método K-médias. A média do índice de silhueta de todas as observações será analisada para

diferentes valores de k , sendo o valor com maior média interpretado como o número ótimo de grupos. Essa abordagem fornece uma validação empírica adicional para a definição da estrutura de agrupamento mais adequada aos dados (Alves, 2023; Cabral, 2022).

7 ANÁLISE FATORIAL EXPLORATÓRIA

A Análise Fatorial Exploratória (AFE) é uma técnica estatística multivariada utilizada para identificar estruturas latentes subjacentes a um conjunto de variáveis observadas. Seu principal objetivo é reduzir a dimensionalidade dos dados, agrupando variáveis correlacionadas em fatores comuns que representam dimensões latentes não diretamente observáveis, mas inferidas a partir das correlações entre as variáveis (Hair et al., 2009; Johnson; Wichern, 2007).

No contexto deste trabalho, a AFE foi escolhida como método estatístico principal para a construção dos rankings de desempenho dos jogadores da NBA por posição. A justificativa para essa escolha reside na capacidade da técnica de sintetizar múltiplas métricas estatísticas em um conjunto reduzido de fatores interpretáveis, o que possibilita a criação de um escore fatorial padronizado representativo do desempenho global de cada jogador (Almas, 2015; Cabral, 2022).

A AFE permite, por exemplo, agrupar métricas relacionadas à eficiência ofensiva, como também apontado por Cabral (2022) ao demonstrar como fatores latentes podem sintetizar múltiplos aspectos do desempenho no basquete, distribuição de jogo ou desempenho defensivo em fatores distintos, facilitando a interpretação dos dados e a comparação entre atletas. A construção dos rankings será baseada nos escores fatoriais padronizados obtidos para cada jogador, de modo que os valores mais altos representam melhor desempenho dentro do grupo posicional analisado (Alves, 2023; Chua, 2023).

A aplicação da AFE é realizada separadamente para cada posição (armador, ala-armador, ala, ala-pivô e pivô), considerando apenas as métricas estatísticas pré-selecionadas para cada função. O procedimento metodológico adotado envolve: (i) verificar a adequação da matriz de correlações por meio dos testes KMO e de esfericidade de Bartlett; (ii) extrair os fatores pelo método de máxima verossimilhança (ML); (iii) aplicar rotação oblíqua do tipo oblimin, permitindo correlação entre os fatores latentes; e (iv) calcular escores fatoriais padronizados para cada jogador pelo método de regressão, que posteriormente são utilizados na construção dos índices de desempenho por posição e do índice geral. O número de fatores é

definido com base em análise paralela e, quando necessário, variáveis com comunalidade inferior a um limiar pré-estabelecido ($h^2 < 0,30$) são removidas, desde que o modelo permaneça com pelo menos três variáveis observadas.

O uso da Análise Fatorial Exploratória contribuirá diretamente para a robustez e a objetividade dos rankings produzidos, reduzindo a influência de variáveis redundantes e capturando de forma mais precisa os diferentes perfis de desempenho existentes entre os jogadores. A técnica também permitirá validar empiricamente a estrutura das métricas selecionadas para cada posição, garantindo maior consistência metodológica ao estudo (Bussab; Morettin, 2003).

8 RESULTADOS

Nesta seção serão apresentados os resultados das análises realizadas e suas respectivas interpretações.

8.1 APRESENTAÇÃO DOS DADOS

Nesta seção, apresenta-se a estrutura da base de dados utilizada para a análise estatística do desempenho dos jogadores da NBA entre as temporadas de 2019-20 e 2023-24. A base contempla uma ampla gama de variáveis, categóricas e numéricas, que descrevem características individuais, estatísticas de atuação em quadra e eventuais prêmios recebidos pelos atletas ao longo das temporadas. A Tabela a seguir organiza essas variáveis, indicando o nome de cada uma, seu tipo (categórica ou numérica) e uma breve descrição de seu conteúdo. Esta caracterização é fundamental para compreender o escopo da análise desenvolvida ao longo deste trabalho, permitindo identificar os principais indicadores de desempenho considerados na construção dos modelos estatísticos e na elaboração dos rankings por posição.

Quadro 1 - Tipos das Variáveis

Variável	Tipo	Descrição
Player	Categórica	Nome do jogador
Age	Numérica	Idade do jogador
Team	Categórica	Equipe da NBA em que o jogador atuou naquela temporada
Pos	Categórica	Posição em quadra(PG, SG, SF, PF, C)
G	Numérica	Total de jogos disputados
Gs	Numérica	Total de jogos como titular
MP	Numérica	Minutos jogados
FG	Numérica	Arremessos convertidos

FGA	Numérica	Tentativas de arremessos
FG%	Numérica	Aproveitamento em arremessos de quadra
3P	Numérica	Cestas de 3 pontos convertidas
3PA	Numérica	Tentativas de cestas de 3 pontos
3P%	Numérica	Aproveitamento nos arremessos de 3 pontos
2P	Numérica	Cestas de 2 pontos convertidas
2PA	Numérica	Tentativas de cestas de 2 pontos
2P%	Numérica	Aproveitamento nas cestas de 2 pontos
eFG%	Numérica	Aproveitamento efetivo de arremessos
FT	Numérica	Lances livres convertidos
FTA	Numérica	Tentativas de lances livres
FT%	Numérica	Aproveitamento em lances livres
ORB	Numérica	Rebotes ofensivos
DRB	Numérica	Rebotes defensivos
TRB	Numérica	Total de rebotes
AST	Numérica	Assistências

STL	Numérica	Roubos de bola
BLK	Numérica	Tocos
TOV	Numérica	Erros de ataque(turnovers)
PF	Numérica	Faltas pessoais
PTS	Numérica	Pontos na temporada
Awards	Categórica	Prêmios da temporada

Fonte: elaborado pelo autor (2025)

8.2 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2019-2020

A análise exploratória da temporada 2019-20 busca descrever o perfil estatístico dos jogadores da NBA com base nas principais métricas de desempenho individual. Por meio de estatísticas descritivas, examinam-se aspectos como participação em jogos, volume ofensivo, eficiência nos arremessos, contribuições defensivas e volume de minutos em quadra. A interpretação desses dados permite destacar tendências, assimetrias e a presença de outliers, contribuindo para uma compreensão preliminar da base que servirá de insumo para as etapas posteriores do estudo.

8.2.1 Estatísticas das variáveis da temporada 2019-2020

A tabela a seguir apresenta um resumo estatístico das variáveis numéricas da temporada 2019-20 da NBA, incluindo medidas como média, desvio padrão, valores mínimo e máximo, além dos quartis. Esses indicadores fornecem uma visão inicial da distribuição dos dados, permitindo identificar padrões gerais de desempenho dos jogadores, variações importantes e possíveis valores extremos.

Tabela 1 - Estatísticas das variáveis 19-20

VARIÁVEL	MÉDIA	D.PADRÃO	MÍN	Q1	MEDIANA	Q3	MAX
G	38,2165	22,9444	1	15,5	41	59	74
GS	17,5376	23,0221	0	0	4	29,5	73
MP	864,1582	707,9150	1	221	734	1435,5	2556

FG	144,8110	146,1559	0	25	95	226,5	685
FGA	315,9201	311,9820	0	59	218	492,5	1514
FG.	0,4429	0,1114	0	0,3992	0,439	0,486	1
X3P	43,7987	51,8433	0	3	24	65,5	299
X3PA	122,7096	136,5641	0	13,5	76	188	843
X3P.	0,3135	0,1334	0	0,271	0,339	0,378	1
X2P	101,0122	112,2199	0	17	60	150,5	596
X2PA	193,2104	210,0888	0	33,5	120	287	1068
X2P.	0,5074	0,1247	0	0,4592	0,509	0,5687	1
eFG.	0,5067	0,1100	0	0,475	0,518	0,557	1
FT	63,0568	83,0957	0	8,5	34	80	692
FTA	81,7081	103,2678	0	13	46	105,5	800
FT.	0,7438	0,1505	0	0,667	0,771	0,836	1
ORB	35,9708	42,3226	0	8	23	50	258
DRB	124,3440	122,2997	0	27,5	88	189	716
TRB	160,3149	159,2342	0	38	110	239,5	916
AST	85,7741	106,2473	0	14	50	112	684
STL	27,5176	25,9977	0	5,5	20	43,5	125
BLK	17,6589	23,1953	0	3	10	23	196
TOV	49,5560	51,8405	0	10	33	72,5	308
PF	74,7281	58,6324	0	21	66	117,5	278
PTS	396,4777	407,8560	0	68	257	609,5	2335

Fonte: Elaborado pelo autor (2025)

Observando a Tabela 1, percebemos diversas informações importantes que descrevem as variáveis presentes na base de dados dos jogadores da NBA da temporada de 2019-2020, são elas:

G (Jogos jogados): A média de 38 partidas indica que muitos atletas participaram de menos da metade da temporada regular (82 jogos), com mediana de 41 e mínimo de apenas 1 jogo. A distribuição é razoavelmente simétrica, sem outliers evidentes.

GS (Jogos como titular): A mediana de apenas 4 jogos revela que a maioria dos atletas não foi titular regularmente, sugerindo alto número de jogadores reservas ou que atuaram em poucos jogos.

MP (Minutos jogados): Alta dispersão, com média de 864 minutos e desvio padrão acima de 700, indica que alguns atletas foram muito pouco utilizados, enquanto outros ultrapassaram 2500 minutos na temporada.

FG / FGA / FG% (Cestas de quadra, tentativas e percentual): A média de 144 cestas com 316 tentativas gera um percentual médio de acerto de 44%, dentro do esperado para uma liga competitiva. O valor máximo de FG chega a 685, com 1514 tentativas, apontando outliers ofensivos intensos.

X3P / X3PA / X3P% (Cestas de 3 pontos): Alta assimetria à direita. Embora a média de cestas de 3 pontos seja 43, jogadores atingem até 299 cestas, revelando especialistas na linha de 3 pontos.

X2P / X2PA / X2P% (Cestas de 2 pontos): Com média de 101 cestas e 193 tentativas, os jogadores mantêm percentual de acerto razoável (50,7%). Alguns atletas superam 1000 tentativas na temporada.

eFG% (Field Goal ajustado): Média próxima de 50%, com dispersão moderada. Essa métrica combina 2P e 3P ponderando o peso maior dos arremessos de 3.

FT / FTA / FT%: Média de 63 lances convertidos em 81 tentativas (FT% $\approx$ 74%), condizente com o padrão da liga. O máximo de 800 tentativas mostra grande protagonismo ofensivo de certos atletas. O desvio padrão elevado evidencia grande variação no volume de lances livres.

ORB / DRB / TRB: Os rebotes totais (TRB) têm média de 160, com máximo de 916. O desvio padrão elevado ($\approx$ 159) confirma grande dispersão

entre especialistas em rebotes e jogadores perimetrais. Rebotes ofensivos são bem menores em média (36), como esperado.

AST: Média de 85 assistências, com máximos superiores a 600 indicam jogadores com forte papel na armação ofensiva.

STL / BLK: Ambas com distribuição assimétrica à direita. A média de 27 roubos e 17 tocos é modesta, mas os máximos de 125 (STL) e 196 (BLK) sugerem atletas defensivamente dominantes.

TOV (Turnovers): Média de 49 erros por temporada. Alguns jogadores superam 300, o que é típico de atletas com alta responsabilidade de criação ofensiva.

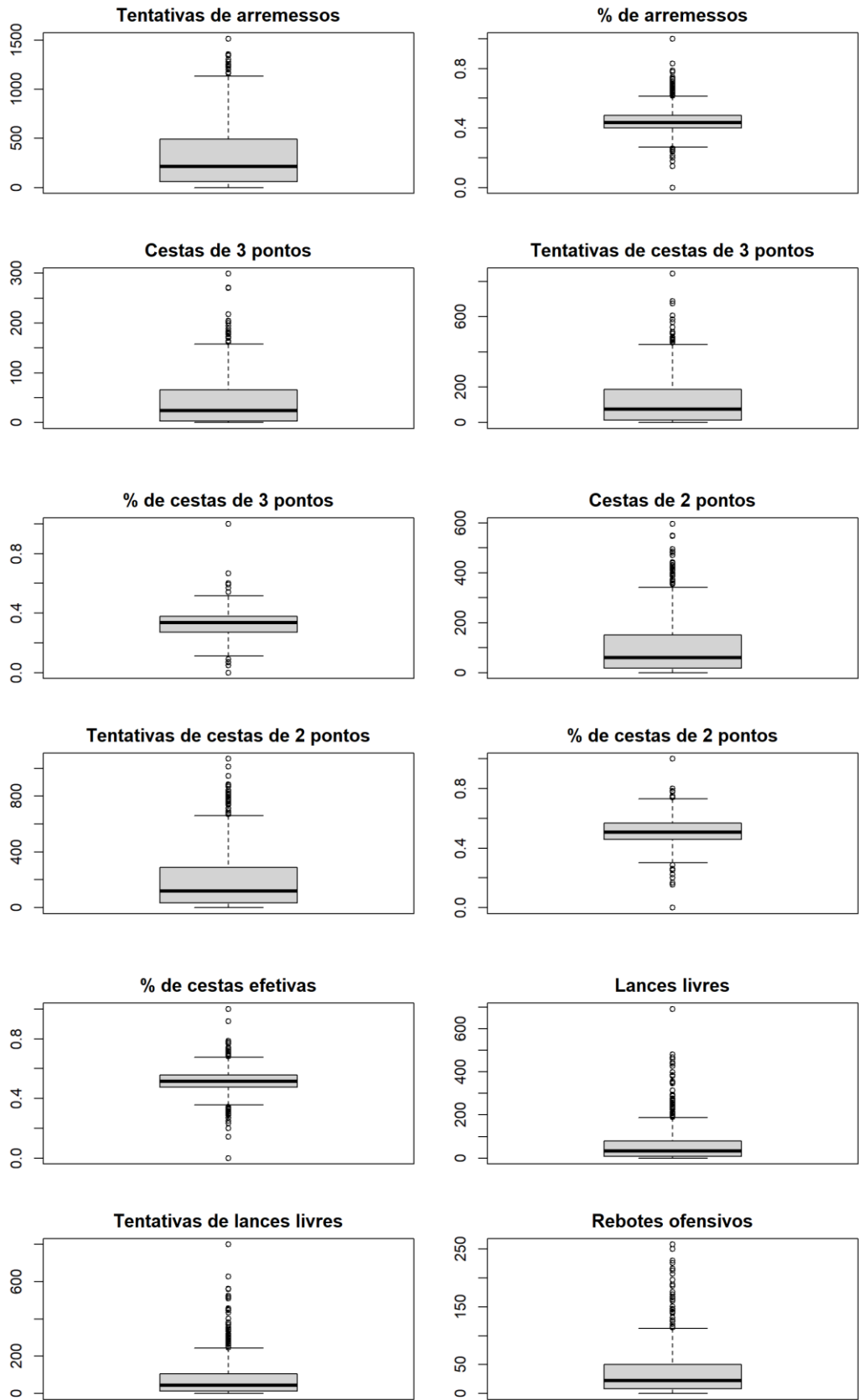
PF (Faltas pessoais): A média de 75 e desvio padrão de 58 revelam variação considerável. Valores máximos de 278 indicam jogadores com grande volume defensivo ou excesso de agressividade.

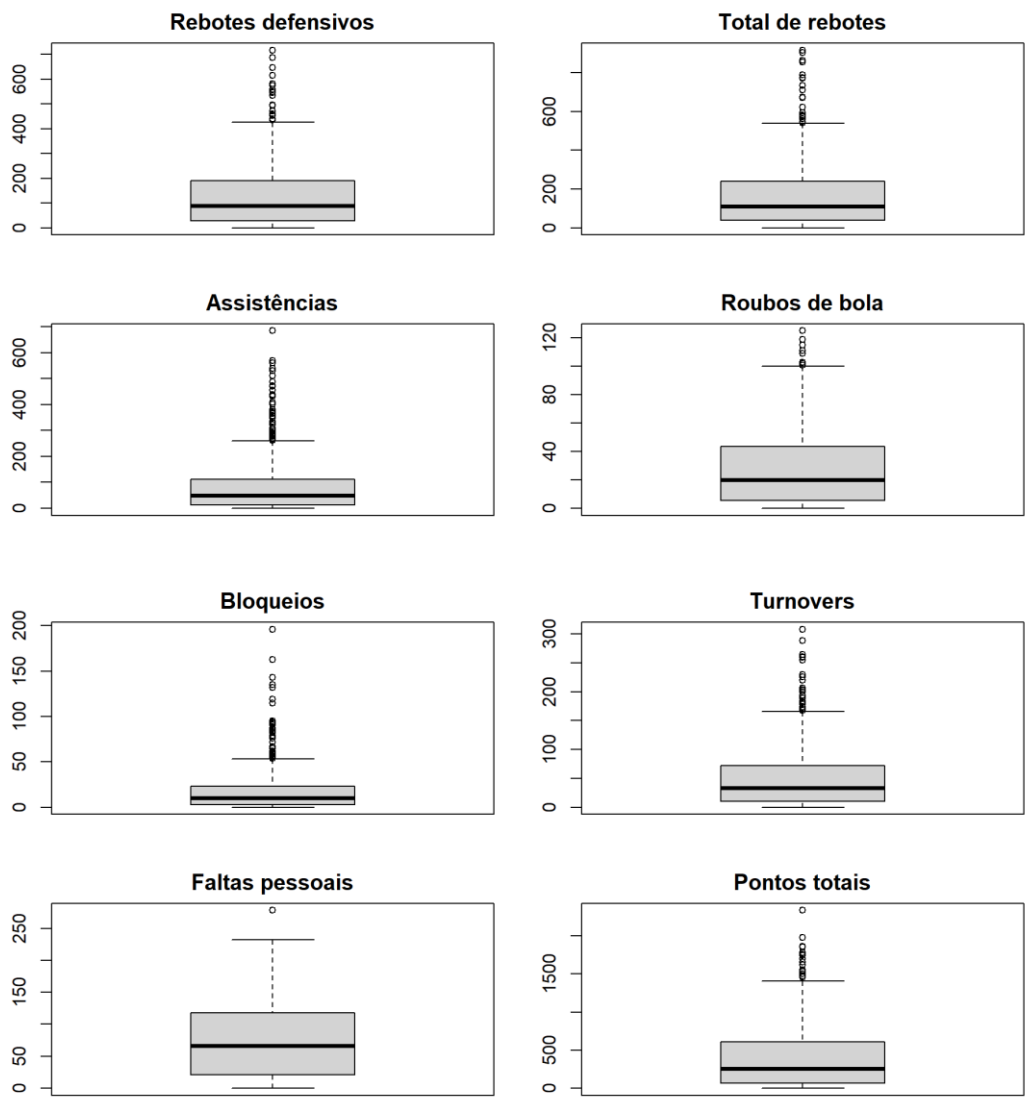
PTS: Com média de 396 e máximo de 2335, é notável a presença de outliers — poucos jogadores concentram alto volume ofensivo, enquanto grande parte atua de forma secundária.

8.2.2 Análise gráfica das variáveis da temporada 2019-2020

Nesta subseção, são apresentadas as análises gráficas das variáveis referentes à temporada 2019-20 da NBA, com o objetivo de explorar visualmente os padrões de desempenho dos jogadores. Foram utilizados boxplots, histogramas e correlogramas para examinar a distribuição, a variabilidade e as correlações entre as métricas analisadas. Esses gráficos possibilitam a identificação de assimetrias, valores discrepantes e relações lineares entre os indicadores, contribuindo para uma melhor compreensão da estrutura dos dados.

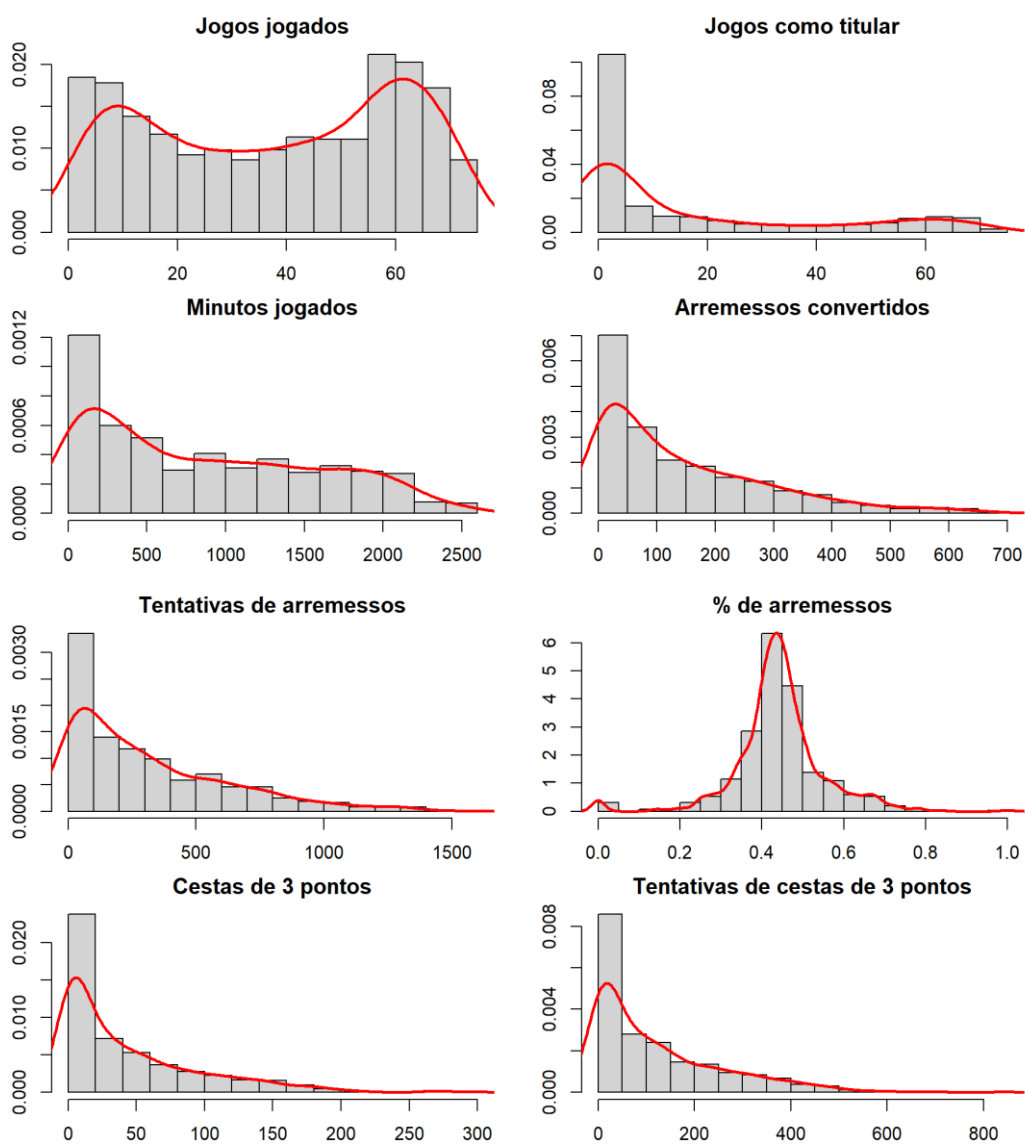
Figura 1: Boxplot's das variáveis 19-20

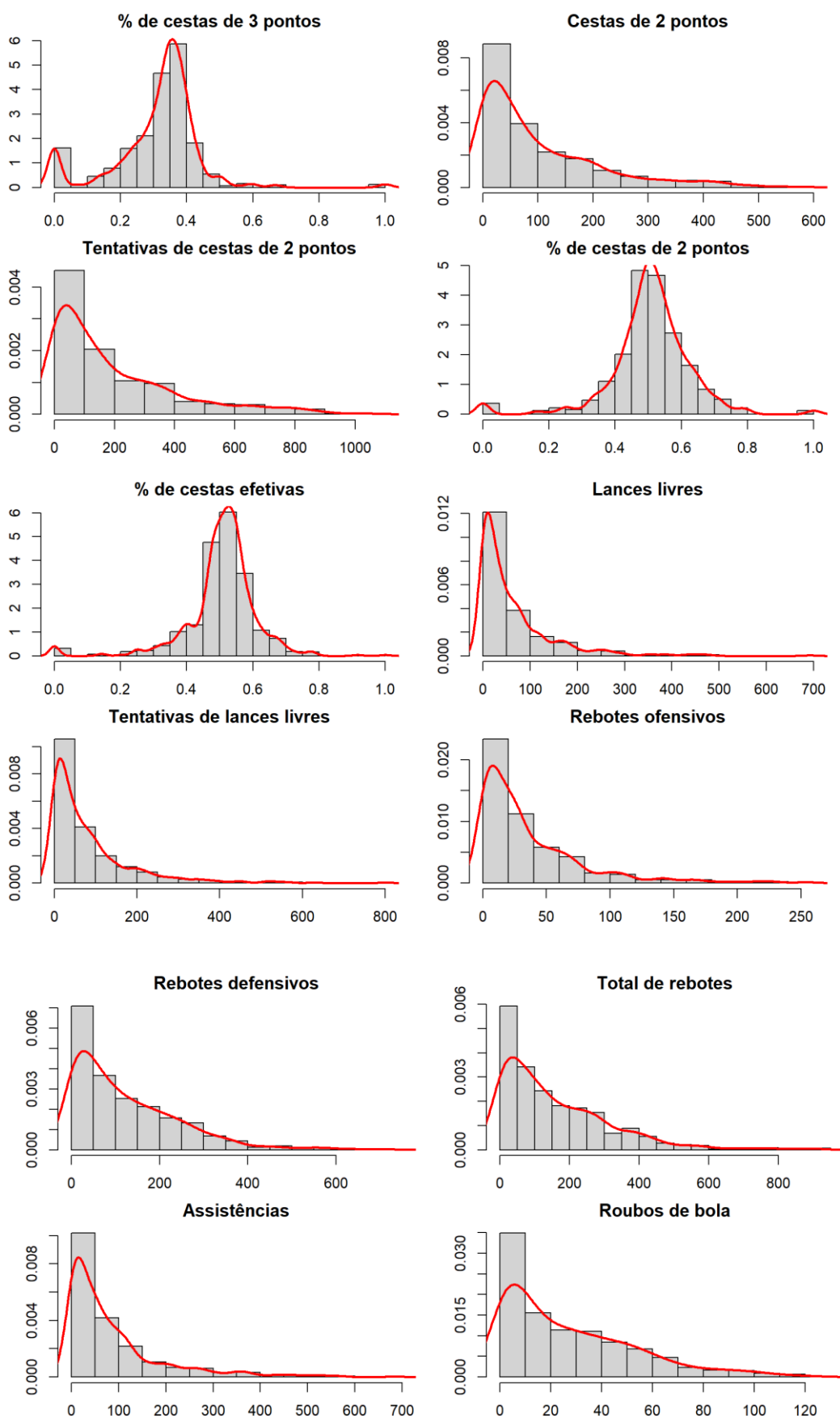


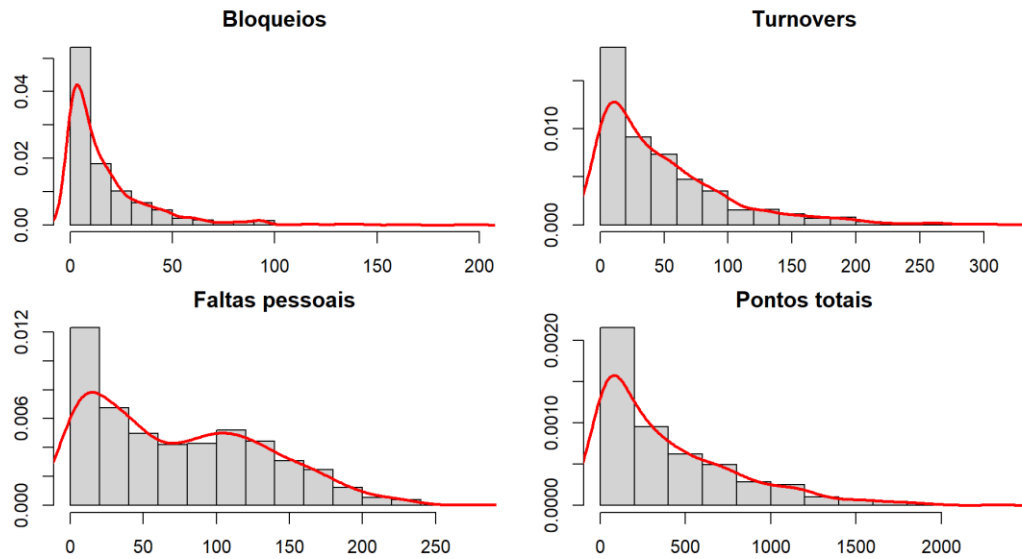


Fonte: elaborado pelo autor (2025)

Figura 2: Histogramas das variáveis 19-20

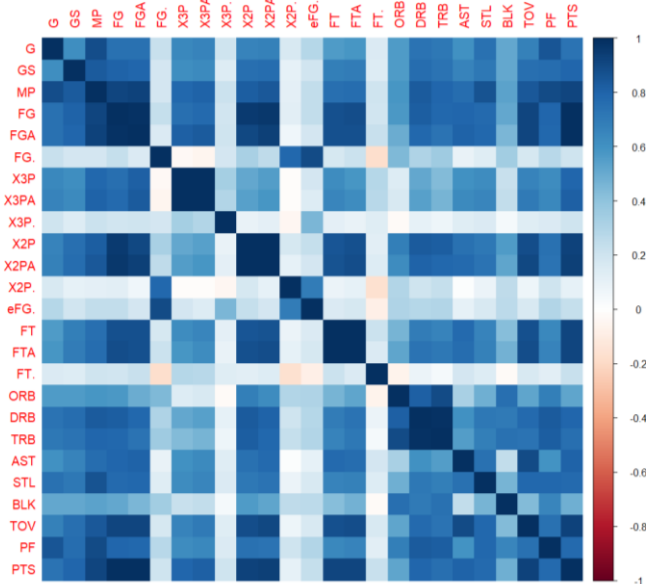






Fonte: elaborado pelo autor (2025)

Figura 3: Correlograma das variáveis 19-20



Fonte: elaborado pelo autor (2025)

Através da análise dos gráficos apresentados, observa-se que as variáveis de participação em jogos revelam estruturas hierárquicas nas equipes: Jogos Jogados mostram distribuição bimodal (picos em 10 e 70 jogos), enquanto Jogos como Titular apresentam forte assimetria à direita, com poucos jogadores concentrando titularidades.

As variáveis de volume ofensivo (Arremessos Convertidos, Tentativas) exibem distribuições fortemente assimétricas à direita, com especialização ainda mais acentuada nos arremessos de três pontos. Em contraste, as métricas de eficiência (Porcentagem de Arremessos, Cestas de 2 Pontos, Cestas Efetivas) apresentam distribuições aproximadamente normais (médias entre 45-55%), sugerindo homogeneidade na eficiência independente da posição.

Rebotes (Ofensivos, Defensivos, Totais) mostram assimetria à direita, com rebotes defensivos mais distribuídos entre posições que os ofensivos. Assistências e Turnovers seguem padrão similar, concentrados em jogadores com maior posse de bola. Estatísticas defensivas revelam que Bloqueios são altamente especializados, enquanto Roubo de Bola apresentam distribuição mais equilibrada entre posições.

A matriz de correlação evidencia grupos de variáveis fortemente relacionadas: arremessos de 3 pontos formam um grupo coeso, assim como pontos, tentativas e arremessos convertidos. Estas correlações refletem a especialização dos jogadores em funções específicas.

Em síntese, a análise revela diversidade de perfis na NBA, com clara especialização em determinados fundamentos, distribuições assimétricas para variáveis de volume e mais simétricas para eficiência, fornecendo insights valiosos para a criação de rankings que devem considerar múltiplas métricas para capturar adequadamente o valor de cada jogador.

8.2.3 Análise de componentes principais(PCA) 2019-2020

Inicialmente, foram importadas as estatísticas da temporada 2019–20, contemplando os principais indicadores de desempenho dos jogadores: jogos disputados (G), jogos como titular (GS), minutos (MP), aproveitamentos de arremessos de campo (FG.), bolas de 3 pontos (X3P, X3P.), arremessos de 2 pontos (X2P, X2P.), eficiência ofensiva ajustada (eFG.), lances livres (FT, FT.), rebotes ofensivos e defensivos (ORB, DRB, TRB), assistências (AST), roubos de bola (STL), tocos (BLK), desperdícios (TOV), faltas (PF) e pontos totais (PTS).

Como critério de corte, foram mantidos apenas os jogadores que disputaram pelo menos 20 jogos e acumularam no mínimo 300 minutos em quadra na temporada. Esse filtro busca garantir que o ranking não seja distorcido por desempenhos pontuais de atletas com baixa participação. Após a coerção das variáveis numéricas e a aplicação desses critérios, obteve-se um conjunto de observações suficientemente robusto para todas as cinco posições (armadores, alas-armadores, alas, alas-pivôs e pivôs).

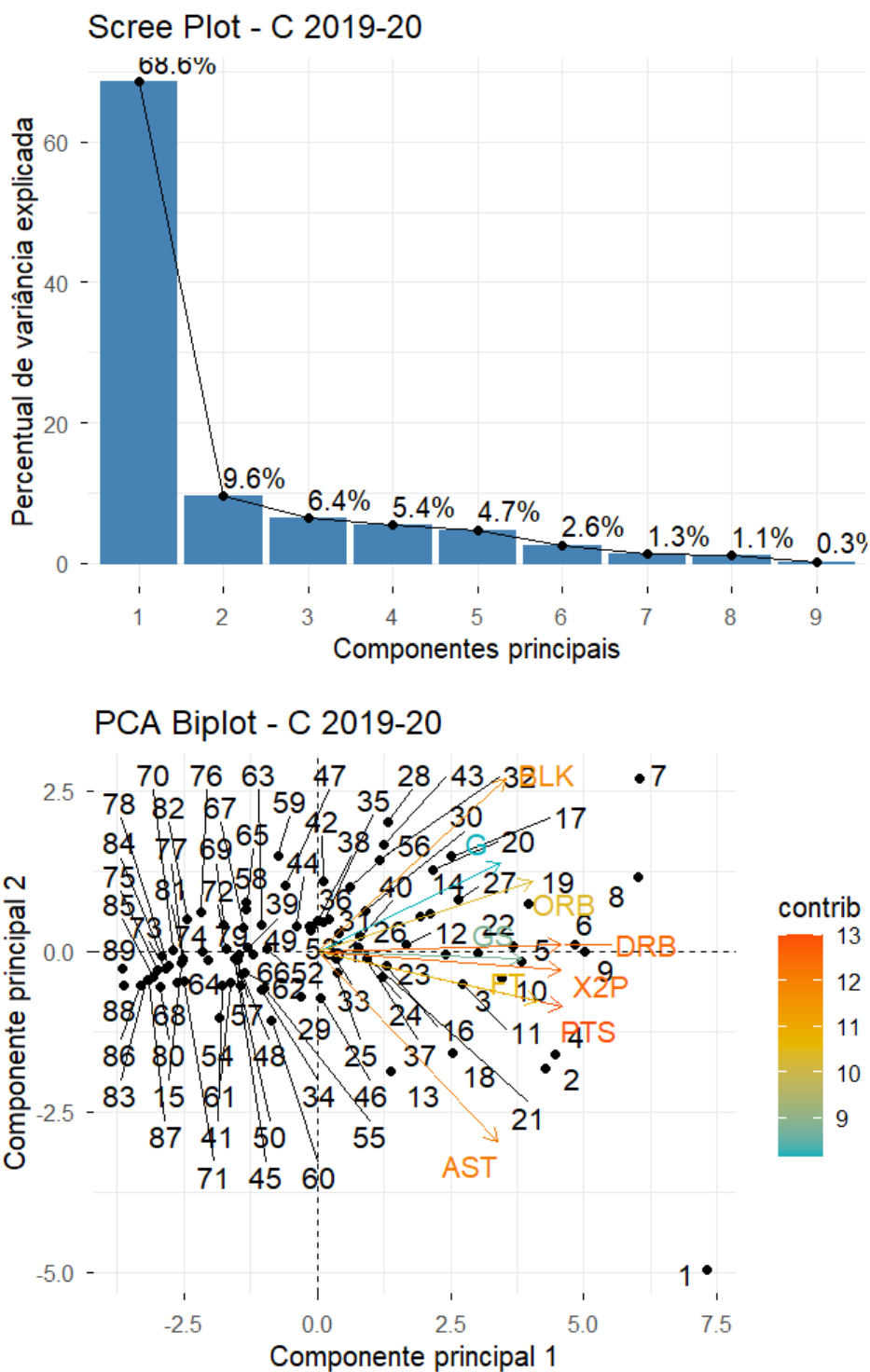
Antes da construção dos índices, foi avaliada a adequação da matriz de correlações para cada conjunto de variáveis (por posição) por meio do índice de adequação amostral de Kaiser-Meyer-Olkin (KMO) e do teste de esfericidade de Bartlett. O KMO compara as magnitudes das correlações simples com as magnitudes das correlações parciais, verificando se as variáveis compartilham variância comum suficiente para justificar a extração de fatores latentes. Valores globais acima de 0,70 são considerados adequados para análise fatorial, pois indicam que a estrutura de correlações não é excessivamente influenciada por variâncias únicas das variáveis (Hair et al., 2009). No presente estudo, todos os conjuntos posicionais apresentaram KMO global superior a esse limiar, confirmando a viabilidade da modelagem por fatores. Complementarmente, o teste de Bartlett examina a hipótese nula de que a matriz de correlações populacional é uma matriz identidade (ou seja, que as variáveis são ortogonais). A rejeição dessa hipótese ($p < 0,05$) indica que existem correlações significativas entre pelo menos alguns pares de variáveis, o que é pressuposto necessário para que a AFE possa identificar estruturas latentes subjacentes (Johnson; Wichern, 2007). Em todas as posições analisadas, o teste de Bartlett foi significativo, corroborando a presença de covariância sistemática entre as métricas de desempenho e legitimando a extração de fatores comuns.

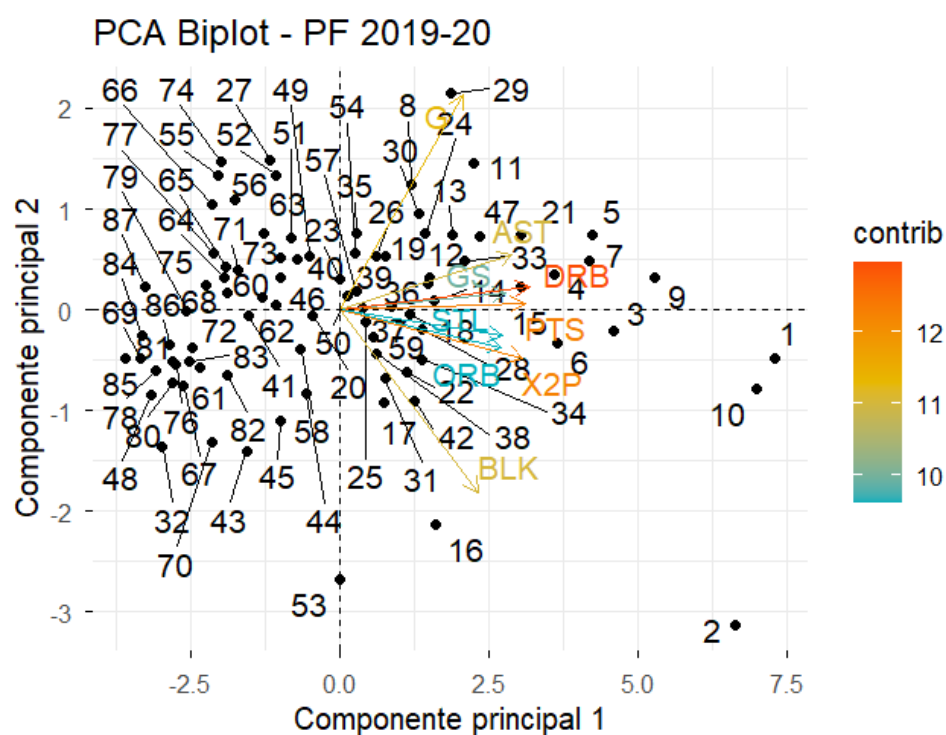
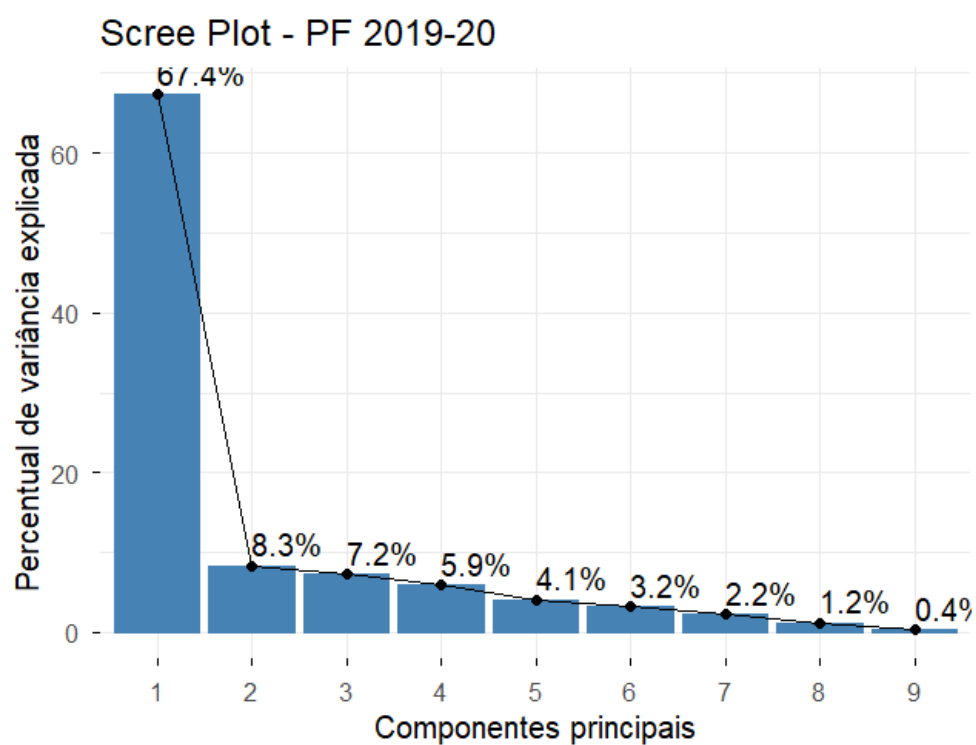
Além do KMO global, foram avaliados os índices MSA individuais (MSA_i), que representam a contribuição de cada variável para a adequação geral da matriz. Variáveis com MSA_i abaixo de 0,50 foram removidas do conjunto inicial, de modo a fortalecer a estrutura fatorial e evitar variáveis que não se integram bem ao fator comum, garantindo maior robustez à solução fatorial final.

Com a matriz depurada, procedeu-se à aplicação da Análise de Componentes Principais (PCA) como ferramenta diagnóstica. Para cada posição, os dados foram previamente padronizados (escores z). Em termos gerais, observou-se que:

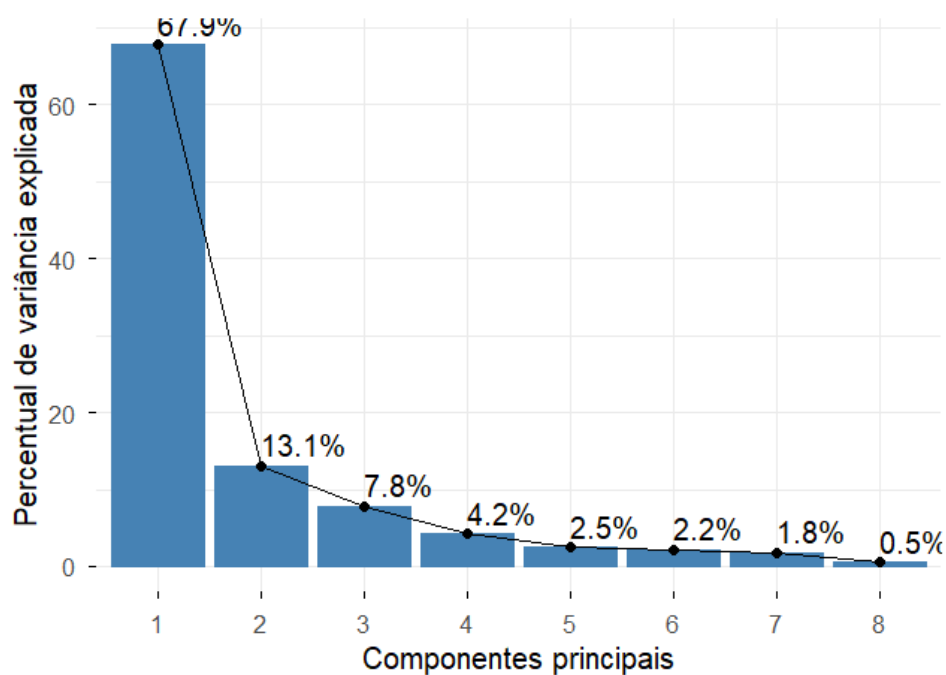
- Os três primeiros componentes principais explicaram uma proporção elevada da variância total (em torno de 75–85%, variando ligeiramente entre posições);
- O scree plot apresentou um padrão típico de “joelho”, com queda acentuada dos autovalores após os primeiros componentes, especialmente o primeiro, sugerindo a existência de um ou dois eixos latentes dominantes.

Figura 4 - Scree plot's e Biplot's por posição 2019-2020

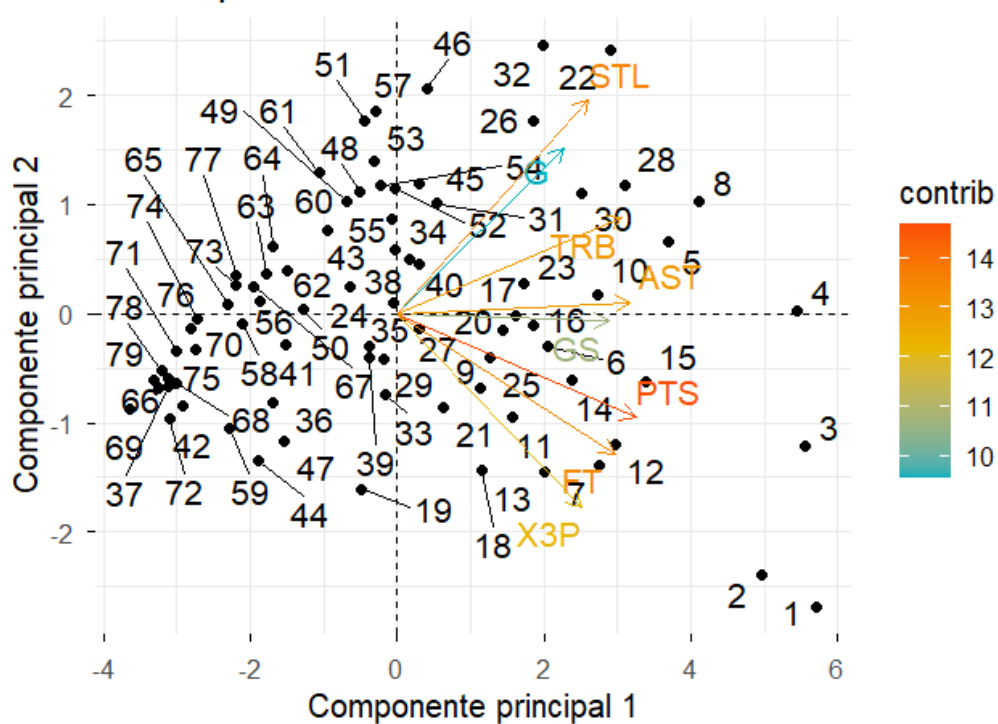




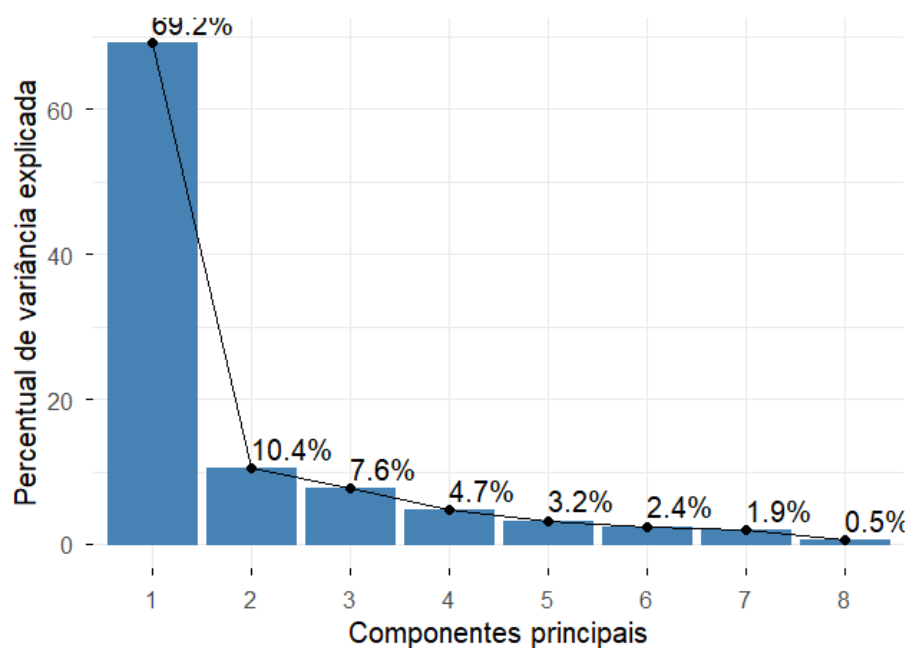
Scree Plot - PG 2019-20



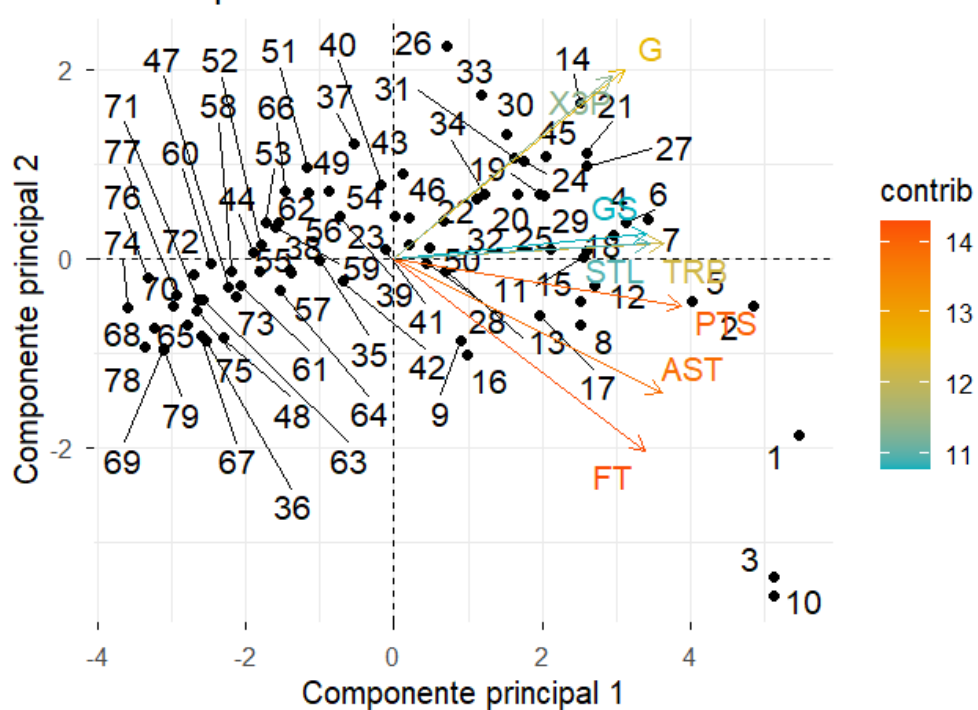
PCA Biplot - PG 2019-20

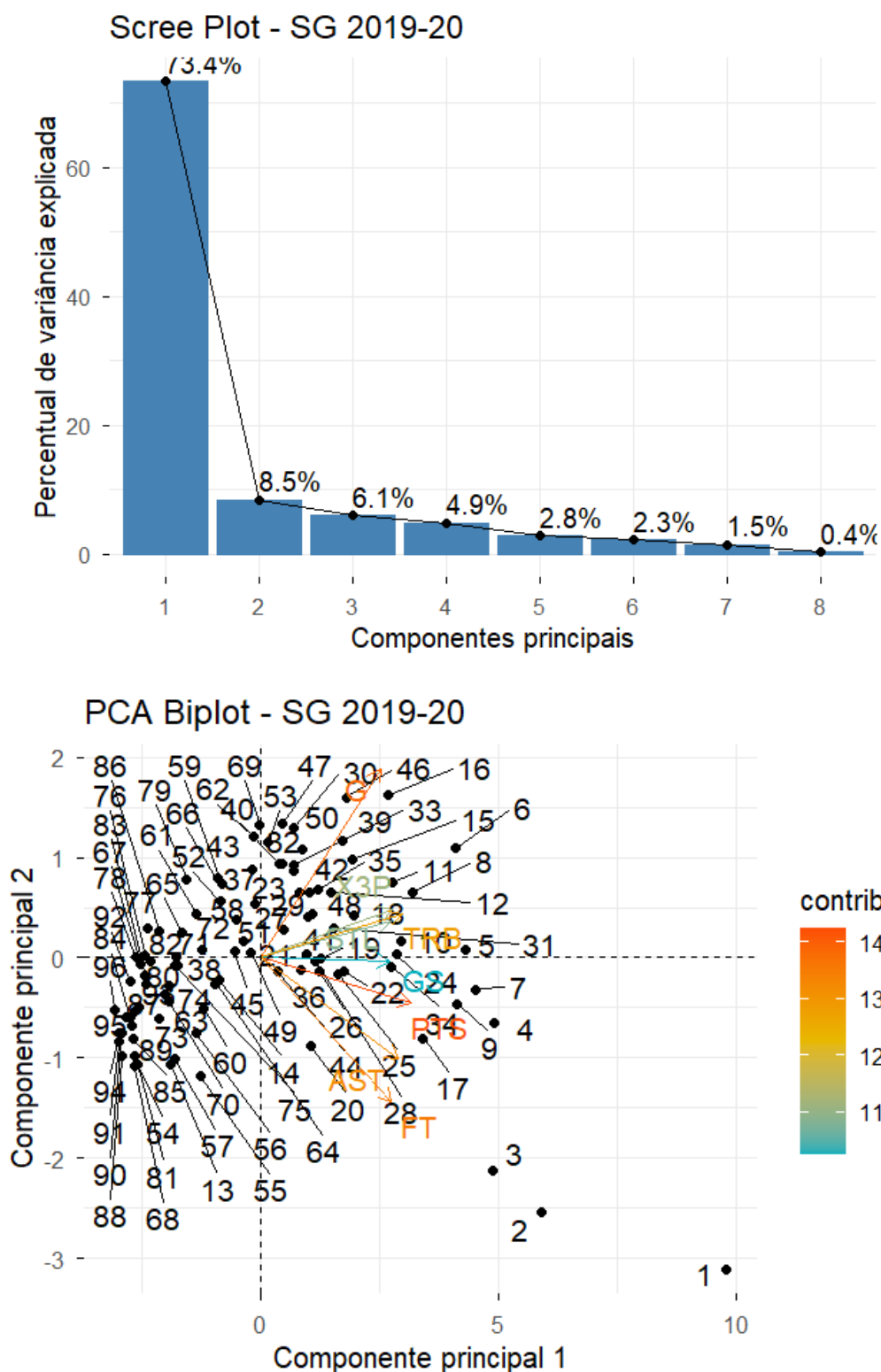


Scree Plot - SF 2019-20



PCA Biplot - SF 2019-20





Fonte: elaborado pelo autor.

Do ponto de vista interpretativo, os primeiros componentes tendem a refletir:

- Para jogadores de perímetro (PG, SG, SF), um eixo principal associado a potência ofensiva e criação de jogadas, com altas cargas em PTS, AST, X3P, FT e, em menor grau, TRB;

- Para jogadores de garrafão (PF, C), um eixo associado a domínio interno, com altas cargas em PTS, X2P, ORB, DRB e BLK.

Em todos os casos, a PCA confirmou a presença de uma estrutura de correlação consistente com a intuição sobre o papel estatístico de cada posição.

8.2.4 Clusterização por K-médias 2019-2020

Como etapa subsequente, os escores dos componentes principais (em geral, os três primeiros PCs) foram utilizados como entrada para o algoritmo de clusterização k-means, com o objetivo de identificar grupos relativamente homogêneos de jogadores dentro de cada posição, considerando suas características multivariadas.

O número de clusters (k) foi avaliado no intervalo de 2 a 8, sendo escolhido com base no coeficiente médio de silhueta. Para cada valor de k, calculou-se:

$$silhueta\acute{m}edia(k) = \frac{1}{n} \sum_{i=1}^n s_i \quad (6)$$

onde s_i é o índice de silhueta do jogador i , que mede o quão bem ele está alocado ao seu cluster em comparação a outros clusters. O valor de k que maximizou a silhueta média foi adotado como k ótimo em cada posição.

Os resultados indicaram, de forma geral:

- k ótimo entre 2 e 4 clusters, dependendo da posição;
- Valores de silhueta média compatíveis com uma separação razoável entre grupos (tipicamente acima de 0,30).

Do ponto de vista interpretativo, os clusters tendem a se organizar em perfis como:

- Grupo de jogadores com estatísticas de elite (altos escores no primeiro componente, combinando pontuação, criação e/ou domínio no garrafão);
- Grupo intermediário, com bons indicadores em algumas dimensões, mas não em todas;
- Grupo de jogadores com estatísticas mais modestas, muitas vezes associados a menor tempo de quadra ou papel mais especializado.

8.2.5 Análise fatorial exploratória (AFE) e construção dos índices por posição 2019-2020

Com os conjuntos de variáveis já avaliados por KMO/Bartlett e explorados via PCA, aplicou-se a Análise Fatorial Exploratória (AFE) para cada posição, com as seguintes características:

- Método de extração: máxima verossimilhança ($fm = "ml"$);
- Rotação: oblimin (rotação oblíqua), permitindo correlação entre fatores;
- Número de fatores: definido por análise paralela ($fa.parallel$), garantindo que apenas fatores com autovalores empíricos superiores aos de dados aleatórios fossem retidos.

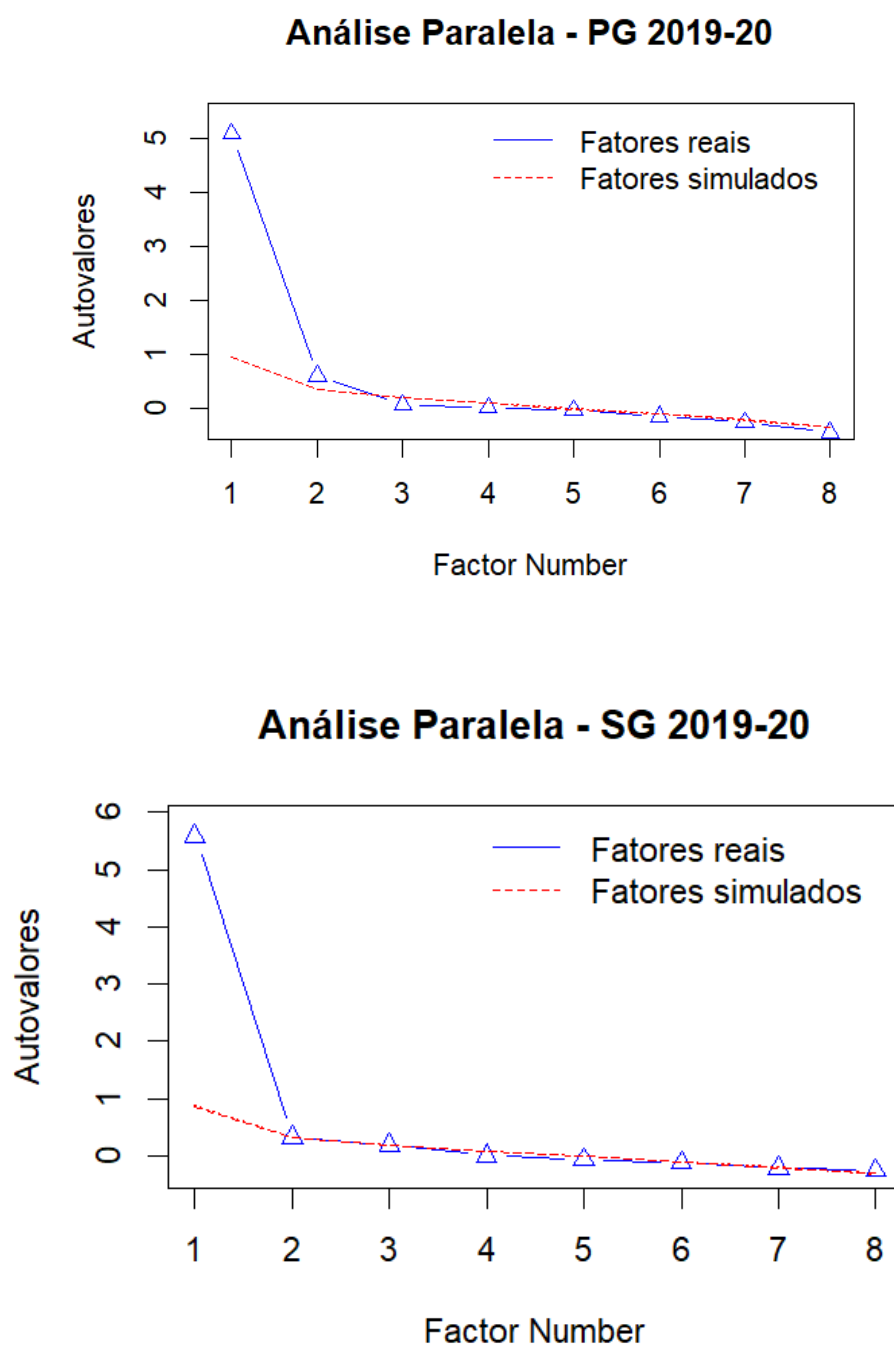
Como procedimento de refinamento, foram adotados critérios clássicos:

- Filtro por KMO individual ($MSAi$) – já aplicado na etapa anterior, removendo variáveis com $MSAi < 0,50$;
- Filtro por comunalidades (h^2) – após a primeira AFE, variáveis com $h^2 < 0,30$ foram candidatas à remoção, desde que o conjunto resultante ainda mantivesse pelo menos três variáveis;
- Reestimação da AFE com o conjunto reduzido, quando necessário, e reavaliação das comunalidades e das cargas fatoriais.

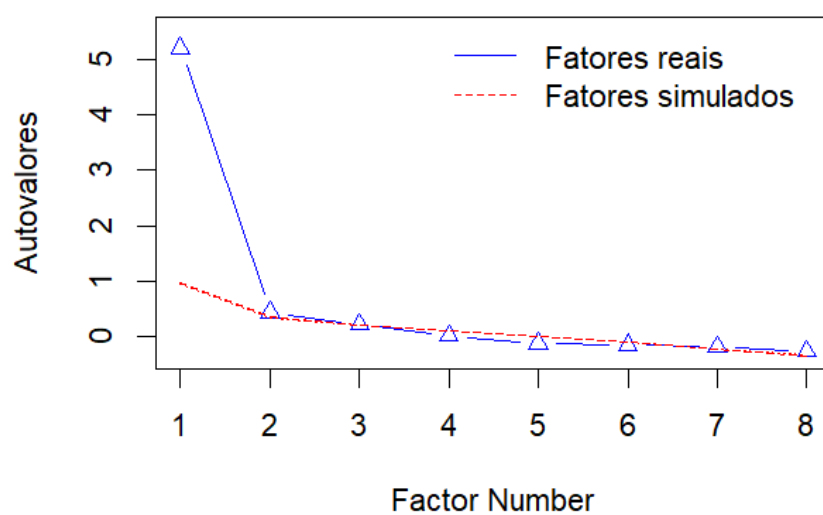
De maneira geral, para as cinco posições:

- A análise paralela sugeriu um número pequeno de fatores (entre 1 e 3);
- O primeiro fator foi claramente dominante, associado às variáveis de maior relevância para o desempenho da posição (por exemplo, PTS, AST, X3P e FT para jogadores de perímetro; PTS, X2P, ORB, DRB e BLK para jogadores de garrafão);
- As comunalidades finais ficaram, em grande parte, acima de 0,50, indicando boa explicação da variância das variáveis originais pelos fatores.

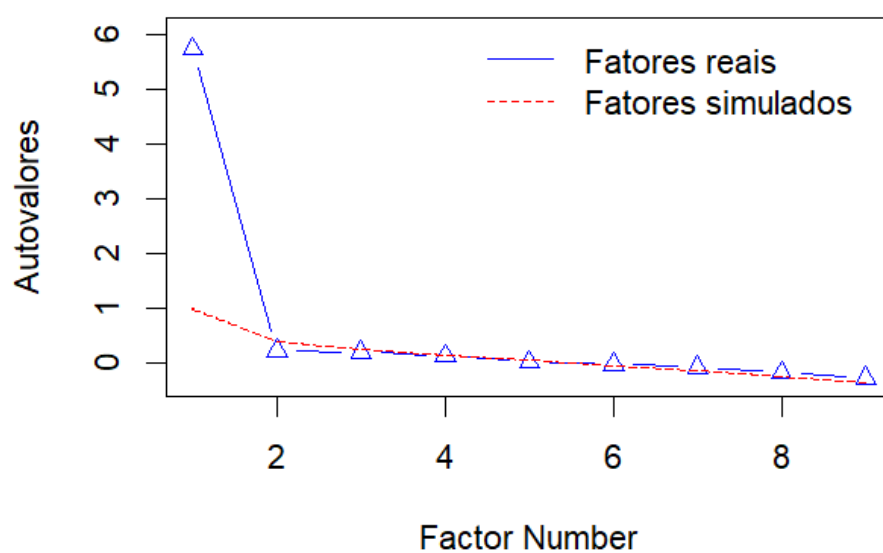
Figura 5 - Análise paralela por posição 2019-2020

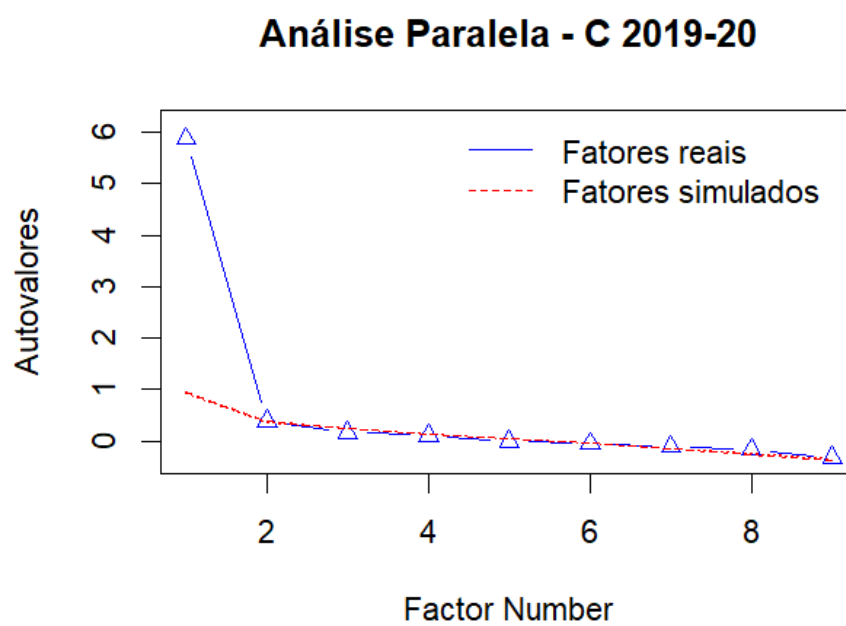


Análise Paralela - SF 2019-20



Análise Paralela - PF 2019-20





Fonte: elaborado pelo autor.

Figura 6 - Diagrama de fatores por posição 2019-2020

Diagrama de Fatores - C 2019-20

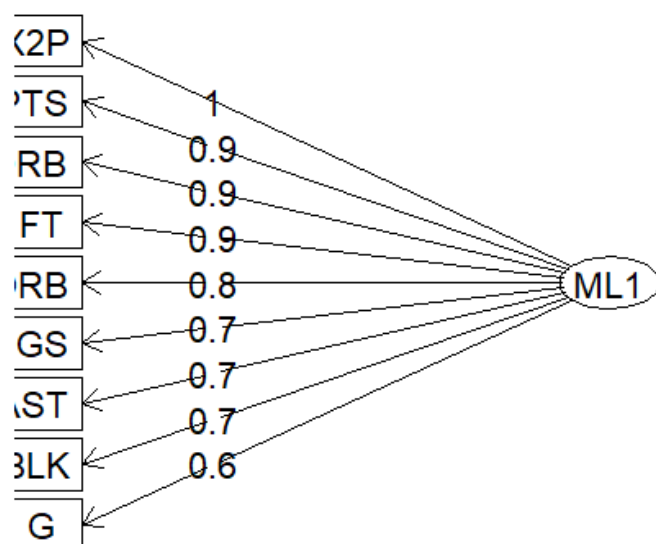


Diagrama de Fatores - PF 2019-20

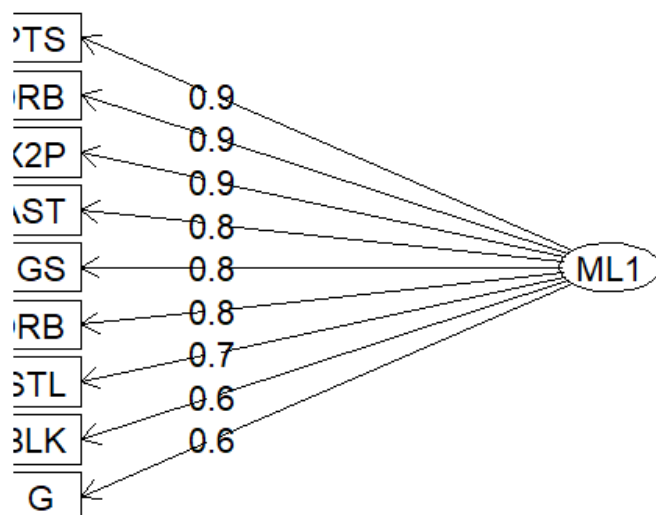


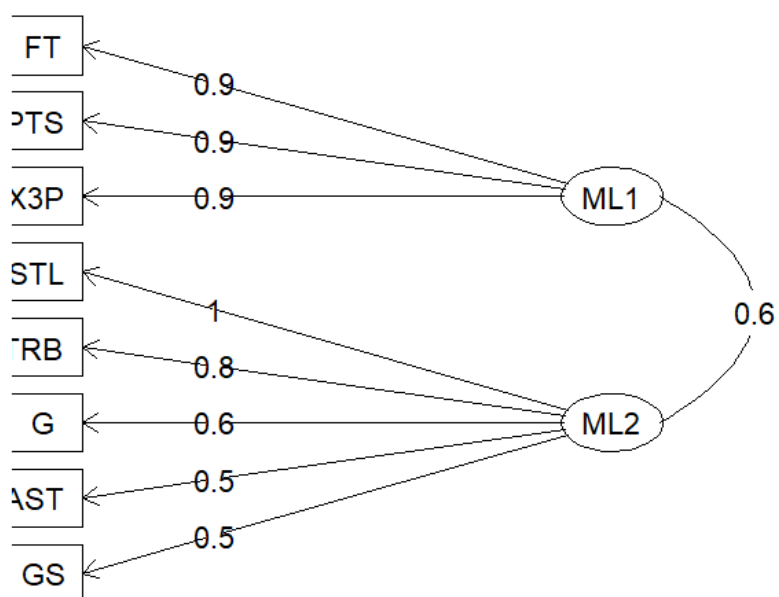
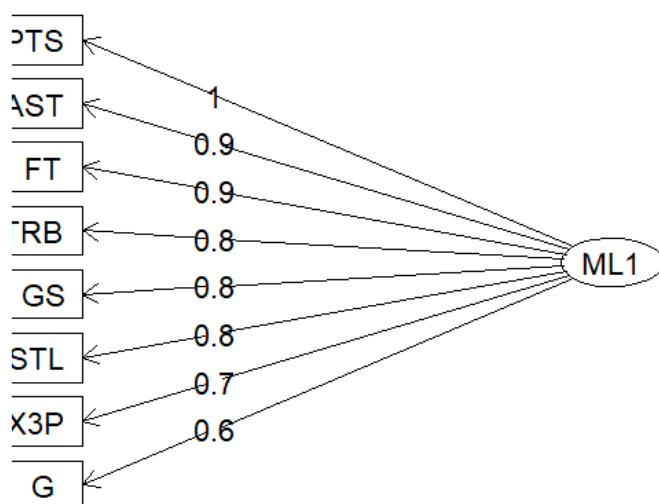
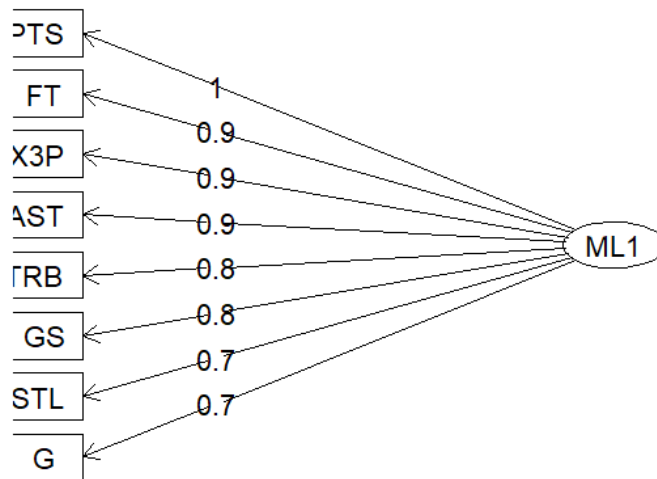
Diagrama de Fatores - PG 2019-20**Diagrama de Fatores - SF 2019-20**

Diagrama de Fatores - SG 2019-20



Fonte: elaborado pelo autor.

A partir dos resultados da AFE, foram obtidos escores fatoriais para cada fator e atleta. Para compor um índice único de desempenho por posição, os escores fatoriais foram ponderados pela proporção de variância explicada de cada fator. Se f_j representa o escore no fator j e w_j a fração de variância explicada pelo fator j , o índice bruto é dado por:

$$Score\ Bruto = \sum_j w_j f_j \quad (7)$$

Em seguida, cada índice foi normalizado para a escala 0–100 por meio da transformação:

$$Score\ Norm = \frac{Score\ Bruto - \min(Score\ Bruto)}{\max(Score\ Bruto) - \min(Score\ Bruto)} \times 100 \quad (8)$$

Esse procedimento garante transparência no cálculo do índice na composição do ranking final.

8.2.6 Rankings por posição 2019-2020

Com os escores normalizados produzidos pela AFE, foram construídos os rankings por posição para a temporada 2019–20. Para cada uma das cinco posições (PG, SG, SF, PF e C), os jogadores foram ordenados em ordem decrescente de `score_norm`, e foram apresentados os Top 10 em tabelas específicas.

De forma geral, observou-se que:

- Nas posições de perímetro, os primeiros colocados tendem a reunir alto volume de pontuação por jogo, grande número de assistências e boa eficiência em arremessos de três e lances livres;
- Nas posições de garrafão, os atletas mais bem ranqueados apresentam elevados índices de rebotes (ofensivos e defensivos), alta produção em pontos próximos ao aro (X2P) e boa taxa de tocos (BLK).

Tabela 2 - Top 10 C 2019-20

Rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
1	Nikola Jokic	C	73	73	2336	19.95	9.74	7.01	100	1
2	Hassan Whiteside	C	67	61	2008	15.52	13.51	1.16	87.21	1
3	Rudy Gobert	C	68	68	2333	15.09	13.47	1.49	86.79	1
4	Andre Drummond	C	57	56	1879	17.72	15.16	2.67	82.05	1
5	Jonas Valanciunas	C	70	70	1845	14.91	11.26	1.87	79.61	1
6	Nikola Vucevic	C	62	62	1998	19.60	10.85	3.58	78.69	1
7	Joel Embiid	C	51	51	1506	22.98	11.63	2.98	77.20	1
8	Montrezl Harrell	C	63	2	1749	18.62	7.06	1.70	75.93	1
9	Andre Drummond	C	49	48	1654	17.76	15.82	2.82	70.96	1
10	Kristaps Porzingis	C	57	57	1814	20.42	9.47	1.79	67.52	1

Fonte: elaborado pelo autor

Tabela 3- Top 10 PF 2019-20

Rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
1	Giannis Antetokounmpo	PF	63	63	1917	29.48	13.59	5.62	100	1
2	Bam Adebayo	PF	72	72	2417	15.92	10.21	5.11	81.37	1
3	Anthony Davis	PF	62	62	2131	26.10	9.31	3.23	80.65	1
4	Domantas Sabonis	PF	62	62	2159	18.50	12.44	5.00	76.22	1
5	Tobias Harris	PF	72	72	2469	19.60	6.88	3.15	68.68	1

6	Jayson Tatum	PF	66	66	2265	23.44	6.97	3.03	68.51	1
7	Julius Randle	PF	64	64	2080	19.50	9.72	3.09	67.60	1
8	Pascal Siakam	PF	60	60	2110	22.85	7.32	3.45	62.82	1
9	Brandon Ingram	PF	62	62	2104	23.82	6.12	4.18	62.61	1
10	Aaron Gordon	PF	62	62	2017	14.42	7.66	3.68	52.35	1

Fonte: elaborado pelo autor.

Tabela 4 – Top 10 PG 2019-20

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
1	LeBron James	PG	67	67	2316	25.34	7.84	10.21	100.00	1
2	Luka Doncic	PG	61	61	2047	28.84	9.39	8.82	98.41	1
3	Damian Lillard	PG	66	66	2474	29.97	4.30	8.03	97.17	1
4	Russell Westbrook	PG	57	57	2049	27.25	7.91	7.04	90.08	1
5	Trae Young	PG	60	60	2120	29.63	4.25	9.33	88.16	1
6	Chris Paul	PG	70	70	2208	17.60	4.99	6.74	79.26	1
7	Ben Simmons	PG	57	57	2017	16.44	7.79	7.98	72.26	1
8	Collin Sexton	PG	65	65	2143	20.75	3.14	2.95	66.73	1
9	Ja Morant	PG	67	67	2074	17.81	3.87	7.28	66.38	1
10	Kyle Lowry	PG	58	58	2098	19.41	5.03	7.47	66.33	1

Fonte: elaborado pelo autor.

Tabela 5 – Top 10 SF 2019-20

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
1	DeMar DeRozan	SF	68	68	2316	22.12	5.51	5.59	100.00	1
2	Kawhi Leonard	SF	57	57	1848	27.07	7.05	4.91	99.80	1
3	Zach LaVine	SF	60	60	2085	25.50	4.82	4.23	93.09	1
4	Jimmy Butler	SF	58	58	1959	19.95	6.66	6.03	90.85	1

5	Khris Middleton	SF	62	59	1853	20.92	6.16	4.27	82.05	1
6	Bojan Bogdanović	SF	63	63	2083	20.24	4.11	2.08	74.05	1
7	Evan Fournier	SF	66	66	2076	18.50	2.59	3.17	74.00	1
8	T.J. Warren	SF	67	67	2202	19.79	4.18	1.48	72.87	1
9	Andrew Wiggins	SF	54	54	1858	21.76	5.07	3.67	69.18	1
10	Paul George	SF	48	48	1419	21.52	5.67	3.85	65.69	1

Fonte: elaborado pelo autor.

Tabela 6 – Top 10 SG 2019-20

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
1	James Harden	SG	68	68	2483	34,34	6,56	7,53	100,00	2
2	Devin Booker	SG	70	70	2512	26,61	4,24	6,51	76,67	2
3	Bradley Beal	SG	57	57	2053	30,54	4,25	6,09	70,51	2
4	Donovan Mitchell	SG	69	69	2364	23,99	4,41	4,26	67,01	2
5	CJ McCollum	SG	70	70	2556	22,23	4,24	4,43	62,36	2
6	Buddy Hield	SG	72	44	2216	19,19	4,56	2,96	56,23	2
7	Shai Gilgeous-Alexander	SG	70	70	2428	19,01	5,89	3,31	55,17	2
8	Jrue Holiday	SG	61	61	2117	19,13	4,80	6,69	48,70	2
9	Dillon Brooks	SG	73	73	2112	16,18	3,33	2,05	47,39	3

10	Jaylen Brown	SG	57	57	1934	20,33	6,39	2,07	46,64	3
----	--------------	----	----	----	------	-------	------	------	-------	---

Fonte: elaborado pelo autor.

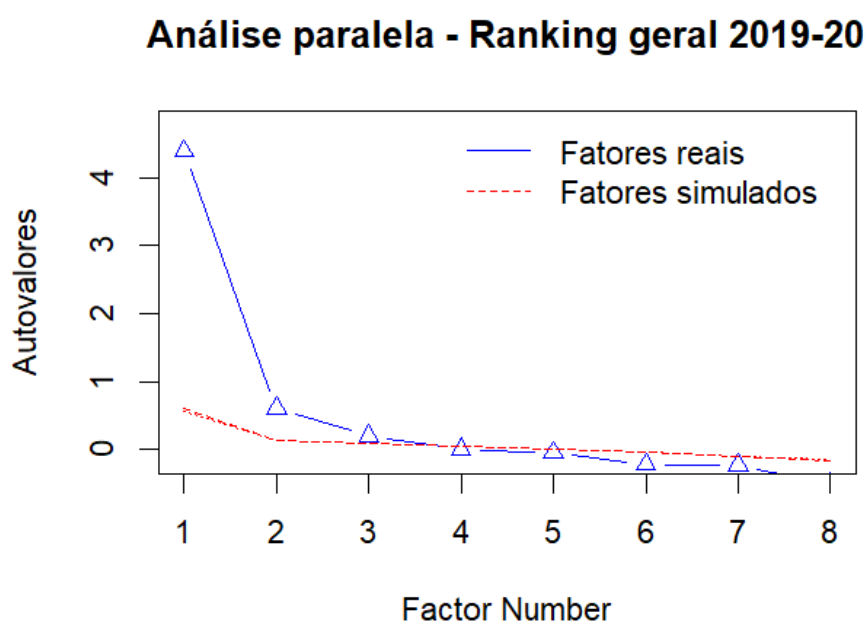
8.2.7 Ranking geral 2019-2020

Como etapa final, foi construído um índice geral de desempenho para a temporada 2019–20, combinando todas as posições em um único ranking. Para isso, considerou-se, para cada jogador elegível, o conjunto de variáveis G, GS, PTS, TRB, AST, STL, BLK e FT, que sintetizam participação, produção ofensiva e contribuições defensivas.

O procedimento foi análogo ao das posições:

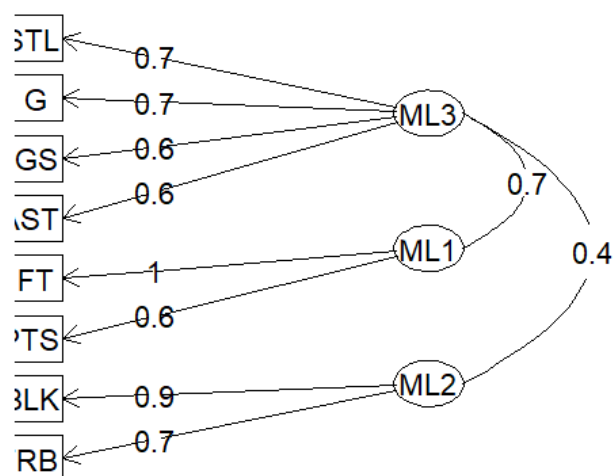
- Padronização das variáveis (escores z);
- Avaliação de KMO global e teste de Bartlett – ambos adequados para AFE;
- Definição do número de fatores via análise paralela;
- Estimação da AFE geral com rotação oblíqua (oblimin);
- Cálculo dos escores fatoriais e ponderação pelos percentuais de variância explicada;
- Normalização do índice geral para a escala 0–100.

Figura 7 - Análise paralela Geral 2019-20



Fonte: elaborado pelo autor.

Figura 8 - Diagrama de fatores Geral 2019-2020

Fatores - Ranking Geral 2019-20

Fonte: elaborado pelo autor.

Tabela 7 – Top 10 Geral 2019-20

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
1	James Harden	SG	68	68	2483	34,34	6,56	7,53	100,00	2
2	Devin Booker	SG	70	70	2512	26,61	4,24	6,51	76,67	2
3	Bradley Beal	SG	57	57	2053	30,54	4,25	6,09	70,51	2
4	Donovan Mitchell	SG	69	69	2364	23,99	4,41	4,26	67,01	2
5	CJ McCollum	SG	70	70	2556	22,23	4,24	4,43	62,36	2
6	Buddy Hield	SG	72	44	2216	19,19	4,56	2,96	56,23	2
7	Shai Gilgeous-Alexander	SG	70	70	2428	19,01	5,89	3,31	55,17	2
8	Jrue Holiday	SG	61	61	2117	19,13	4,80	6,69	48,70	2

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	Score_norm	cluster
9	Dillon Brooks	SG	73	73	2112	16,18	3,33	2,05	47,39	3
10	Jaylen Brown	SG	57	57	1934	20,33	6,39	2,07	46,64	3

Fonte: elaborado pelo autor.

A tabela 7 apresenta o Top 10 geral da temporada segundo o índice proposto. Em seguida, foi realizada uma comparação direta com o MVP oficial da temporada (Giannis Antetokounmpo). O script identifica a posição do MVP no ranking geral e o respectivo score normalizado.

8.3 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2020-2021

Nesta seção, é realizada a análise exploratória dos dados da temporada 2020-21 da NBA, com o objetivo de compreender a distribuição das variáveis estatísticas entre os atletas. São observadas características como tempo de jogo, frequência como titular, aproveitamentos ofensivos e ações defensivas, buscando identificar comportamentos distintos entre os jogadores. A análise permite evidenciar padrões de desempenho e variações relevantes que serão consideradas nas próximas etapas do trabalho.

8.3.1 Estatísticas das variáveis dos dados da temporada 2020-2021

Nesta subseção, são descritas as estatísticas descritivas das variáveis da temporada 2020-21 da NBA, com destaque para medidas de tendência central, dispersão e amplitude. A análise fornece uma base sólida para interpretar o comportamento dos jogadores ao longo da temporada, além de evidenciar a diversidade de atuações e o perfil estatístico da liga nesse período.

Tabela 8 - Estatísticas das variáveis 20-21

VARIÁVEL	MÉDIA	D.PADRÃO	MÍN	Q1	MEDIANA	Q3	MÁX
G	37,3687	21,2691	1	19	37	57	72
GS	16,9418	21,6037	0	0	5	29	72
MP	833,7404	673,7783	3	237	683	1333	2667

FG	141,23541	142,9285	0	31	95	208	732
FGA	303,7744	297,9567	0	70	212	456	1396
FG.	0,4435	0,1124	0	0,397	0,4395	0,4952	1
X3P	43,2269	51,6356	0	4	23	66	337
X3PA	118,4978	132,3180	0	15	66	188	801
X3P.	0,3146	0,1284	0	0,271	0,338	0,388	1
X2P	98,0085	110,0184	0	19	59	137	640
X2PA	185,2765	202,0527	0	39	115	258	1056
X2P.	0,5067	0,1299	0	0,458	0,51	0,573	1
eFG.	0,5089	0,1129	0	0,476	0,52	0,564	1
FT	58,3262	74,2927	0	10	33	75	484
FTA	75,1432	91,9077	0	13	44	102	581
FT.	0,7519	0,1464	0	0,683	0,775	0,845	1
ORB	34,1815	42,1904	0	7	20	43	297
DRB	119,2737	116,3931	0	31	86	181	720
TRB	153,4553	152,1111	0	39	108	228	960
AST	85,1900	104,9851	0	15	49	108	763
STL	26,2936	23,8924	0	7	19	40	128
BLK	16,9900	21,5063	0	3	10	22	190
TOV	45,7177	47,6618	0	10	32	65	312
PF	66,9489	52,6660	0	22	56	107	237
PTS	384,0241	392,3153	0	85	256	582	2015

Fonte: elaborado pelo autor (2025)

Observando a tabela de estatísticas das variáveis 20-21, diversas informações relevantes podem ser extraídas, oferecendo uma visão sobre as características dos jogadores da NBA na temporada 2020-2021:

a) Jogos e Participação: O número de jogos disputados apresenta média de 37,37 jogos por jogador (mínimo 1, máximo 72). Quanto aos jogos como titular, a mediana de apenas 5 jogos, contrastando com o máximo de 72, revela clara distinção entre titulares fixos e reservas.

b) Minutos Jogados: Média de 833,74 minutos por temporada, variando de 3 a 2.667 minutos. A mediana de 683 minutos indica que metade dos jogadores tem participação moderada, enquanto um grupo seleta acumula tempo significativamente maior.

c) Eficiência Ofensiva: A porcentagem de arremessos de quadra tem média de 44,35% com mediana de 43,95%, sugerindo distribuição relativamente simétrica. Nas cestas de três pontos, a média de 31,46% com mediana de 33,8% demonstra eficiência considerável dos especialistas.

d) Volume Ofensivo: Os arremessos convertidos apresentam média de 141,24 cestas por temporada (máximo 732), enquanto as tentativas têm média de 303,77 (máximo 1.396). Nos pontos totais, a média de 384,02 contrasta com o máximo de 2.015, evidenciando o papel dominante dos principais pontuadores.

e) Rebotes: Média de 153,47 rebotes totais por temporada (máximo 960), com predominância dos defensivos (média 119,27) sobre os ofensivos (média 34,18). A mediana de rebotes totais (108) significativamente abaixo da média indica concentração em jogadores específicos.

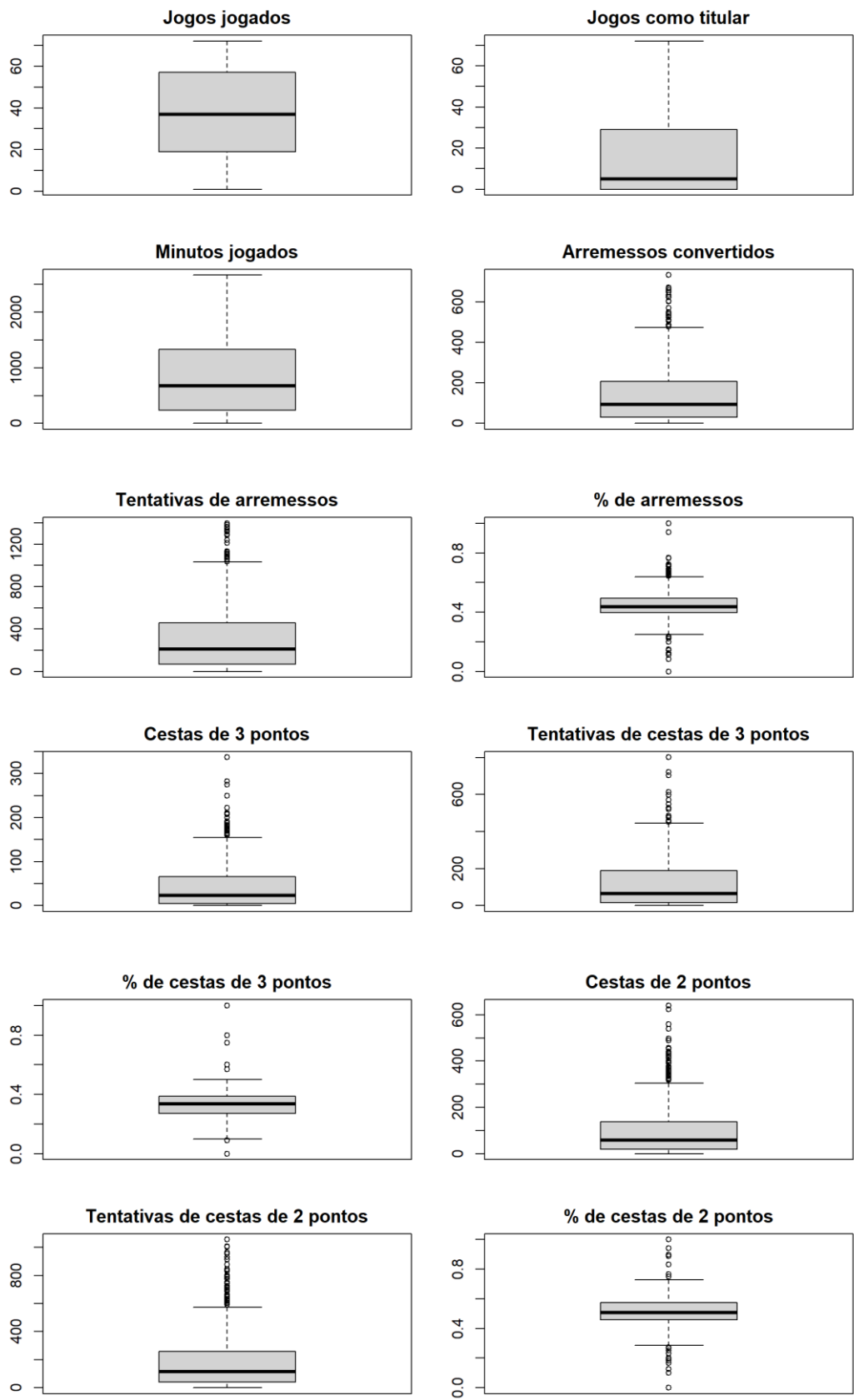
f) Estatísticas Defensivas e de Criação: As assistências apresentam média de 85,19 (máximo 763), evidenciando especialização dos armadores. Roubo de bola (média 26,29) e bloqueios (média 16,99) mostram medianas abaixo das médias, indicando que poucos jogadores se destacam nestas estatísticas.

Esta análise revela a diversidade de perfis na NBA, com clara especialização em determinados fundamentos e distribuição assimétrica para variáveis de volume.

8.3.2 Análise gráfica das variáveis da temporada 2020-2021

A seguir, são exibidas as representações gráficas das variáveis da temporada 2020-21 da NBA. Através de boxplots, histogramas e correlogramas, busca-se compreender o comportamento estatístico dos dados, destacando a presença de outliers, a forma das distribuições e a força das correlações entre variáveis. Essa visualização gráfica é essencial para complementar as análises numéricas e fundamentar as próximas etapas.

Figura 9 - Boxplots das variáveis 20-21



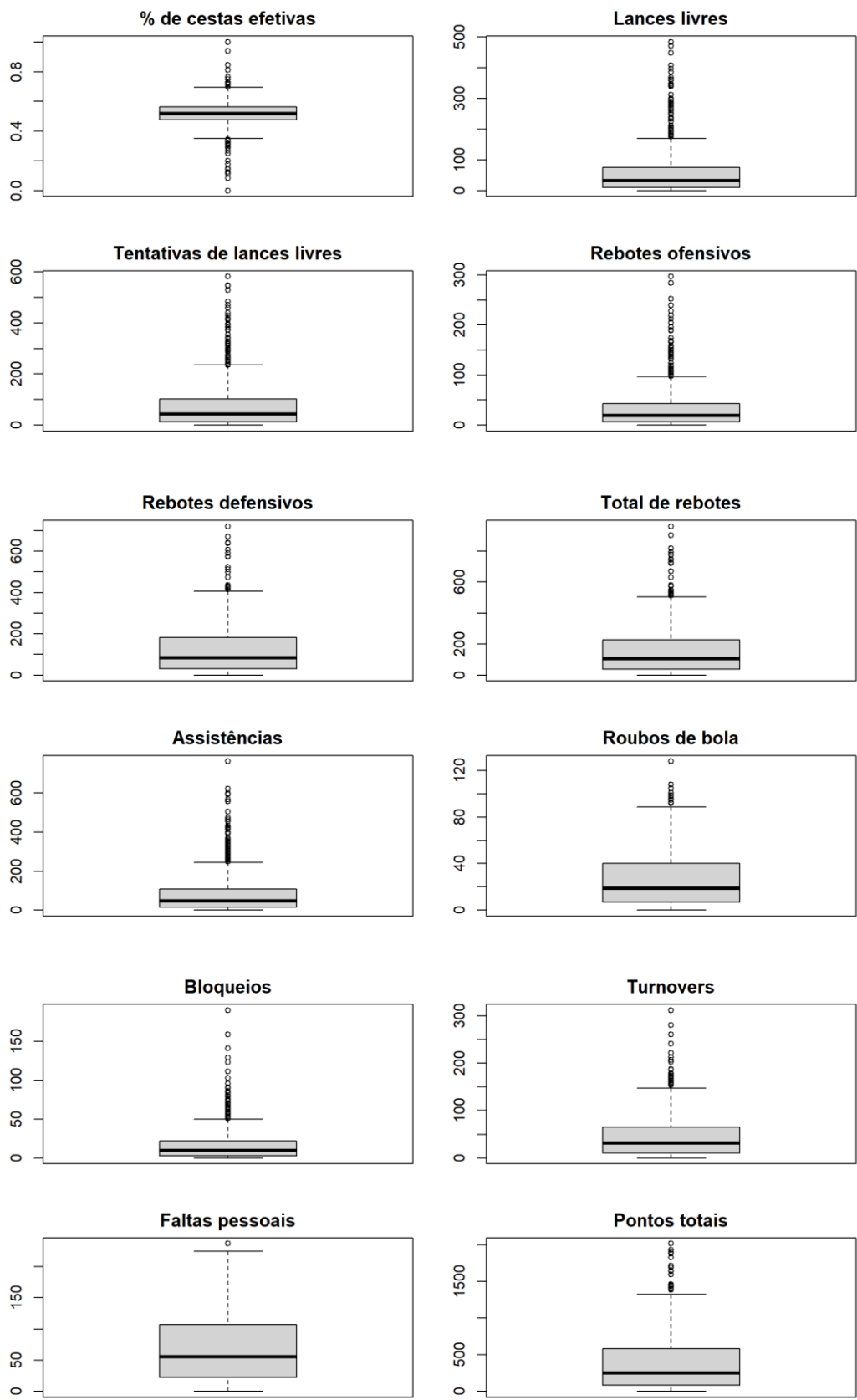
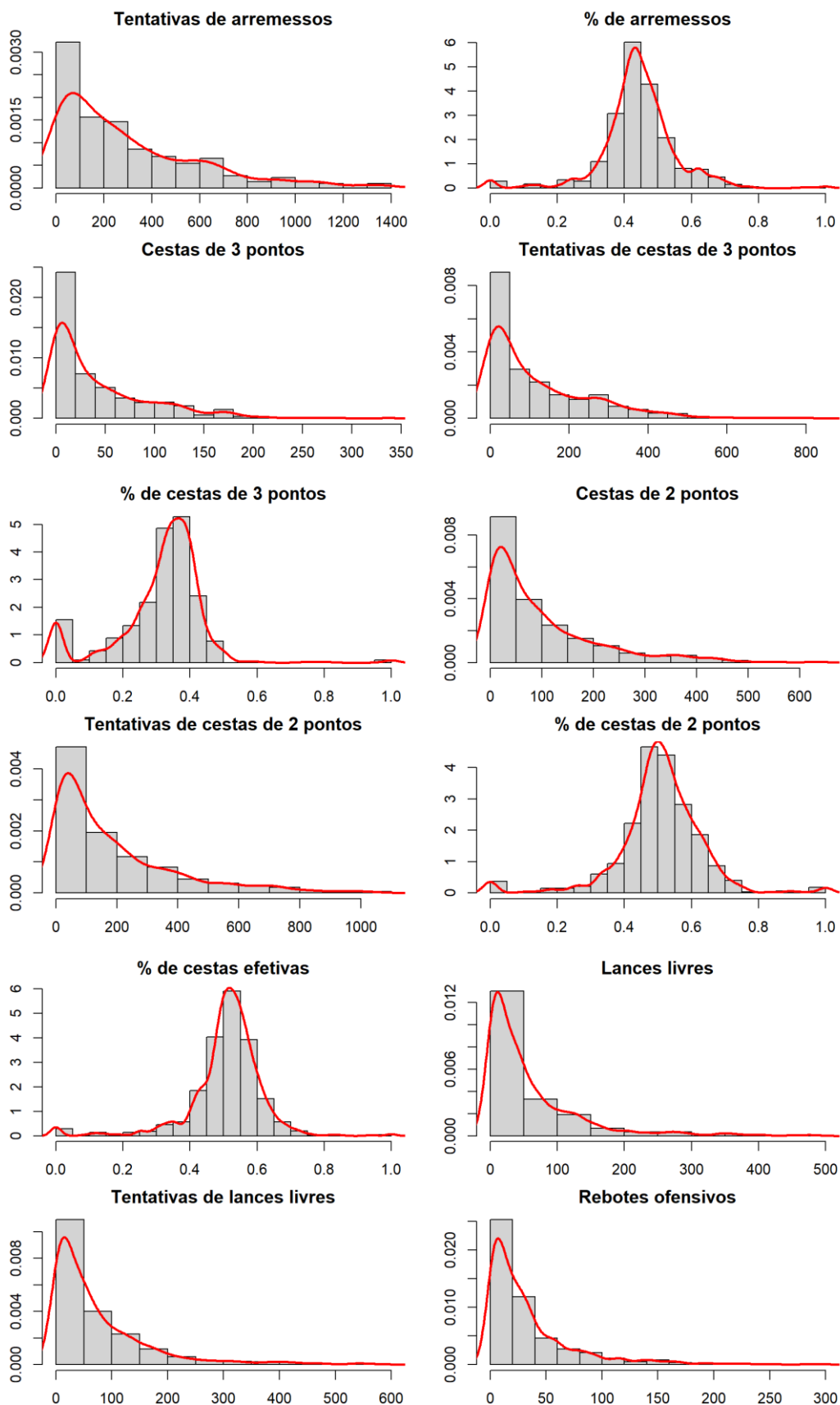
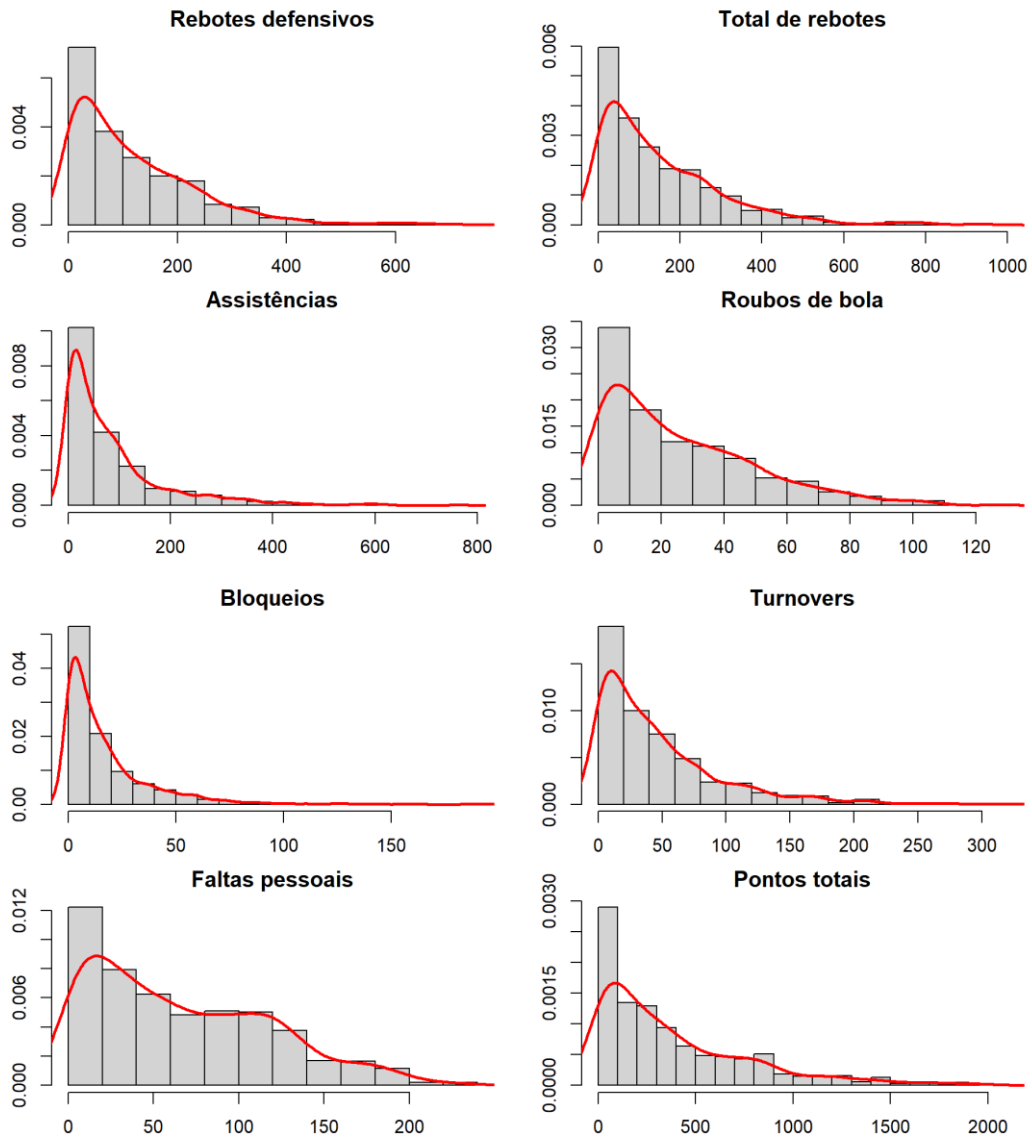


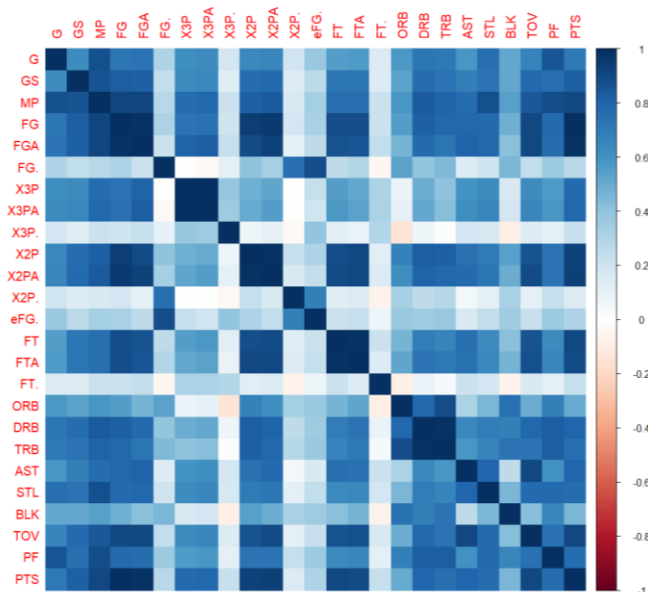
Figura 10 - Histogramas das variáveis 20-21





Fonte: elaborado pelo autor (2025)

Figura 11 - Correlograma das variáveis 20-21



Fonte: elaborado pelo autor (2025)

Através da análise dos gráficos apresentados, observa-se que as variáveis de participação em jogos revelam padrões distintos: Jogos Jogados mostra distribuição multimodal (picos em 20, 40 e 60 jogos), enquanto Jogos como Titular apresenta forte assimetria à direita, evidenciando a hierarquia nas equipes.

Os Minutos Jogados exibem distribuição assimétrica à direita, com mediana em torno de 683 minutos e outliers ultrapassando 2000 minutos, destacando a concentração de tempo em jogadores específicos.

As variáveis de volume ofensivo (Tentativas de Arremessos, Cestas de 3 e 2 Pontos) apresentam distribuições fortemente assimétricas à direita, com poucos jogadores concentrando grande parte das ações ofensivas. Em contraste, as métricas de eficiência (Porcentagem de Arremessos, Cestas de 3 e 2 Pontos) mostram distribuições mais simétricas (médias entre 34-51%), sugerindo homogeneidade na eficiência independente da posição.

Rebotes (Ofensivos, Defensivos, Totais) revelam assimetria à direita, com medianas de 20, 86 e 108 respectivamente, e outliers chegando a 960 rebotes totais, refletindo a especialização de pivôs e alas-pivôs. Assistências e Turnovers seguem padrão similar, concentrados em jogadores com maior posse de bola, com outliers de assistências ultrapassando 700.

Estatísticas defensivas como Bloqueios e Roubos de Bola apresentam forte assimetria à direita (medianas 10 e 19), enquanto Faltas Pessoais mostram distribuição mais homogênea (mediana 56).

Em síntese, a análise revela diversidade de perfis na NBA, com clara especialização em determinados fundamentos, distribuições assimétricas para variáveis de volume e mais simétricas para eficiência, fornecendo insights valiosos para a criação de rankings por posição e geral.

8.3.3 Análise de componentes principais 2020-2021

Para a temporada 2020–21, após a etapa de filtragem com os critérios mínimos de participação — pelo menos 20 jogos e 300 minutos totais — a base de dados ficou composta por 478 jogadores. Esse filtro garante que apenas atletas com participação minimamente relevante ao longo da temporada sejam considerados no processo de construção dos índices, evitando que desempenhos pontuais (outliers com poucos minutos) distorçam os resultados dos modelos multivariados.

A PCA foi aplicada separadamente para cada posição (PG, SG, SF, PF, C), usando sempre um conjunto de variáveis relevantes para o papel de cada função em quadra. Antes da PCA, foram avaliadas as medidas de adequação da matriz de correlações:

- KMO global sempre acima de 0,78, indicando boa adequação amostral:
 - PG: 0,825
 - SG: 0,795
 - SF: 0,822
 - PF: 0,789
 - C: 0,809
- Teste de esfericidade de Bartlett altamente significativo em todas as posições (p-values da ordem de 10^{-140} ou menores), rejeitando a hipótese de matriz identidade e confirmando a existência de estrutura latente.

A variância explicada pelos três primeiros componentes principais foi elevada em todas as posições:

- Armadores (PG): 89,09%
- Alas-armadores (SG): 86,13%
- Alas (SF): 89,93%
- Alas-pivôs (PF): 83,94%

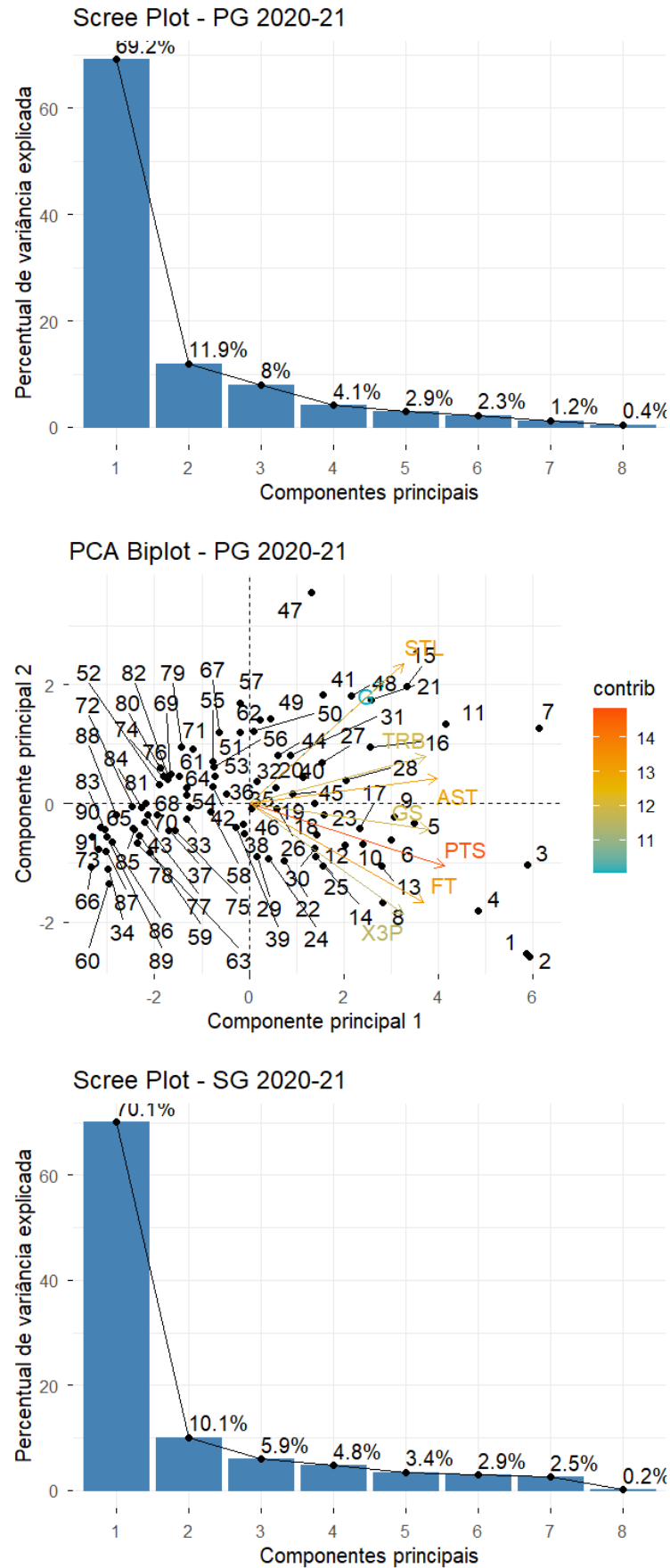
- Pivôs (C): 85,87%

Esses percentuais mostram que poucas combinações lineares das estatísticas já capturam grande parte da informação original, o que é desejável para fins de redução de dimensionalidade e visualização.

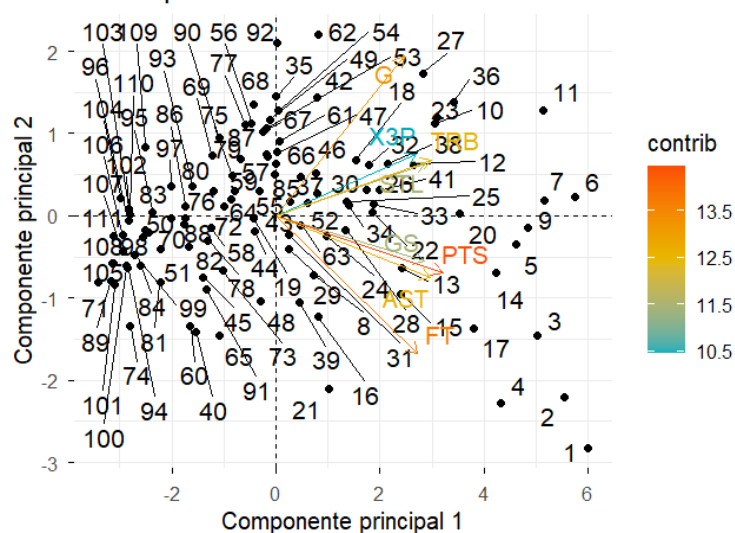
Em termos conceituais:

- Nos armadores (PG), os componentes principais sintetizam combinações de:
 - volume e organização do jogo (G, GS, AST, TRB, STL) e
 - capacidade de pontuar e arremessar de fora (PTS, X3P, FT).
- Nos alas-armadores (SG) e alas (SF), as combinações refletem sobretudo:
 - pontuação,
 - eficiência de arremessos (X3P, FT) e
 - produção geral (rebotes, assistências, roubos).
- Em PF e C, os componentes concentram:
 - pontuação próxima à cesta (X2P),
 - rebotes ofensivos e defensivos (ORB, DRB),
 - tocos (BLK) e
 - volume de participação (G, GS)

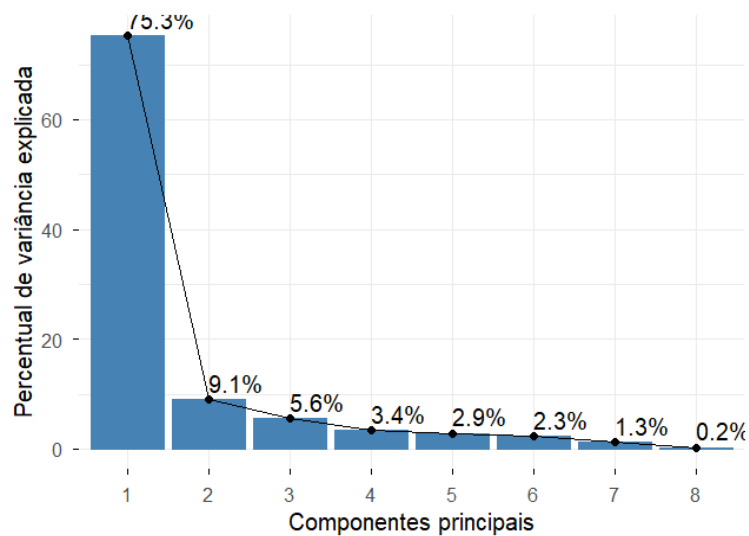
Figura 12 - Biplots e Screeplots de cada posição 2020-2021.



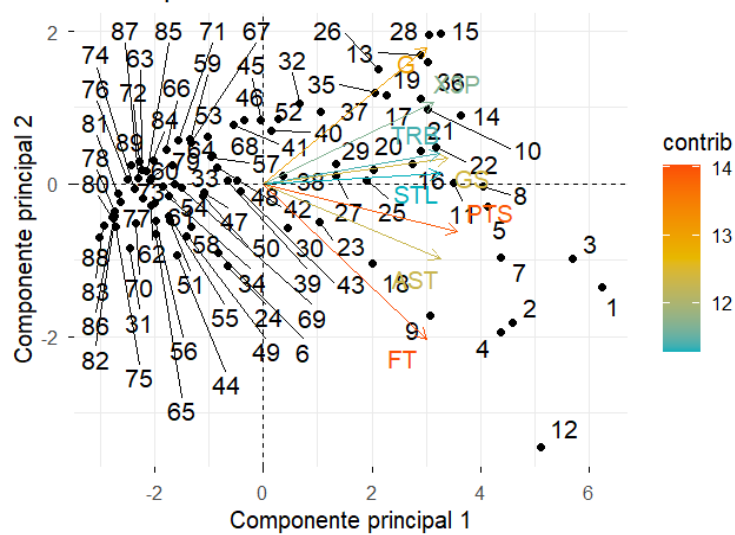
PCA Biplot - SG 2020-21



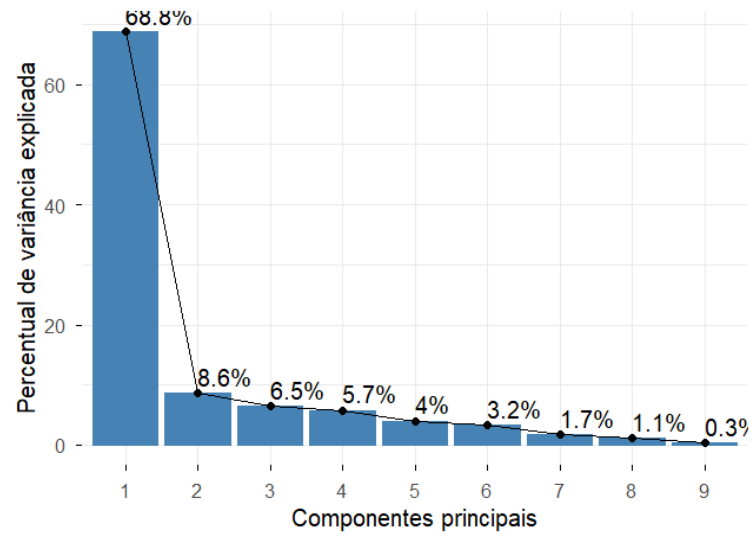
Scree Plot - SF 2020-21



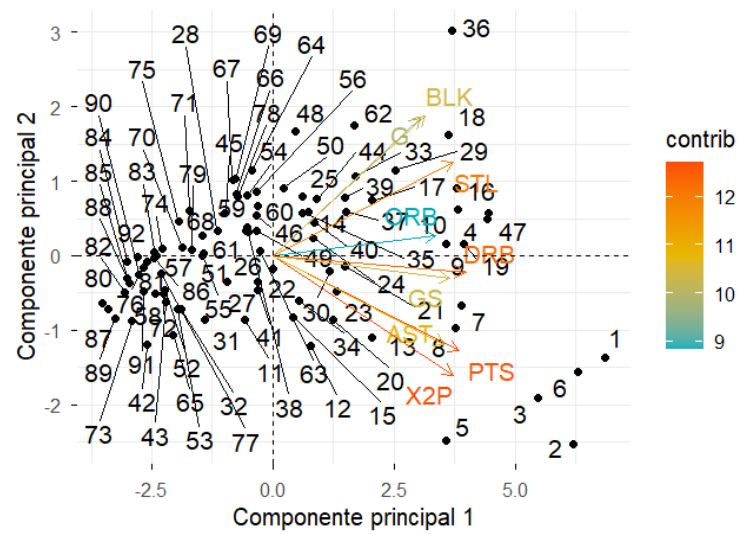
PCA Biplot - SF 2020-21

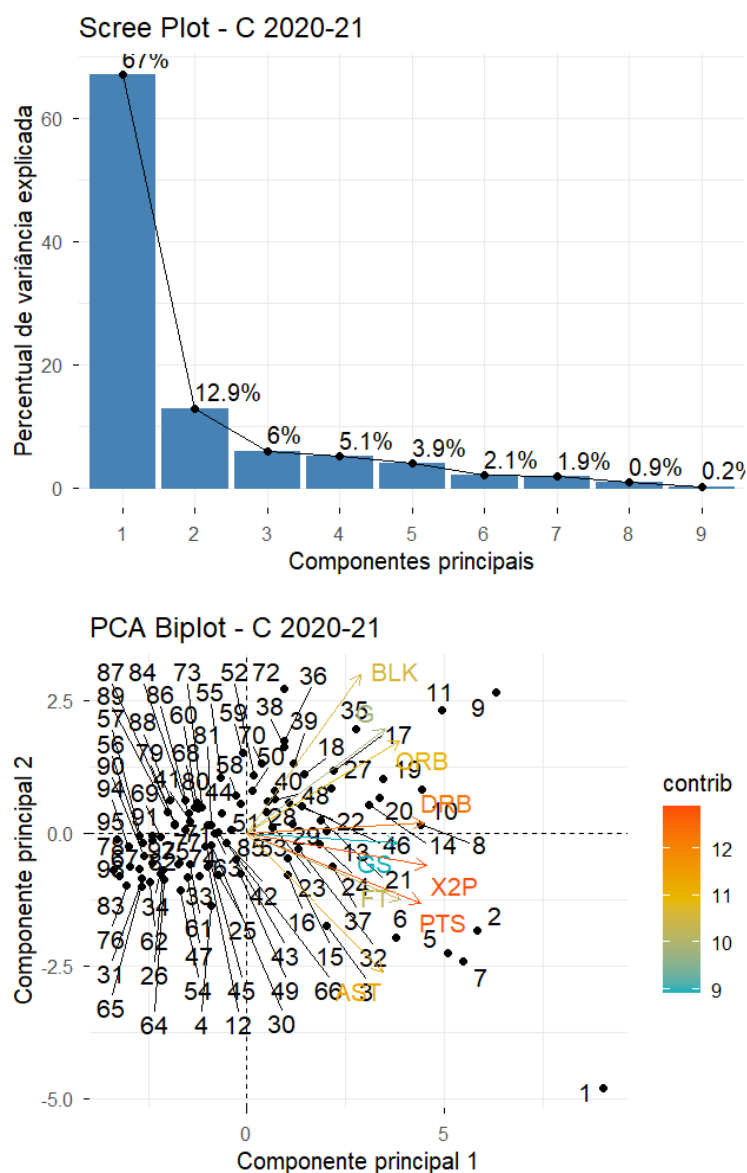


Scree Plot - PF 2020-21



PCA Biplot - PF 2020-21





Fonte: elaborado pelo autor.

8.3.4 Clusterização por k-médias 2020-2021

A partir dos escores dos primeiros componentes principais (geralmente os três primeiros), foi aplicado o algoritmo k-means para identificar perfis de jogadores dentro de cada posição. O número ótimo de clusters foi escolhido via índice de silhueta, calculado para k de 2 a 8.

Em todas as posições, o critério de silhueta sugeriu $k = 2$:

- PG: $k^* = 2$, silhueta média $\approx 0,457$
- SG: $k^* = 2$, silhueta média $\approx 0,488$
- SF: $k^* = 2$, silhueta média $\approx 0,606$

- PF: $k^* = 2$, silhueta média $\approx 0,506$
- C: $k^* = 2$, silhueta média $\approx 0,472$

As silhuetas médias em torno de 0,45–0,60 indicam um nível de separação moderado a bom entre os grupos, em linha com o que se espera em dados esportivos, onde não há “clusters perfeitos”, mas sim perfis de atuação que se sobrepõem parcialmente. Em geral:

- Um cluster tende a agrupar jogadores com maior protagonismo ofensivo e maior carga de minutos.
- O outro cluster agrupa atletas mais coadjuvantes, com participação estatística mais modesta.

8.3.5 Análise fatorial exploratória por posição 2020-2021

Depois da PCA e do k-means, foi aplicada a Análise Fatorial Exploratória (AFE) em cada posição, utilizando:

- método de estimação: máxima verossimilhança ($fm = "ml"$),
- rotação oblíqua: oblimin, permitindo correlação entre fatores,
- número de fatores: definido por análise paralela ($fa.parallel$).

Armadores (PG)

- Análise paralela: sugere 2 fatores.
- Comunalidades (h^2): de 0,36 (G) a 0,99 (PTS), indicando que a maior parte da variância dos indicadores é bem explicada pelo modelo fatorial.
- Os fatores podem ser interpretados como:
 1. Fator 1 – Eficiência de arremessos e pontuação
 - a. Altas cargas em X3P, PTS, FT.
 2. Fator 2 – Volume e contribuição all-around
 - a. Altas cargas em AST, TRB, STL, além de G e GS.
- A variância explicada cumulativa é de aproximadamente 75%, o que indica bom poder de síntese da estrutura latente.

Alas-armadores (SG)

- KMO global: 0,795; Bartlett significativo.
- Análise paralela: sugere 2 fatores.
- Comunalidades elevadas, com FT e PTS explicados quase integralmente ($h^2 \sim 0,99$).

Os fatores podem ser lidos como:

1. Fator 1 – Eficiência de lances livres e pontuação (FT, PTS, AST, GS).

2. Fator 2 – Volume ofensivo e contribuição estatística de perímetro (X3P, G, TRB, STL).

Alas (SF)

- KMO global: 0,822; Bartlett altamente significativo.
- Análise paralela: sugere 1 fator.
- Comunalidades: todas altas (0,52–0,96), indicando forte unidimensionalidade.

Nesse caso, a estrutura é essencialmente um fator geral de desempenho:

- Altas cargas em quase todas as variáveis: PTS, GS, FT, AST, X3P, TRB, STL, G.

Ou seja, o desempenho dos SF é sintetizado por um score global que valoriza fortemente pontuação, minutos titulares, eficiência de arremesso e contribuição em rebotes, assistências e roubos.

Alas-pivôs (PF)

- KMO global: 0,789; Bartlett significativo.
- Análise paralela: sugere 1 fator.
- Comunalidades: variando de 0,43 (G) a 0,86 (DRB).

A interpretação é novamente de um fator geral de desempenho de frontcourt, com ênfase em:

- rebotes defensivos (DRB),
- pontuação (PTS),
- aproveitamento em dois pontos (X2P),
- tocos e rebotes ofensivos (BLK, ORB),
- participação em jogos e minutos como titular (G, GS).

Pivôs (C)

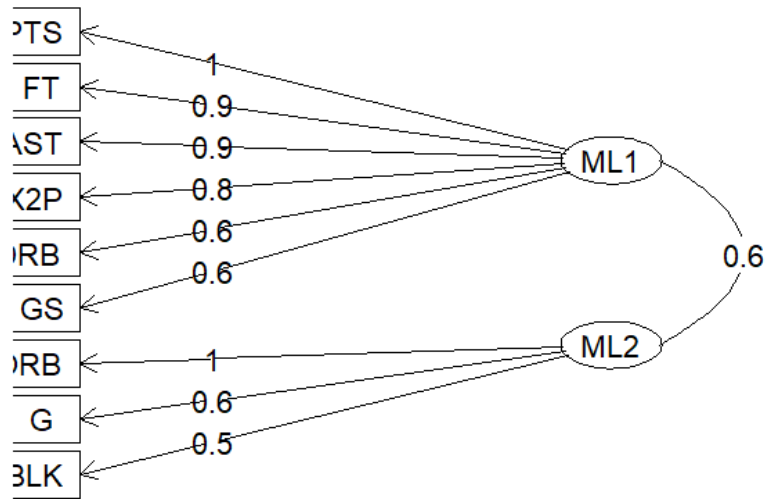
- KMO global: 0,809; Bartlett significativo.
- Análise paralela: sugere 2 fatores.
- Comunalidades altas para a maioria das variáveis, especialmente ORB, PTS e X2P (próximas de 1).

A estrutura pode ser interpretada como:

1. Fator 1 – Pontuação e eficiência ofensiva do garrafão
 - a. Altas cargas em PTS, FT, AST, X2P, DRB, GS.
2. Fator 2 – Presença física e proteção de aro
 - a. Cargas em ORB, G, BLK, além de contribuição em rebotes.

Figura 13 - Análises paralelas e Diagramas de fatores por posição 2020-2021

Diagrama de Fatores - C 2020-21



Análise Paralela - C 2020-21

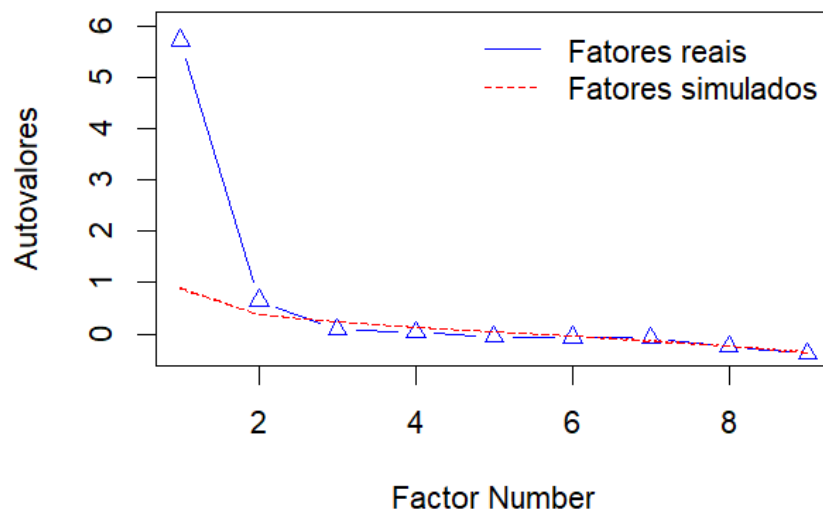
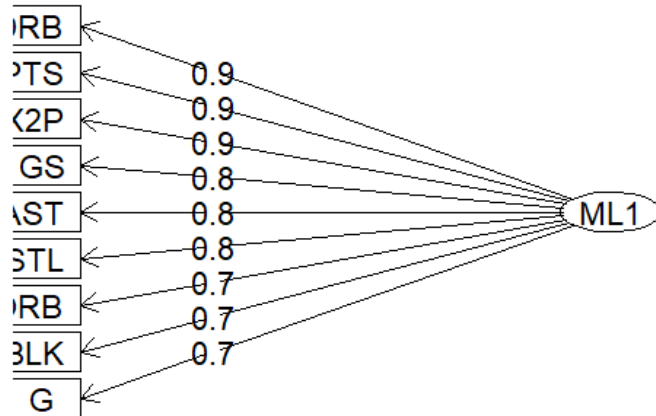


Diagrama de Fatores - PF 2020-21



Análise Paralela - PF 2020-21

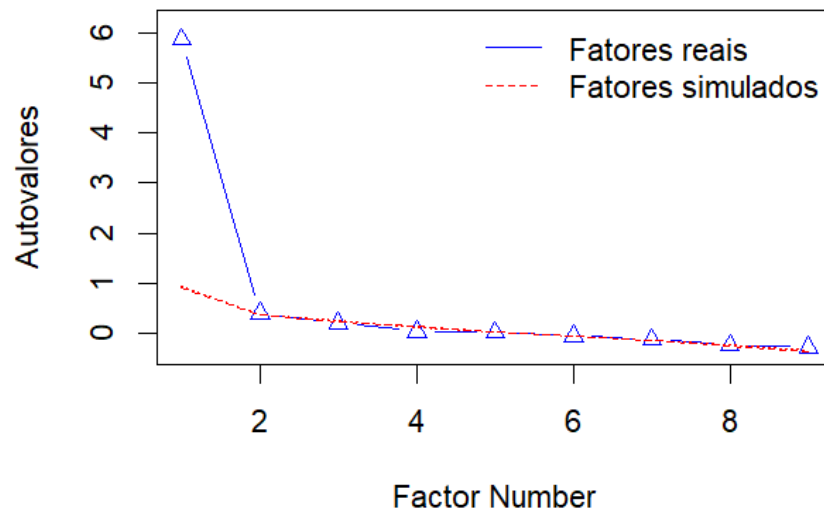
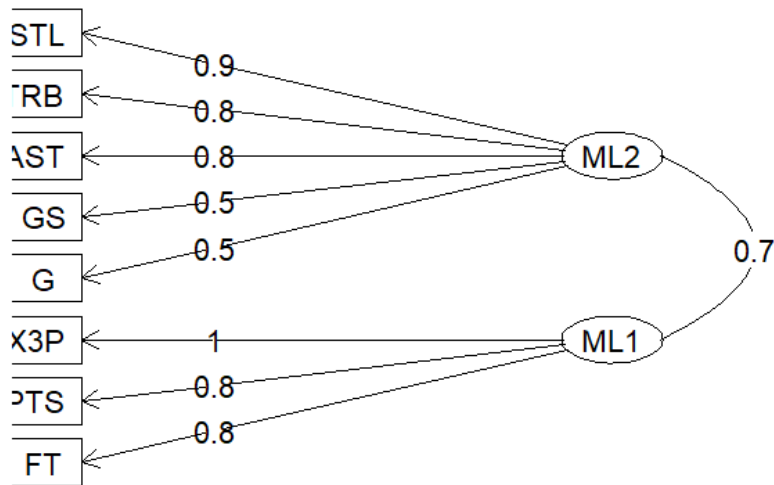


Diagrama de Fatores - PG 2020-21



Análise Paralela - PG 2020-21

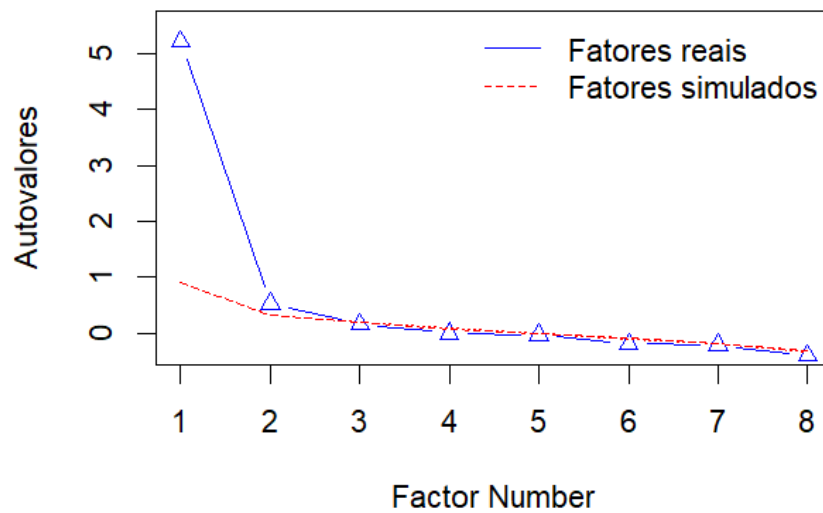
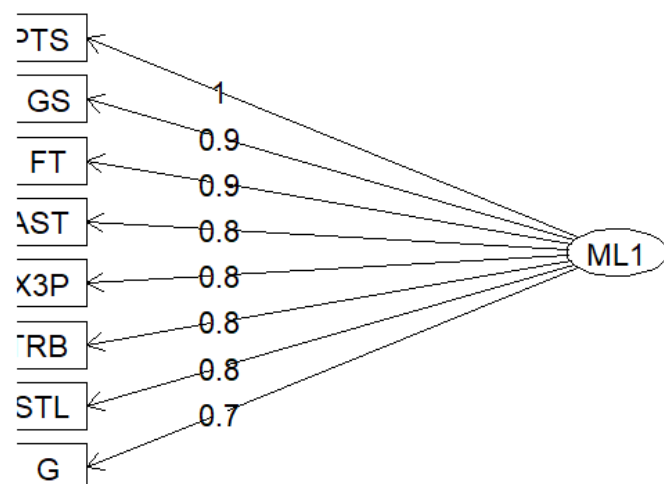


Diagrama de Fatores - SF 2020-21



Análise Paralela - SF 2020-21

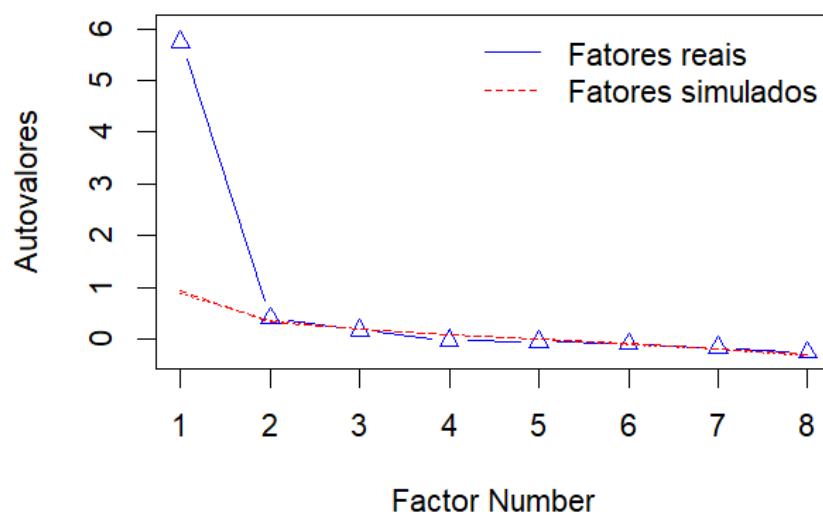
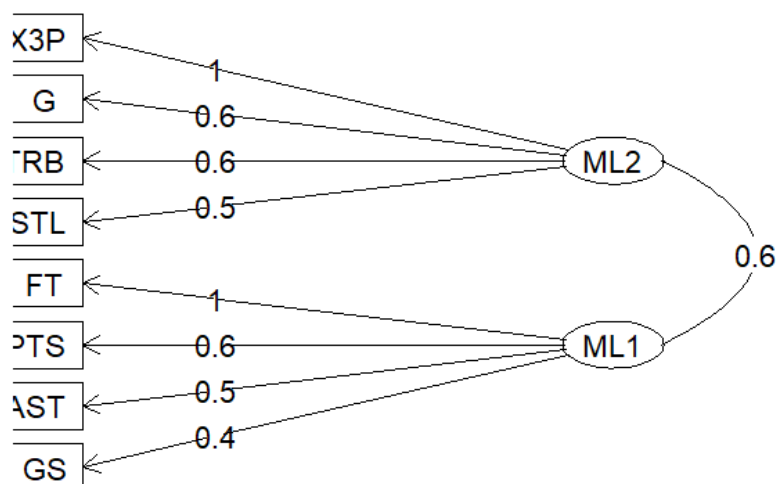
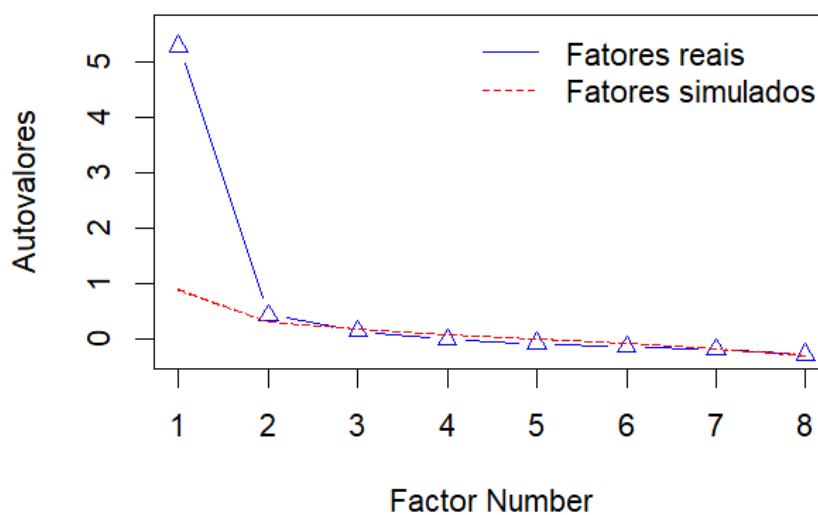


Diagrama de Fatores - SG 2020-21



Análise Paralela - SG 2020-21



Fonte: elaborado pelo autor.

8.3.6 Construção de rankings por posição 2020-2021

A partir das AFE por posição, foram construídos índices de desempenho específicos para PG, SG, SF, PF e C, com duas etapas principais:

1. Cálculo dos escores fatoriais (via método de regressão).
2. Combinação dos fatores usando pesos proporcionais à variância explicada (pesos = $V_{\text{accounted}}[2,] / \text{soma}$). Como por exemplo a equação à seguir:

$$Score_{PG} = 0.334Z_G + 0.471Z_{GS} + 0.512Z_{PTS} + 0.517Z_{AST} + 0.419Z_{STL} + 0.365Z_{X3P} + 0.456Z_{FT} + 0.491Z_{TRB} \quad (9)$$

Em todas as posições:

- O score bruto (score_ranking) é transformado para uma escala 0–100:

$$score_{norm} = \frac{scoreranking - \min(scoreranking)}{\max(scoreranking) - \min(scoreranking)} \times 100 \quad (10)$$

O ranking final é obtido ordenando os jogadores em ordem decrescente de score_norm, com rank = 1 representando o melhor jogador da posição.

Os resultados numéricos (não listados aqui por completo) geram, para cada posição, um Top 10 que sintetiza os atletas com maior impacto estatístico dentro do papel desempenhado em quadra.

Tabela 8 - Top10 C 2020-21

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Nikola Jokic	C	72	72	2488	26,36	10,83	8,32	100,00	2
2	Nikola Vucevic	C	70	70	2348	23,41	11,67	3,84	80,80	2
3	Clint Capela	C	63	63	1898	15,17	14,33	0,78	73,23	2
4	Jonas Valančiūnas	C	62	61	1755	17,06	12,52	1,81	71,35	2
5	Rudy Gobert	C	71	71	2187	14,30	13,52	1,25	68,90	2
6	Joel Embiid	C	51	51	1585	28,45	10,57	2,84	68,20	2
7	Deandre Ayton	C	69	69	2115	14,45	10,54	1,43	66,02	2
8	Enes Freedom	C	72	35	1758	11,19	11,04	1,17	64,86	2
9	Bam Adebayo	C	64	64	2143	18,70	8,95	5,41	63,42	2
10	Karl-Anthony Towns	C	50	50	1689	24,78	10,58	4,50	61,71	2

Fonte: elaborado pelo autor.

Tabela 9 -Top 10 C 2020-21

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Nikola Jokic	C	72	72	2488	26,36	10,83	8,32	100,00	2
2	Nikola Vucevic	C	70	70	2348	23,41	11,67	3,84	80,80	2
3	Clint Capela	C	63	63	1898	15,17	14,33	0,78	73,23	2
4	Jonas Valančiūnas	C	62	61	1755	17,06	12,52	1,81	71,35	2
5	Rudy Gobert	C	71	71	2187	14,30	13,52	1,25	68,90	2
6	Joel Embiid	C	51	51	1585	28,45	10,57	2,84	68,20	2
7	Deandre Ayton	C	69	69	2115	14,45	10,54	1,43	66,02	2
8	Enes Freedom	C	72	35	1758	11,19	11,04	1,17	64,86	2
9	Bam Adebayo	C	64	64	2143	18,70	8,95	5,41	63,42	2
10	Karl-Anthony Towns	C	50	50	1689	24,78	10,58	4,50	61,71	2

Fonte: elaborado pelo autor.

Tabela 10 – Top 10 PF 2020-21

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Giannis Antetokounmpo	PF	61	61	2013	28,15	11,00	5,85	100,00	1
2	Julius Randle	PF	71	71	2667	24,11	10,18	6,01	99,01	1
3	Domantas Sabonis	PF	62	62	2231	20,32	11,97	6,69	93,39	1
4	Zion Williamson	PF	61	61	2026	27,00	7,23	3,70	80,70	1
5	Tobias Harris	PF	62	62	2014	19,52	6,76	3,55	68,78	1
6	Andrew Wiggins	PF	71	71	2364	18,59	4,89	2,35	68,26	1
7	Draymond Green	PF	63	63	1982	7,05	7,11	8,86	67,04	1
8	Pascal Siakam	PF	56	56	2006	21,36	7,23	4,46	66,42	1
9	Kelly Olynyk	PF	70	62	1997	13,50	6,99	2,90	66,00	1

10	DeMar DeRozan	PF	61	61	2056	21,57	4,25	6,92	65,44	1
----	---------------	----	----	----	------	-------	------	------	-------	---

Fonte: elaborado pelo autor.

Tabela 11 – Top 10 PG 2020-21

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Luka Doncic	PG	66	66	2262	27,73	7,98	8,59	100,00	1
2	Russell Westbrook	PG	65	65	2369	22,23	11,54	11,74	99,65	1
3	Stephen Curry	PG	63	63	2152	31,98	5,48	5,76	95,12	1
4	Damian Lillard	PG	67	67	2398	28,78	4,22	7,54	94,83	1
5	Trae Young	PG	63	63	2125	25,30	3,89	9,43	85,06	1
6	De'Aaron Fox	PG	58	58	2036	25,19	3,50	7,19	76,10	1
7	Chris Paul	PG	70	70	2199	16,41	4,46	8,89	74,38	1
8	Kyrie Irving	PG	54	54	1886	26,87	4,76	6,00	72,18	1
9	Ja Morant	PG	63	63	2053	19,11	4,00	7,38	67,51	1
10	Dejounte Murray	PG	67	67	2139	15,69	7,06	5,42	67,47	1

Fonte: elaborado pelo autor.

Tabela 12 – Top 10 SF 2020-21

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Jayson Tatum	SF	64	64	2290	26,44	7,38	4,31	100,00	2
2	Khris Middleton	SF	68	68	2269	20,37	5,97	5,44	86,58	2
3	Brandon Ingram	SF	61	61	2093	23,77	4,90	4,85	84,50	2
4	Kawhi Leonard	SF	52	52	1773	24,85	6,48	5,17	77,38	2
5	Paul George	SF	54	54	1821	23,31	6,65	5,19	76,15	2
6	Norman Powell	SF	69	58	2205	18,59	3,12	1,87	75,21	2
7	Jimmy Butler	SF	52	52	1745	21,46	6,90	7,10	75,03	2
8	Bojan Bogdanović	SF	72	72	2216	17,01	3,90	1,89	73,58	2
9	Jerami Grant	SF	54	54	1829	22,31	4,63	2,81	68,83	2
10	Dillon Brooks	SF	67	67	1997	17,18	2,93	2,34	67,78	2

Fonte: elaborado pelo autor.

Tabela 13 – Top 10 SG 2020-21

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Bradley Beal	SG	60	60	2147	31,30	4,72	4,42	100,00	2
2	Devin Booker	SG	67	67	2270	25,55	4,19	4,31	92,19	2
3	Zach LaVine	SG	58	58	2034	27,43	4,98	4,86	90,00	2
4	Terry Rozier	SG	69	69	2383	20,39	4,38	4,25	83,61	2
5	Jaylen Brown	SG	58	58	1999	24,66	5,98	3,36	81,20	2
6	Anthony Edwards	SG	72	55	2314	19,33	4,67	2,93	79,48	2
7	Collin Sexton	SG	60	60	2115	24,33	3,12	4,37	77,17	2
8	Buddy Hield	SG	71	71	2433	16,56	4,70	3,62	74,81	2
9	Jordan Clarkson	SG	68	1	1818	18,40	4,03	2,47	73,42	2
10	RJ Barrett	SG	72	72	2511	17,56	5,75	3,01	70,92	2

Fonte: elaborado pelo autor.

8.3.7 Ranking geral 2020-2021

Para construir um índice geral de desempenho, combinando todas as posições, foi utilizada uma AFE única sobre as variáveis:

- G, GS, PTS, TRB, AST, STL, BLK, FT .

Os resultados principais foram:

- KMO global: 0,81 (boa adequação).
- MSA individuais: entre 0,63 (BLK) e 0,92 (GS), todos acima de 0,60.
- Análise paralela: sugere 3 fatores.
- AFE (3 fatores, rotação oblimin):
 - Fator ML2: FT, PTS, AST, GS, TRB (desempenho ofensivo geral e volume).
 - Fator ML1: STL e G (pressão defensiva no perímetro e presença em quadra).
 - Fator ML3: BLK e TRB (proteção de aro e controle dos rebotes).

A variância explicada cumulativa pelos três fatores foi de aproximadamente 77%, indicando que a estrutura latente é capaz de resumir a maior parte da heterogeneidade entre os jogadores.

Interpretação dos pesos:

- Maior ênfase em PTS, GS e FT, refletindo a importância de:

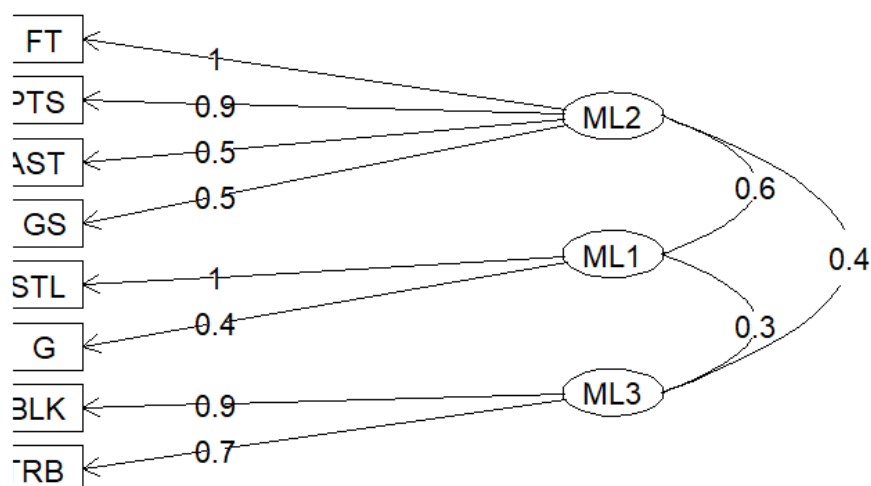
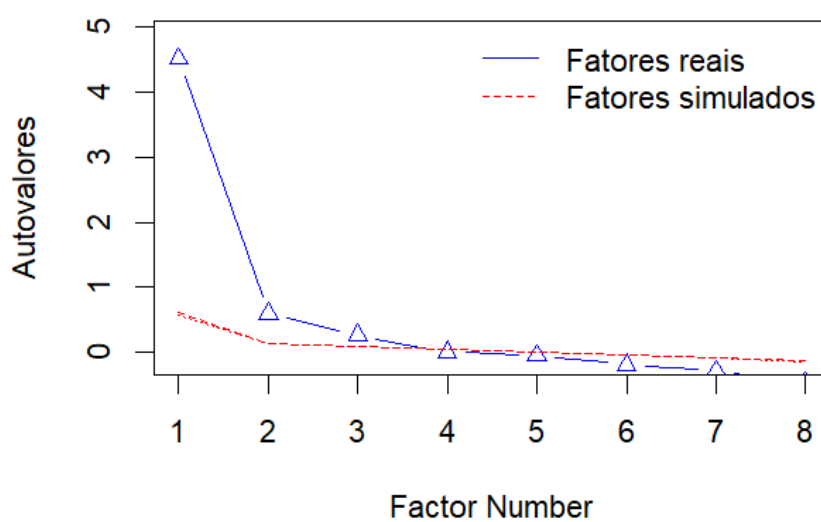
- pontuar com volume,
 - atuar muitos minutos como titular,
 - manter alta eficiência em lances livres.
- TRB e AST aparecem com pesos intermediários, capturando contribuição em rebotes e organização do jogo.
- STL e BLK também entram na equação, mas com peso ligeiramente menor, complementando o caráter defensivo do índice.

Após o cálculo do `score_geral_norm` (0–100) e a ordenação dos jogadores, foi obtido o ranking geral da temporada 2020–21. A checagem com o prêmio oficial de MVP mostra que:

- Nikola Jokic aparece na 1ª posição do ranking geral.
- Seu score normalizado é 100, por construção (jogador com melhor índice).

Essa evidência é extremamente importante para validar empiricamente o índice: um jogador reconhecido oficialmente como MVP, com narrativa estatística e qualitativa muito forte, aparece exatamente no topo do ranking construído por meio de técnicas multivariadas.

Figura 14 - Diagrama de fatores e Análise paralela gerais 2020-2021

Fatores - Ranking Geral 2020-21**Análise paralela - Ranking geral 2020-21**

Fonte: elaborado pelo autor.

Tabela 14 – Top 10 Geral 2020-21

rank_geral	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_geral_norm	cluster
1	Nikola Jokic	C	72	72	2488	26,36	10,83	8,32	100,00	2
2	Giannis Antetokounmpo	PF	61	61	2013	28,15	11,00	5,85	90,14	1
3	Stephen Curry	PG	63	63	2152	31,98	5,48	5,76	83,05	1
4	Russell Westbrook	PG	65	65	2369	22,23	11,54	11,74	82,97	1
5	Julius Randle	PF	71	71	2667	24,11	10,18	6,01	82,54	1
6	Luka Doncic	PG	66	66	2262	27,73	7,98	8,59	82,31	1
7	Rudy Gobert	C	71	71	2187	14,30	13,52	1,25	81,20	2
8	Nikola Vucevic	C	70	70	2348	23,41	11,67	3,84	80,51	2
9	Damian Lillard	PG	67	67	2398	28,78	4,22	7,54	79,32	1
10	Jayson Tatum	SF	64	64	2290	26,44	7,38	4,31	79,10	2

Fonte: elaborado pelo autor.

8.4 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2021-2022

A seguir, apresenta-se a análise exploratória dos dados da temporada 2021-22. Esta etapa visa descrever estatisticamente o comportamento das variáveis referentes ao desempenho dos atletas, avaliando desde métricas básicas, como pontos e rebotes, até indicadores de eficiência ofensiva e defensiva. O detalhamento dos dados permite reconhecer a diversidade de atuações dentro da liga, além de identificar possíveis valores discrepantes que refletem especializações ou estilos de jogo distintos.

8.4.1 Estatísticas das variáveis da temporada 2021-2022

A tabela apresentada a seguir resume os principais indicadores estatísticos das variáveis da temporada 2021-22 da NBA, abrangendo métricas ofensivas e defensivas dos jogadores. A distribuição das variáveis permite observar diferenças de volume e eficiência entre atletas, constituindo uma etapa essencial para fundamentar os próximos passos.

Tabela 15 - Estatísticas das variáveis 21-22

VARIÁVEL	MÉDIA	D.PADRÃO	MÍN	Q1	MEDIANA	Q3	MAX
G	36,7056	25,8980	1	12	36,5	61	82
GS	16,6724	23,8171	0	0	4	25	82
MP	825,1884	775,7863	1	121	577,5	1414,5	2854
FG	137,7339	154,7395	0	13,75	81	216,25	774
FGA	299,6305	330,8996	0	35	182,5	466,5	1564
FG.	0,4342	0,1377	0	0,3852	0,441	0,5	1
X3P	42,3472	55,2006	0	2	17	63,25	285
X3PA	120,0431	148,7174	0	7	54	186,25	750
X3P.	0,3035	0,1380	0	0,259	0,331	0,376	1
X2P	95,3866	116,5516	0	8	51	140	724
X2PA	179,5874	214,9275	0	18	100	263,25	1393
X2P.	0,5055	0,1578	0	0,451	0,516	0,579	1
eFG.	0,4975	0,1420	0	0,4642	0,517	0,563	1
FT	57,6490	81,0439	0	4	27	78,25	654
FTA	74,460	100,8416	0	7	36,5	102,25	803
FT.	0,7476	0,1561	0	0,6742	0,765	0,846	1
ORB	35,4544	46,01754	0	4	20	46	349
DRB	116,4273	128,7407	0	15	74	176,5	813
TRB	151,8817	168,1790	0	20,75	100	226	1019
AST	84,7068	113,3896	0	7	41	113,25	737
STL	26,0701	27,2314	0	3	17	42	138
BLK	16,0775	21,7708	0	1	9	22	177
TOV	44,8325	52,9628	0	5	28	65	303
PF	67,4470	61,6198	0	12	52	110	286
PTS	375,4642	426,8222	0	36,5	214	592,5	2155

Fonte: elaborado pelo autor (2025)

Observando a tabela de estatísticas das variáveis, diversas informações relevantes podem ser extraídas, oferecendo uma visão sobre as características dos jogadores da NBA na temporada 2021-2022:

a) Jogos e Participação: O número de jogos disputados apresenta média de 36,71 jogos por jogador (mínimo 1, máximo 82). Quanto aos jogos como titular, a mediana de apenas 4 jogos, contrastando com o máximo de 82, revela clara distinção entre titulares fixos e reservas, evidenciando a hierarquia nas equipes.

b) Minutos Jogados: Média de 825,19 minutos por temporada, variando de 1 a 2.854 minutos. A mediana de 577,5 minutos indica que metade dos jogadores tem participação moderada, enquanto um grupo seleto acumula tempo significativamente maior.

c) Eficiência Ofensiva: A porcentagem de arremessos de quadra tem média de 43,43% com mediana de 44,1%, sugerindo distribuição relativamente simétrica. Nas cestas de três pontos, a média de 30,35% com mediana de 33,1% demonstra eficiência considerável dos especialistas em arremessos de longa distância.

d) Volume Ofensivo: Os arremessos convertidos apresentam média de 137,73 cestas por temporada (máximo 774), enquanto as tentativas têm média de 299,63 (máximo 1.564). Nos pontos totais, a média de 375,46 contrasta com o máximo de 2.155, evidenciando o papel dominante dos principais pontuadores.

e) Rebotes: Média de 151,88 rebotes totais por temporada (máximo 1.019), com predominância dos defensivos (média 116,43) sobre os ofensivos (média 35,45). A mediana de rebotes totais (100) significativamente abaixo da média indica concentração em jogadores específicos, tipicamente pivôs e alas-pivôs.

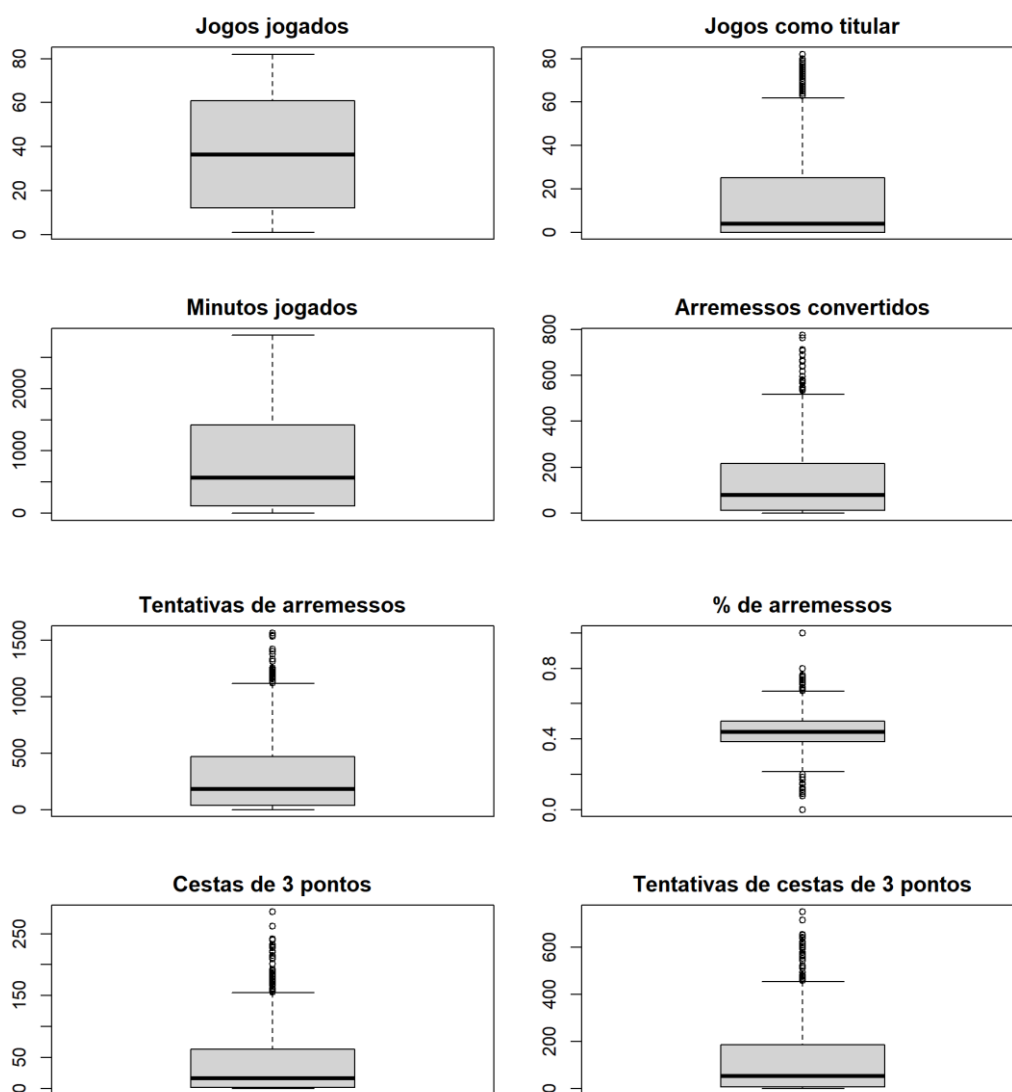
f) Estatísticas Defensivas e de Criação: As assistências apresentam média de 84,71 (máximo 737), evidenciando especialização dos armadores. Roubo de bola (média 26,07) e bloqueios (média 16,08) mostram medianas abaixo das médias, indicando que poucos jogadores se destacam nestas estatísticas defensivas.

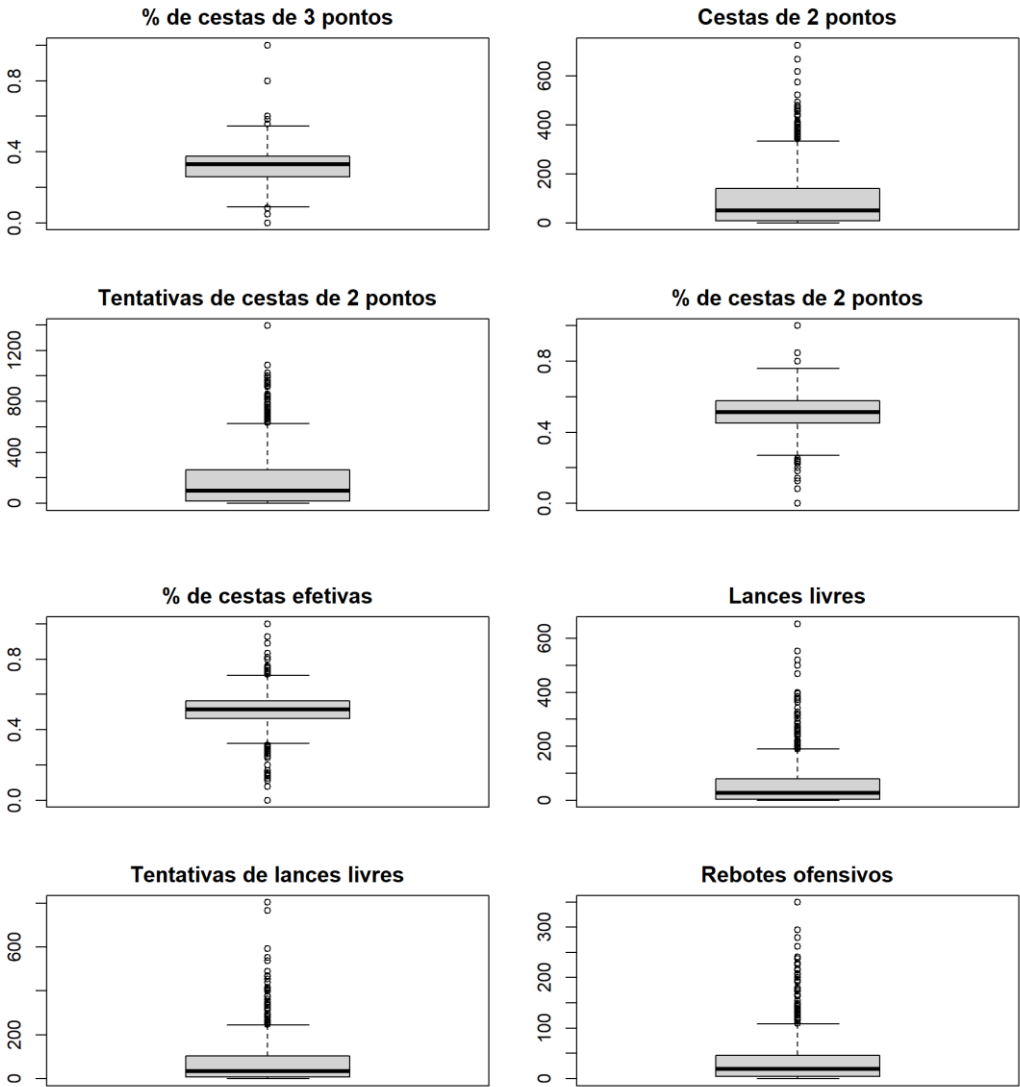
Esta análise revela a diversidade de perfis na NBA, com clara especialização em determinados fundamentos e distribuição assimétrica para variáveis de volume, contrastando com distribuições mais equilibradas para variáveis de eficiência.

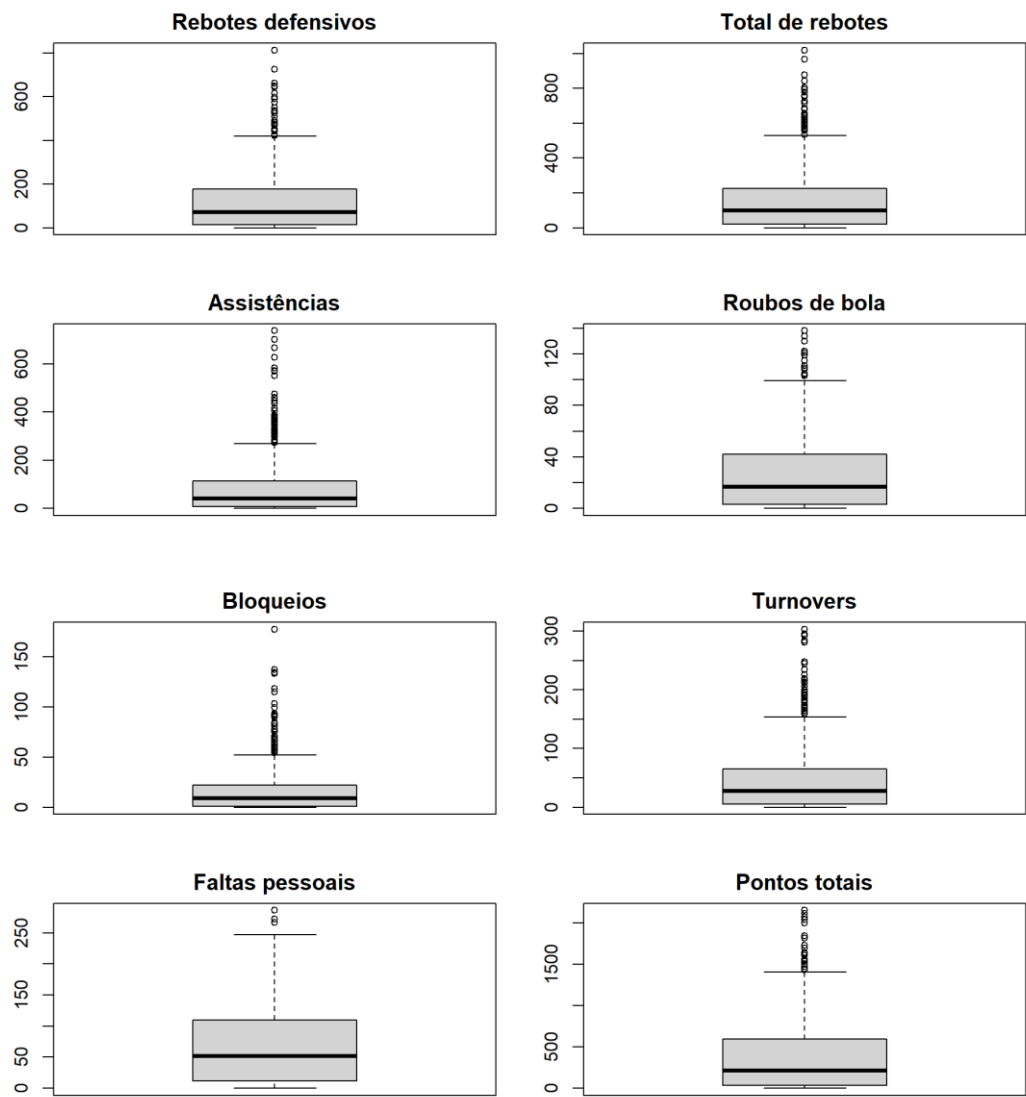
8.4.2 Análise gráfica das variáveis da temporada 2021-2022

Nesta seção, as variáveis estatísticas da temporada 2021-22 da NBA são analisadas por meio de gráficos exploratórios, incluindo boxplots, histogramas e correlogramas. Os boxplots evidenciam diferenças por posição, os histogramas mostram a concentração dos valores ao longo das distribuições, e os correlogramas apontam as inter-relações entre os principais indicadores. Essa abordagem gráfica permite uma leitura mais intuitiva das características do conjunto de dados.

Figura 15 - Boxplots das variáveis 21-22

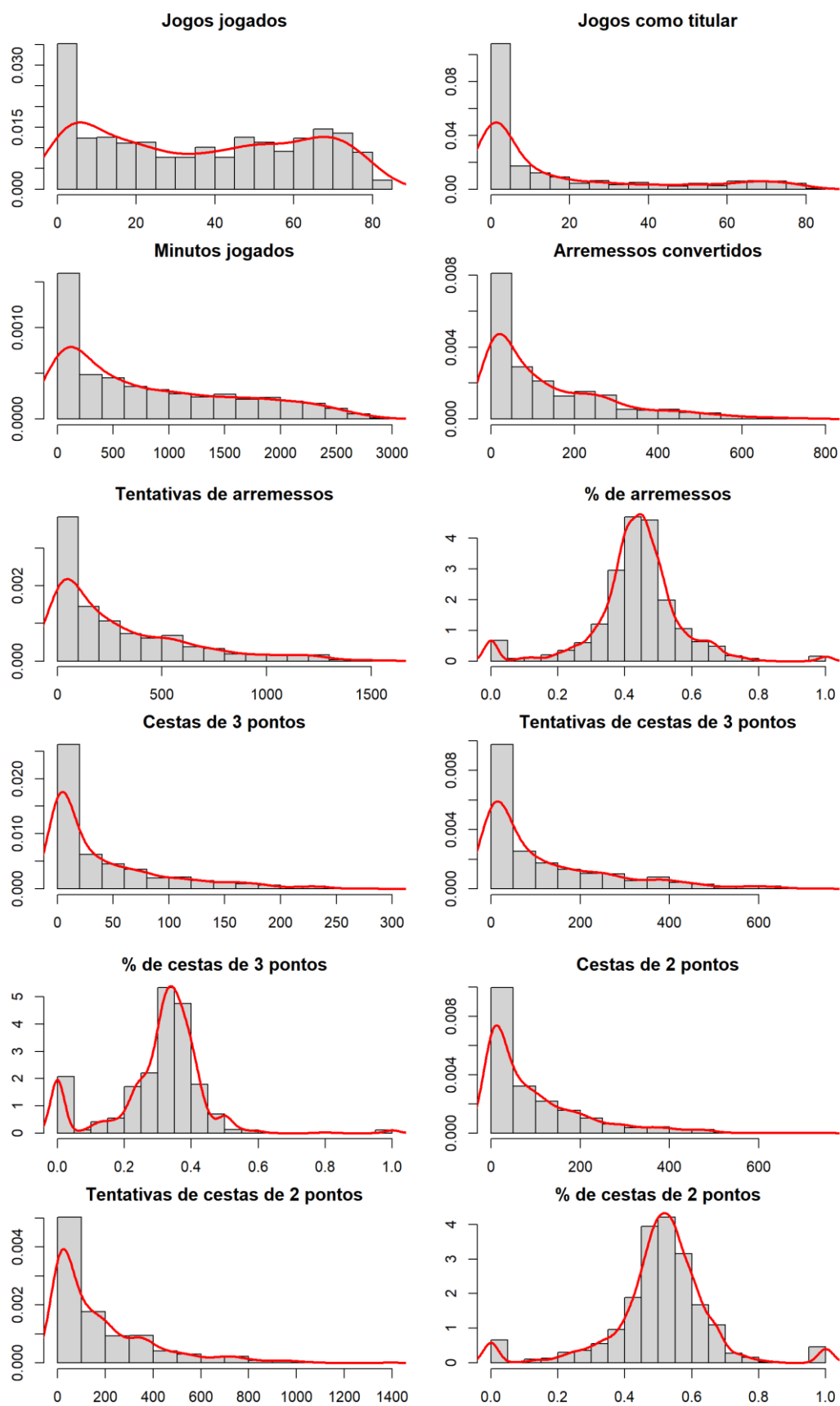


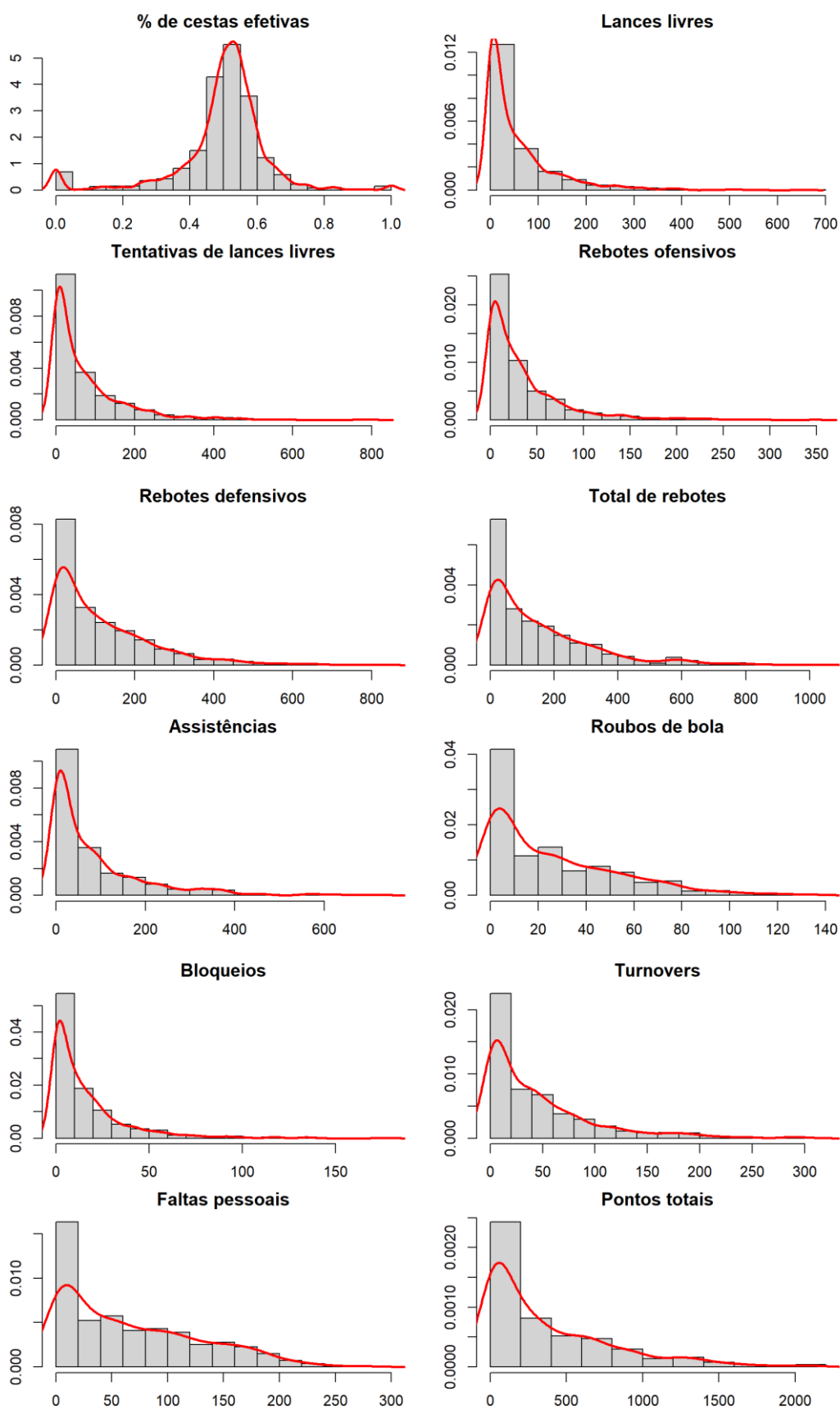




Fonte: elaborado pelo autor (2025)

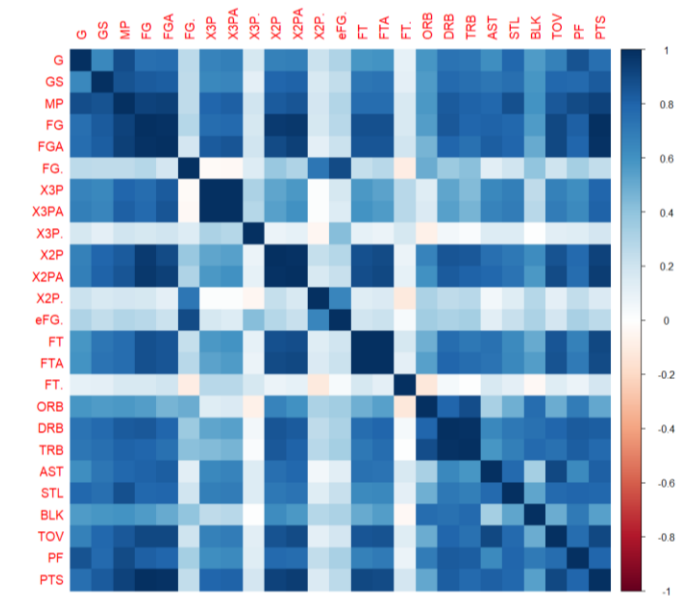
Figura 16 - Histograma das variáveis 21-22





Fonte: elaborado pelo autor (2025)

Figura 17 - Correlograma das variáveis 21-22



Fonte: elaborado pelo autor (2025)

Através da análise dos gráficos apresentados, observa-se que as variáveis de participação em jogos revelam padrões distintos: Jogos Jogados mostra distribuição com mediana em torno de 37 jogos, enquanto Jogos como Titular apresenta forte assimetria à direita, evidenciando a hierarquia nas equipes.

Os Minutos Jogados exibem distribuição assimétrica à direita, com mediana em torno de 577 minutos e outliers ultrapassando 2000 minutos, destacando a concentração de tempo em jogadores específicos.

As variáveis de volume ofensivo (Tentativas de Arremessos, Cestas de 3 e 2 Pontos) apresentam distribuições fortemente assimétricas à direita, com poucos jogadores concentrando grande parte das ações ofensivas. Em contraste, as métricas de eficiência (Porcentagem de Arremessos, Cestas de 3 e 2 Pontos) mostram distribuições mais simétricas (médias entre 30-51%), sugerindo homogeneidade na eficiência independente da posição.

Rebotes (Ofensivos, Defensivos, Totais) revelam assimetria à direita, com medianas de 20, 74 e 100 respectivamente, e outliers chegando a 1019 rebotes totais, refletindo a especialização de pivôs e alas-pivôs. Assistências e Turnovers seguem padrão similar, concentrados em jogadores com maior posse de bola, com outliers de assistências ultrapassando 700.

Estatísticas defensivas como Bloqueios e Roubos de Bola apresentam forte assimetria à direita (medianas 9 e 17), enquanto Faltas Pessoais mostram distribuição mais homogênea (mediana 52).

Em síntese, a análise da temporada 2021-2022 revela diversidade de perfis na NBA, com clara especialização em determinados fundamentos, distribuições assimétricas para variáveis de volume e mais simétricas para eficiência, fornecendo insights valiosos para a criação de rankings por posição e geral.

8.4.3 Análise de componentes principais 2021-2022

Para a temporada 2021–22, após a aplicação dos critérios mínimos de participação — pelo menos 20 jogos e 300 minutos totais — a base de dados foi composta por 498 jogadores. Esse filtro garante que apenas atletas com participação significativa ao longo da temporada sejam considerados, evitando que desempenhos pontuais (outliers com poucos minutos) distorçam os resultados das análises multivariadas.

A PCA (Análise de Componentes Principais) foi realizada separadamente para cada posição (PG, SG, SF, PF, C), utilizando um conjunto de variáveis relevantes para o papel específico de cada função em quadra. Antes de aplicar a PCA, avaliamos as medidas de adequação da matriz de correlações.

- O KMO global para a posição PG foi de 0.893, indicando boa adequação amostral. Os valores de KMO por variável para as demais posições também foram superiores a 0.7, reforçando a adequação da amostra para a análise fatorial.
 - PG: 0.893
 - SG: 0.839
 - SF: 0.827
 - PF: 0.772
 - C: 0.849
- O teste de esfericidade de Bartlett foi altamente significativo para todas as posições ($p < 10^{-192}$), confirmando que as variáveis têm correlações suficientes para justificar a análise multivariada.

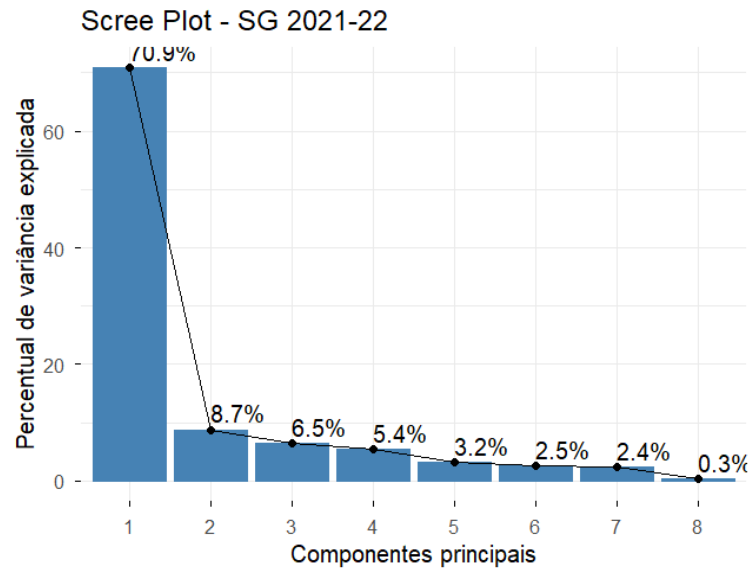
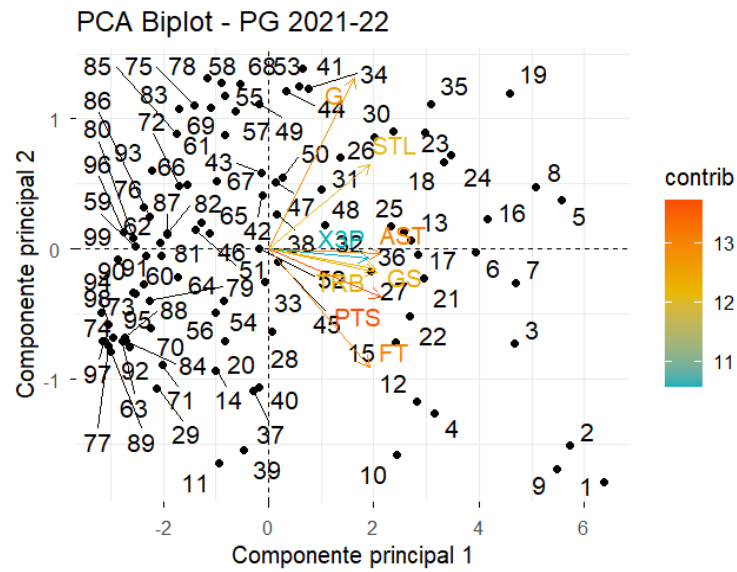
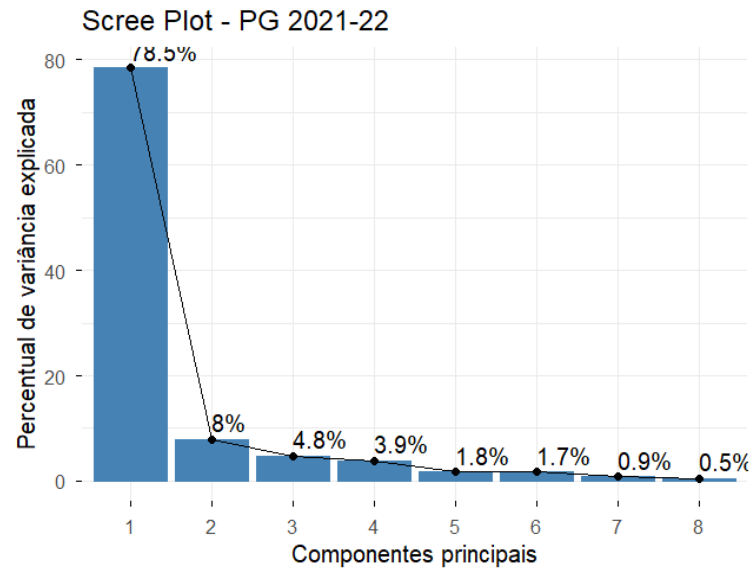
A variância explicada pelos três primeiros componentes principais foi bastante elevada em todas as posições, mostrando que a redução de dimensionalidade foi eficaz em capturar a maior parte da informação. Para os armadores (PG), por exemplo, a variância explicada pelos

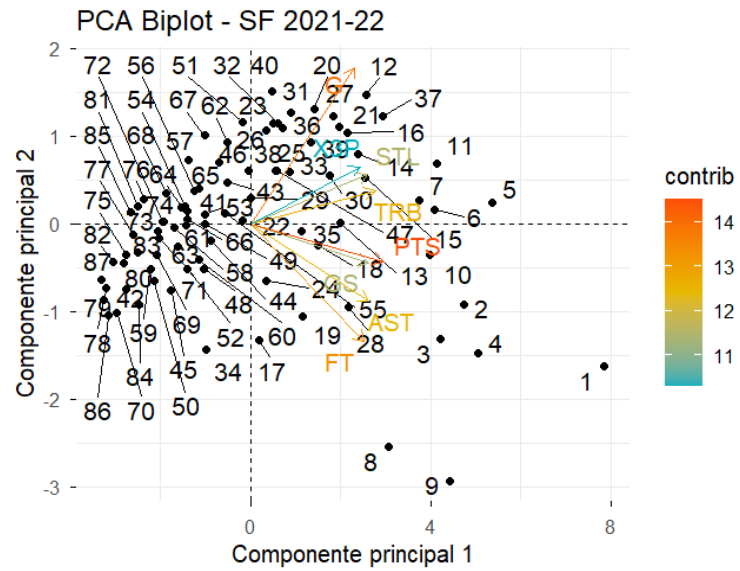
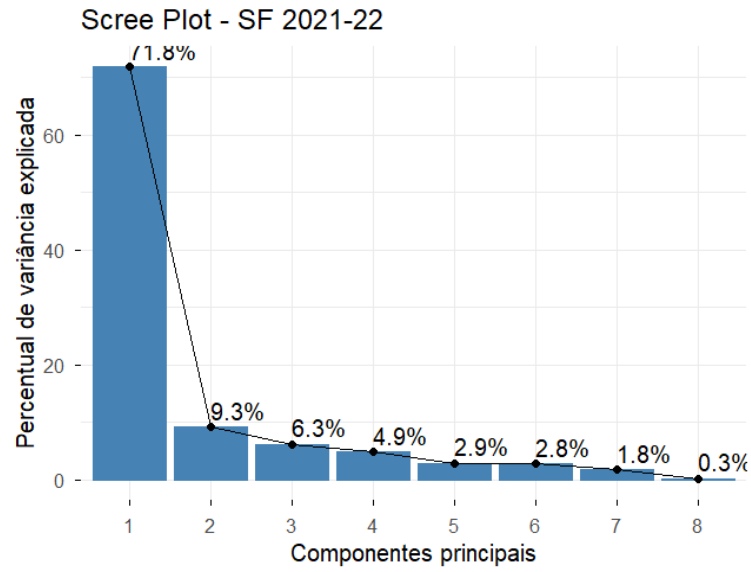
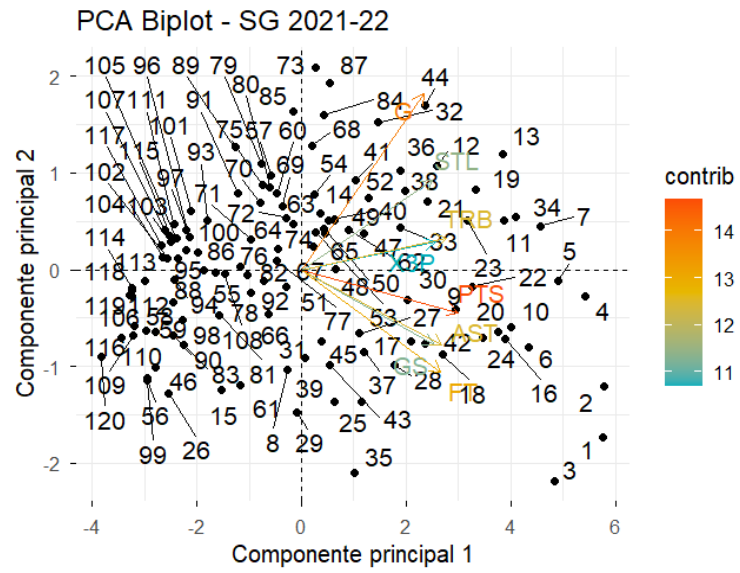
3 primeiros PCs foi de 91,26%. Esse alto valor demonstra que a maior parte da variação nos dados pode ser explicada por esses três componentes.

Para visualizar esse comportamento, o screeplot gerado para os armadores mostra claramente que os primeiros componentes capturam a maior parte da variação, com uma queda acentuada após o terceiro componente, confirmando que a escolha de três componentes principais é a mais adequada.

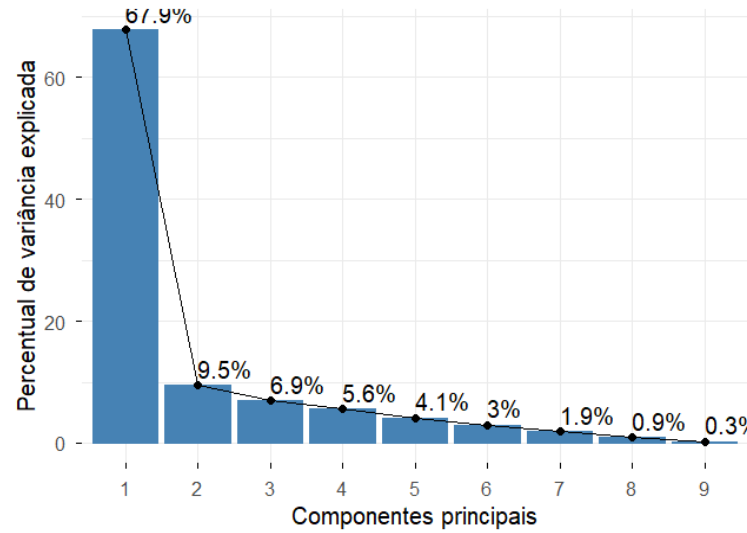
Com base nos componentes principais, as variáveis relacionadas à organização do jogo e volume (como G, GS, AST, TRB, STL) e à capacidade de pontuar e arremessar de fora (como PTS, X3P, FT) se destacam para os armadores (PG). Para os alas-armadores (SG) e alas (SF), os componentes principais refletem a pontuação, eficiência de arremessos e produção geral, enquanto para as posições de alas-pivôs (PF) e pivôs (C), os componentes destacam a pontuação próxima à cesta, rebotes ofensivos e defensivos e toques (BLK), além da participação nos jogos (G, GS).

Figura 18 - Scree Plots e Biplots por posição da temporada 2021-2022

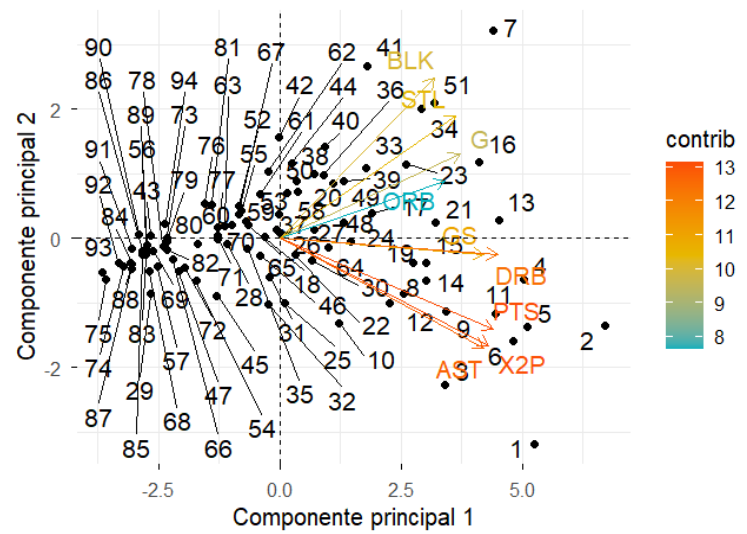


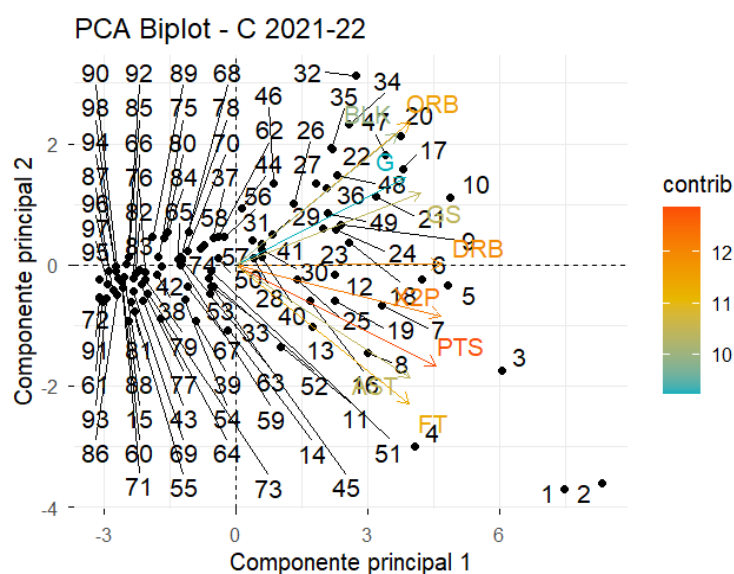
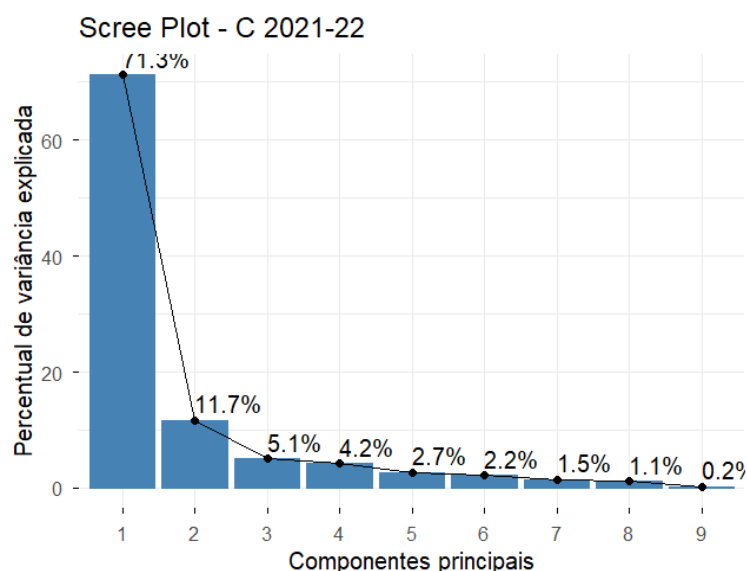


Scree Plot - PF 2021-22



PCA Biplot - PF 2021-22





Fonte: elaborado pelo autor.

8.4.4 Clusterização por k-médias 2021-2022

Com os escores dos três primeiros componentes principais, foi aplicado o algoritmo K-means para identificar perfis de jogadores dentro de cada posição. O número ótimo de clusters foi determinado pelo índice de silhueta, que sugeriu $k = 2$ em todas as posições.

As silhuetas médias para cada posição indicam que os clusters são moderadamente separados, com valores variando de 0.48 a 0.57, o que é esperado em dados esportivos, onde os jogadores podem ter características parcialmente sobrepostas. Os clusters refletem a diferenciação entre jogadores com maior protagonismo ofensivo e maior carga de minutos jogados (Cluster 1) e jogadores com contribuições mais modestas (Cluster 2).

Um gráfico de biplot da clusterização mostra claramente como os jogadores se posicionam nos dois clusters com base nos componentes principais. Para os armadores (PG), por exemplo, o Cluster 1 agrupa jogadores com maior volume de jogo e impacto ofensivo, enquanto o Cluster 2 é formado por jogadores mais coadjuvantes, com menores contribuições em termos de pontos e assistências. Esse gráfico pode ser inserido logo após a análise de K-means para ilustrar visualmente a separação entre os clusters.

8.4.5 Análise fatorial exploratória por posição 2021-2022

Após a PCA e o K-means, foi aplicada a Análise Fatorial Exploratória (AFE) em cada posição. A AFE foi realizada utilizando:

- Método de estimação: máxima verossimilhança ($fm = "ml"$).
- Rotação oblíqua: oblimin, permitindo a correlação entre fatores.
- Número de fatores: definido por análise paralela ($fa.parallel$).

Armadores (PG):

- Análise paralela sugeriu 2 fatores.
- Comunalidades (h^2) variaram de 0.456 (G) a 0.93 (PTS), indicando que a maior parte da variância dos indicadores é bem explicada pelo modelo fatorial.
- Fatores:
 1. Fator 1 – Eficiência de arremessos e pontuação (altas cargas em X3P, PTS, FT).
 2. Fator 2 – Volume e contribuição all-around (altas cargas em AST, TRB, STL, além de G e GS).

A variância explicada cumulativa foi de 76%, mostrando que a estrutura latente explica a maior parte da heterogeneidade entre os jogadores.

Alas-armadores (SG):

- KMO global: 0.839; Teste de Bartlett altamente significativo.
- Análise paralela sugeriu 2 fatores.
- Fatores:
 1. Fator 1 – Eficiência de lances livres e pontuação (FT, PTS, AST, GS).
 2. Fator 2 – Volume ofensivo e contribuição estatística de perímetro (X3P, G, TRB, STL).

Alas (SF):

- KMO global: 0.827; Análise paralela sugeriu 1 fator.

- A estrutura é um fator geral de desempenho, refletindo pontuação, minutos como titular, eficiência de arremesso, rebotes, assistências e roubos de bola.

Alas-pivôs (PF):

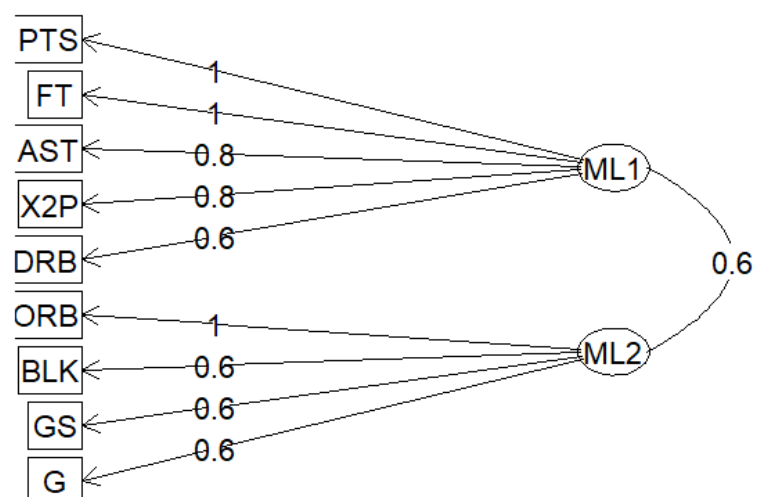
- KMO global: 0.772; Análise paralela sugeriu 1 fator.
- O fator reflete principalmente rebotes defensivos, pontuação, aproveitamento em dois pontos (X2P), tocos e rebotes ofensivos, com um foco nas variáveis defensivas.

Pivôs (C):

- KMO global: 0.849; Análise paralela sugeriu 2 fatores.
- Fatores:
 1. Fator 1 – Pontuação e eficiência ofensiva do garrafão (altas cargas em PTS, FT, AST, X2P, DRB, GS).
 2. Fator 2 – Presença física e proteção de aro (altas cargas em ORB, G, BLK, e contribuição em rebotes).

Figura 19 - Diagramas de fatores e Análises paralelas de cada posição 2021-2022

Diagrama de Fatores - C 2021-22



Análise Paralela - C 2021-22

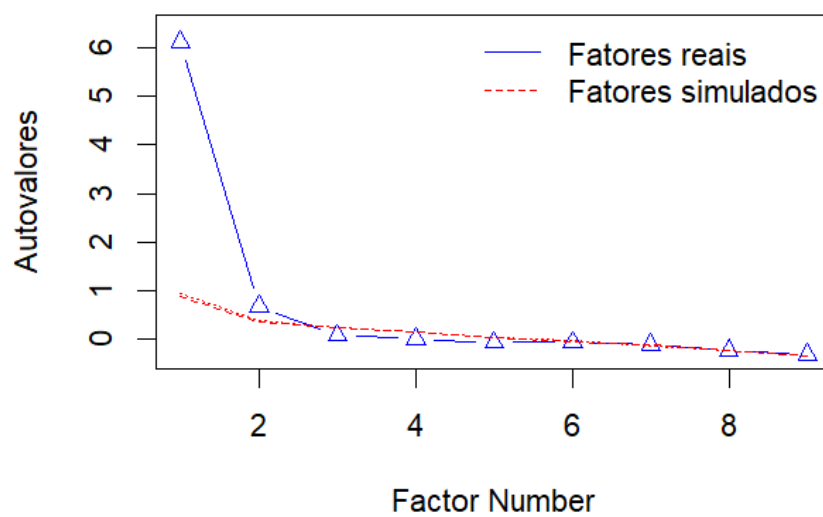
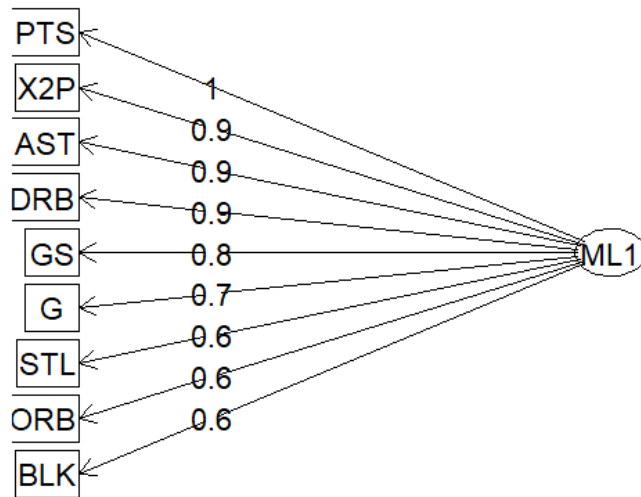


Diagrama de Fatores - PF 2021-22



Análise Paralela - PF 2021-22

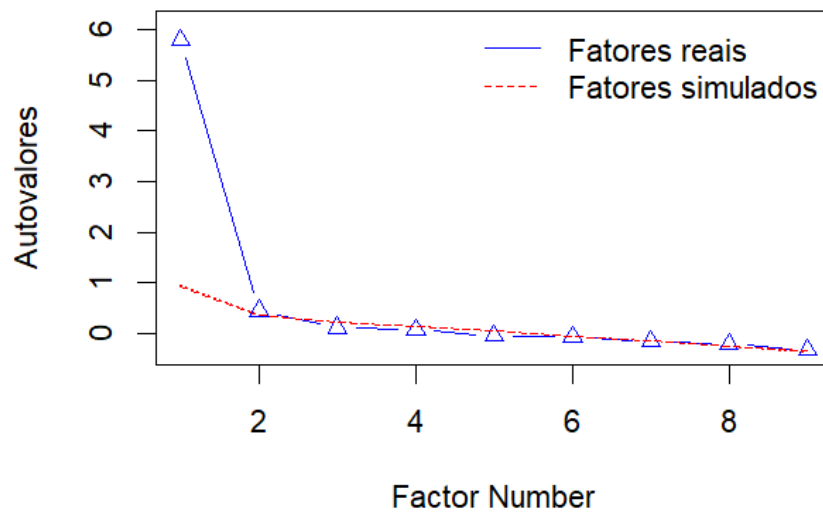
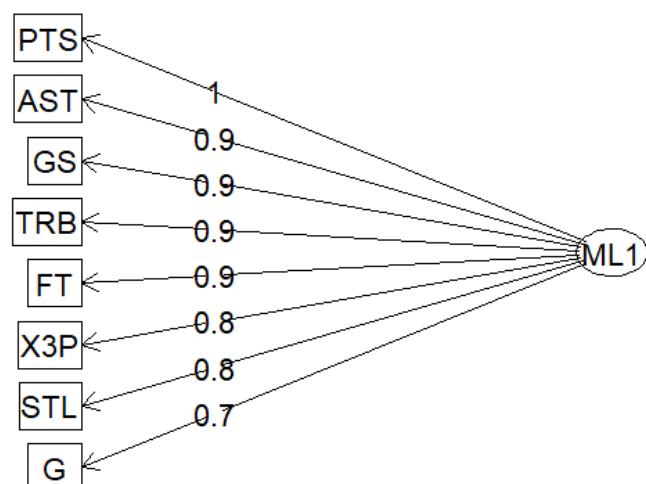


Diagrama de Fatores - PG 2021-22



Análise Paralela - PG 2021-22

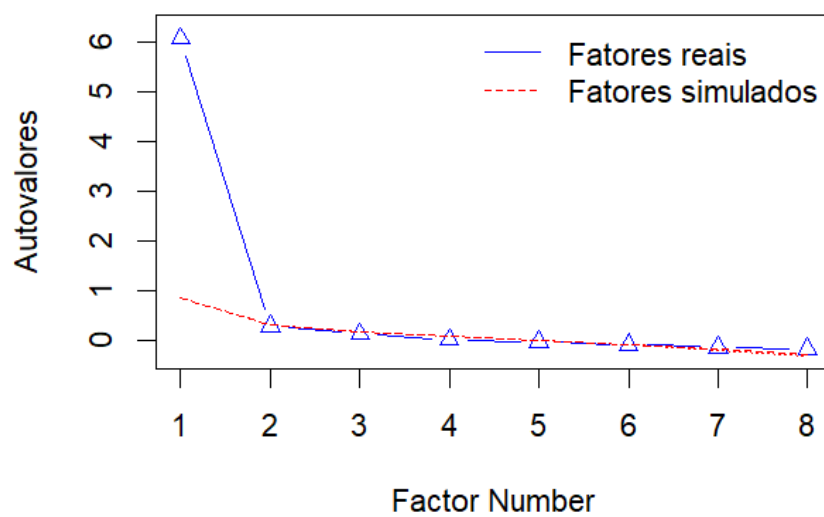
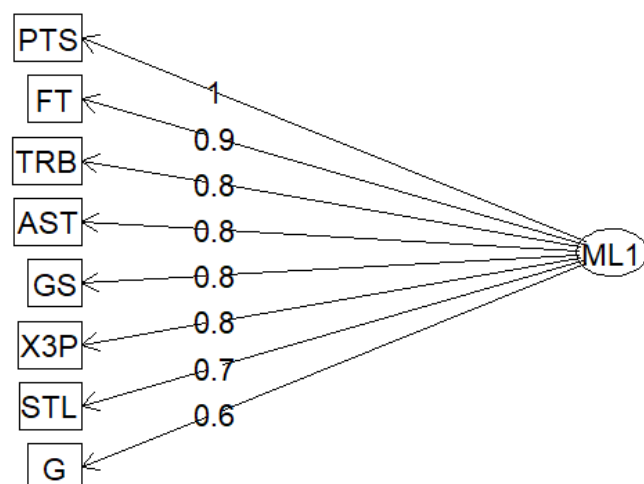


Diagrama de Fatores - SF 2021-22



Análise Paralela - SF 2021-22

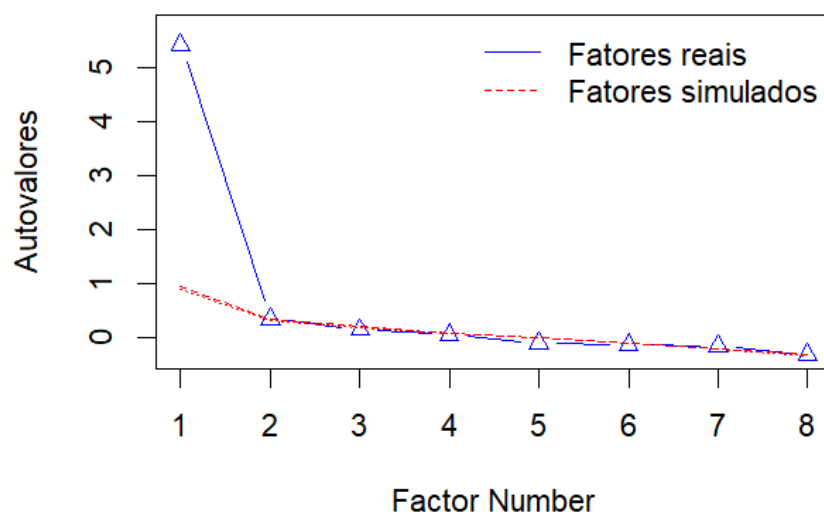
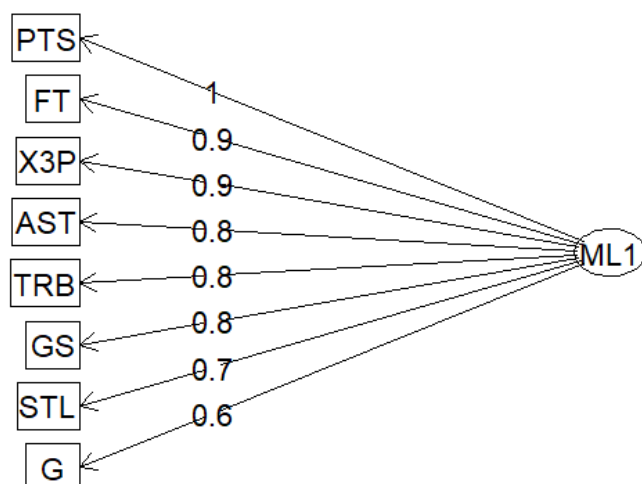
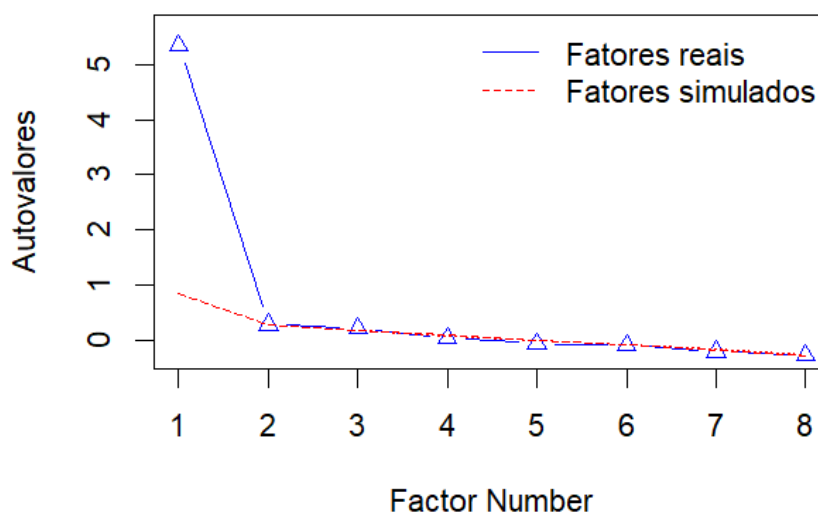


Diagrama de Fatores - SG 2021-22



Análise Paralela - SG 2021-22



Fonte: elaborado pelo autor.

8.4.6 Construção dos rankings por posição 2021-2022

A partir dos resultados da AFE por posição, os rankings foram construídos com duas etapas principais:

1. Cálculo dos escores fatoriais usando o método de regressão.
2. Combinação dos fatores utilizando os pesos proporcionais à variância explicada.

As equações finais dos rankings para cada posição foram:

- PG: A equação final do índice para armadores foi baseada nas variáveis mais relevantes para o papel dos PGs, com maior ênfase em PTS, AST, FT e TRB.
- SG: Para os alas-armadores, a equação incluiu PTS, FT, X3P, TRB, e AST.
- SF: O fator único para os alas foi composto por PTS, GS, FT, AST, X3P.
- PF e C: As equações para essas posições destacaram rebotes, toques, pontuação, e eficiência de arremesso.

Tabela 16 – Top 10 C 2021-22

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Nikola Jokic	C	74	74	2476	27,08	13,77	7,89	100,00	2
2	Joel Embiid	C	68	68	2297	30,57	11,71	4,18	92,70	2
3	Karl-Anthony Towns	C	74	74	2476	24,57	9,82	3,64	87,13	2
4	Jonas Valančiūnas	C	74	74	2240	17,76	11,39	2,57	74,89	2
5	Rudy Gobert	C	66	66	2120	15,56	14,67	1,09	68,26	2
6	LeBron James	C	56	56	2084	30,27	8,20	6,23	67,00	2
7	Nikola Vucevic	C	73	73	2418	17,64	11,01	3,23	66,21	2
8	Clint Capela	C	74	73	2042	11,08	11,85	1,24	64,31	2
9	Jakob Poeltl	C	68	67	1970	13,46	9,32	2,78	64,03	2
10	Steven Adams	C	76	75	1999	6,95	10,00	3,37	57,93	2

Fonte: elaborado pelo autor.

Tabela 17 – Top 10 PF 2021-22

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Giannis Antetokounmpo	PF	67	67	2204	29,88	11,61	5,79	100,00	1
2	DeMar DeRozan	PF	76	76	2743	27,87	5,16	4,92	97,81	1
3	Pascal Siakam	PF	68	68	2578	22,81	8,53	5,29	81,79	1
4	Miles Bridges	PF	80	80	2837	20,16	6,99	3,75	78,96	1
5	Julius Randle	PF	72	72	2544	20,10	9,94	5,14	77,26	1

6	Kevin Durant	PF	55	55	2047	29,87	7,40	6,38	75,23	1
7	Domantas Sabonis	PF	62	61	2136	18,89	12,13	5,21	70,14	1
8	Scottie Barnes	PF	74	74	2617	15,32	7,53	3,46	65,45	1
9	Tobias Harris	PF	73	73	2543	17,22	6,79	3,45	65,39	1
10	Evan Mobley	PF	69	69	2331	14,99	8,26	2,52	61,46	1

Fonte: elaborado pelo autor.

Tabela 18 – Top 10 PG 2021-22

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Trae Young	PG	76	76	2652	28,36	3,74	9,70	100,00	1
2	Luka Dončić	PG	65	65	2301	28,42	9,12	8,74	91,67	1
3	James Harden	PG	65	65	2419	22,03	7,69	10,26	86,20	1
4	LaMelo Ball	PG	75	75	2422	20,11	6,68	7,61	84,40	1
5	Dejounte Murray	PG	68	68	2366	21,15	8,28	9,22	82,41	1
6	Russell Westbrook	PG	78	78	2678	18,47	7,44	7,05	79,98	1
7	Stephen Curry	PG	64	64	2211	25,47	5,23	6,31	76,44	1
8	Tyrese Haliburton	PG	77	77	2695	15,34	4,04	8,16	73,45	1
9	Darius Garland	PG	68	68	2430	21,66	3,29	8,57	72,47	1
10	Fred VanVleet	PG	65	65	2462	20,31	4,45	6,68	69,06	1

Fonte: elaborado pelo autor.

Tabela 19 – Top 10 SF 2021-22

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Jayson Tatum	SF	76	76	2731	26,92	8,01	4,39	100,00	1
2	Jaylen Brown	SF	66	66	2220	23,62	6,12	3,50	74,66	1
3	RJ Barrett	SF	70	70	2417	20,03	5,83	2,97	67,04	1
4	Khris Middleton	SF	66	66	2141	20,08	5,39	5,42	63,74	1
5	Saddiq Bey	SF	82	82	2704	16,11	5,38	2,84	63,59	1
6	Keldon Johnson	SF	75	74	2392	17,04	6,09	2,11	60,77	1

7	Andrew Wiggins	SF	73	73	2329	17,21	4,45	2,21	59,40	1
8	Brandon Ingram	SF	55	55	1869	22,65	5,82	5,58	58,95	1
9	Jimmy Butler	SF	57	57	1931	21,39	5,89	5,47	58,57	1
10	Franz Wagner	SF	79	79	2429	15,15	4,51	2,92	56,80	1

Fonte: elaborado pelo autor.

Tabela 20 – Top 10 SG 2021-22

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Devin Booker	SG	68	68	2345	26,79	5,03	4,84	100,00	2
2	Donovan Mitchell	SG	67	67	2266	25,87	4,21	5,34	95,16	2
3	Zach LaVine	SG	67	67	2328	24,40	4,60	4,52	89,54	2
4	Anthony Edwards	SG	72	72	2466	21,29	4,76	3,82	83,98	2
5	Terry Rozier	SG	73	73	2458	19,29	4,32	4,47	76,83	2
6	Jordan Poole	SG	76	51	2283	18,49	3,42	4,00	76,67	2
7	Desmond Bane	SG	76	76	2266	18,21	4,39	2,74	75,25	2
8	Tyler Herro	SG	66	10	2151	20,71	4,98	3,98	73,63	2
9	Jalen Brunson	SG	79	61	2524	16,27	3,91	4,77	69,58	2
10	Gary Trent Jr.	SG	70	69	2448	18,33	2,73	2,01	69,38	2

Fonte: elaborado pelo autor.

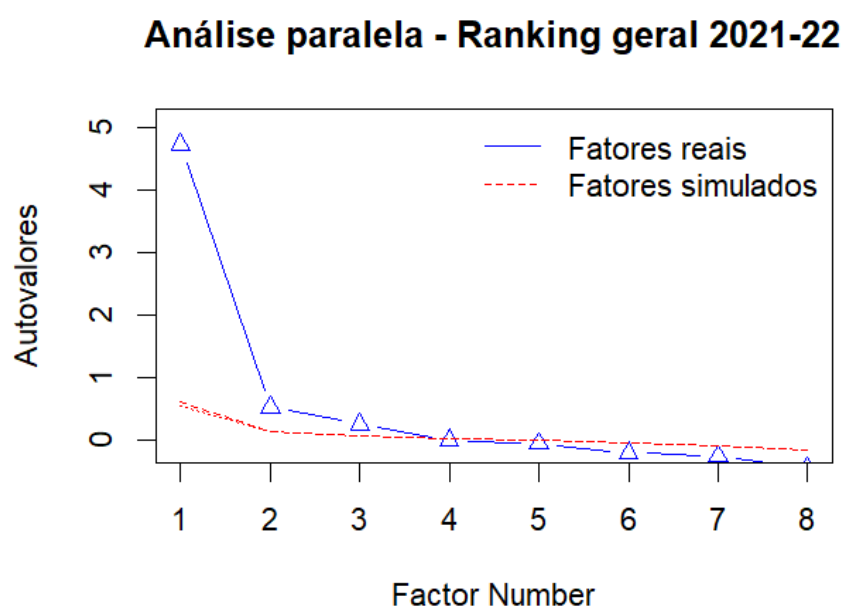
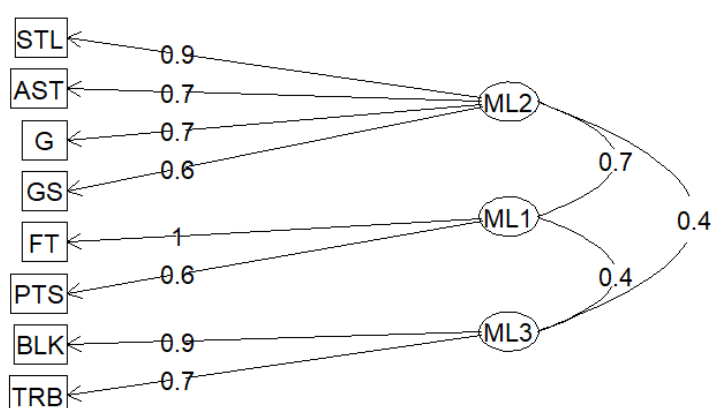
8.4.7 Ranking geral 2021-2022

Após a construção dos rankings por posição, um ranking geral foi criado, combinando todas as posições. A análise fatorial para o ranking geral sugeriu 3 fatores, com 78% de variância explicada cumulativamente.

A equação final do ranking geral incluiu variáveis como PTS, GS, AST, TRB, STL, BLK, FT, refletindo os aspectos mais críticos para o desempenho geral dos jogadores.

O MVP da temporada 2021-2022 foi Nikola Jokic, que ocupou a 2ª posição no ranking geral, com um score normalizado de 96.83. Isso valida o modelo de análise estatística, já que o jogador mais reconhecido oficialmente aparece próximo do topo do ranking, reforçando a eficácia do índice de desempenho criado.

Figura 20 - Análise paralela e diagrama de fatores Geral 2021-2022

**Fatores - Ranking Geral 2021-22**

Fonte: elaborado pelo autor.

Tabela 21 – Top 10 SG 2021-22

rank_geral	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_geral_norm	cluster
1	Joel Embiid	C	68	68	2297	30,57	11,71	4,18	100,00	2
2	Nikola Jokic	C	74	74	2476	27,08	13,77	7,89	96,83	2
3	Giannis Antetokounmpo	PF	67	67	2204	29,88	11,61	5,79	94,04	1
4	Trae Young	PG	76	76	2652	28,36	3,74	9,70	82,19	1

5	Karl-Anthony Towns	C	74	74	2476	24,57	9,82	3,64	81,91	2
6	Jayson Tatum	SF	76	76	2731	26,92	8,01	4,39	81,40	1
7	DeMar DeRozan	PF	76	76	2743	27,87	5,16	4,92	80,85	1
8	James Harden	PG	65	65	2419	22,03	7,69	10,26	77,76	1
9	Luka Dončić	PG	65	65	2301	28,42	9,12	8,74	77,21	1
10	Jaren Jackson Jr.	PF	78	78	2126	16,31	5,82	1,10	71,76	1

Fonte: elaborado pelo autor.

8.5 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2022-2023

A análise da temporada 2022-23 tem como foco explorar as principais métricas individuais dos jogadores da NBA, destacando suas características estatísticas gerais. A partir da descrição de variáveis como pontuação, assistências, arremessos convertidos e participação defensiva, busca-se identificar o grau de variabilidade presente na base de dados e compreender o perfil técnico dos atletas. Este diagnóstico inicial é essencial para fundamentar as análises multivariadas posteriores.

8.5.1 Estatísticas das variáveis da temporada 2022-2023

A seguir, são apresentados os resultados descritivos das variáveis da temporada 2022-23, incluindo medidas como média, mediana, desvio padrão e extremos. Esta síntese estatística auxilia na identificação de tendências relevantes e na caracterização do desempenho dos jogadores durante a temporada, servindo como base para as análises analíticas subsequentes.

Tabela 22 - Estatísticas das variáveis 22-23

VARIÁVEL	MÉDIA	D.PADRÃO	MÍN	Q1	MEDIANA	Q3	MAX
G	43,3372	24,7291	1	22	45	65,5	83
GS	20,0692	25,7663	0	0	6	36,5	83
MP	984,4212	800,2363	1	266,5	797	1663,5	2963
FG	169,3873	169,1577	0	36	108	253,5	728

FGA	357,3770	350,7376	0	79	237	529	1559
FG.	0,4642	0,1127	0	0,415	0,454	0,506	1
X3P	50,7952	57,2180	0	4,5	30	77	301
X3PA	140,6067	151,7023	0	16,5	89	218	731
X3P.	0,3298	0,1281	0	0,297	0,349	0,389	1
X2P	118,5920	131,9451	0	20	65	182	662
X2PA	216,7702	237,1170	0	38	121	321,5	1213
X2P.	0,5326	0,1380	0	0,485	0,536	0,594	1
eFG.	0,5353	0,1086	0	0,496	0,537	0,579	1
FT	73,6494	98,3249	0	11	36	96	669
FTA	94,2798	120,8667	0	15	47	128	772
FT.	0,7526	0,1499	0	0,687	0,769	0,8445	1
ORB	42,1561	49,1871	0	9	27	57	274
DRB	133,8792	130,3787	0	32	94	196,5	744
TRB	176,0353	172,7937	0	45	126	260,5	973
AST	102,9705	122,3577	0	18	57	137,5	741
STL	29,6980	27,0790	0	7	22	45	128
BLK	18,7187	24,5847	0	4	11	24,5	193
TOV	54,2636	55,4331	0	12	36	81	300
PF	81,1944	64,0565	0	26	68	123	279
PTS	463,2194	471,4232	0	95,5	291	689	2225

Fonte: elaborado pelo autor (2025)

Observando a tabela de estatísticas das variáveis, diversas informações relevantes podem ser extraídas, oferecendo uma visão sobre as características dos jogadores da NBA na temporada 2022-2023:

a) Jogos e Participação: O número de jogos disputados apresenta média de 43,34 jogos por jogador (mínimo 1, máximo 83). Quanto aos jogos como titular, a mediana de apenas 6 jogos, contrastando com o máximo de 83, revela clara distinção entre titulares fixos e reservas, evidenciando a hierarquia nas equipes.

b) Minutos Jogados: Média de 984,42 minutos por temporada, variando de 1 a 2.963 minutos. A mediana de 797 minutos indica que metade dos jogadores tem participação moderada, enquanto um grupo seleto acumula tempo significativamente maior.

c) Eficiência Ofensiva: A porcentagem de arremessos de quadra tem média de 46,43% com mediana de 45,4%, sugerindo distribuição relativamente simétrica. Nas cestas de três pontos, a média de 32,98% com mediana de 34,9% demonstra eficiência considerável dos especialistas em arremessos de longa distância.

d) Volume Ofensivo: Os arremessos convertidos apresentam média de 169,39 cestas por temporada (máximo 728), enquanto as tentativas têm média de 357,38 (máximo 1.559). Nos pontos totais, a média de 463,22 contrasta com o máximo de 2.225, evidenciando o papel dominante dos principais pontuadores.

e) Rebotes: Média de 176,04 rebotes totais por temporada (máximo 973), com predominância dos defensivos (média 133,88) sobre os ofensivos (média 42,16). A mediana de rebotes totais (126) significativamente abaixo da média indica concentração em jogadores específicos, tipicamente pivôs e alas-pivôs.

f) Estatísticas Defensivas e de Criação: As assistências apresentam média de 102,97 (máximo 741), evidenciando especialização dos armadores. Roubo de bola (média 29,70) e bloqueios (média 18,72) mostram medianas abaixo das médias, indicando que poucos jogadores se destacam nestas estatísticas defensivas.

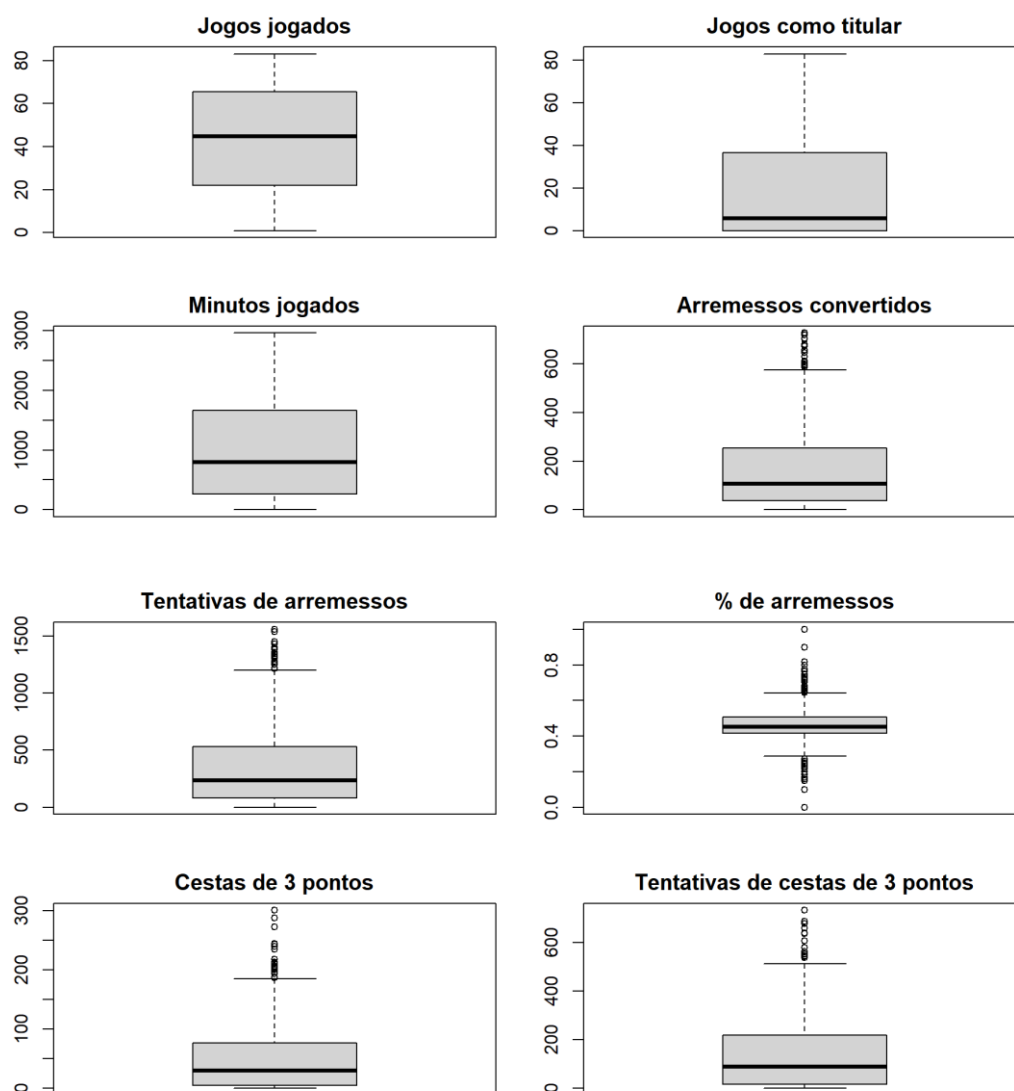
Esta análise revela a diversidade de perfis na NBA, com clara especialização em determinados fundamentos e distribuição assimétrica para variáveis de volume, contrastando com distribuições mais equilibradas para variáveis de eficiência.

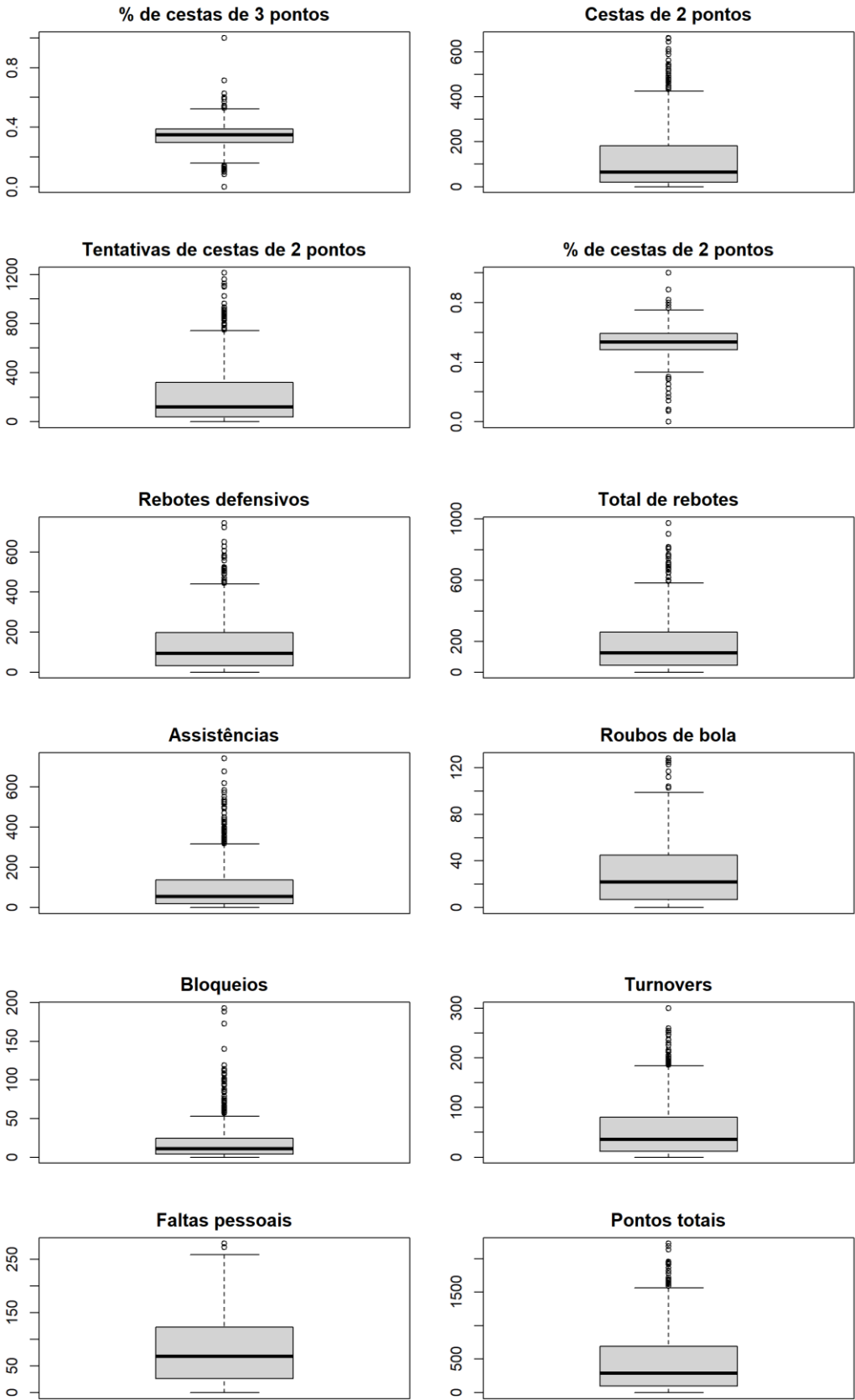
8.5.2 Análise gráfica das variáveis da temporada 2022-2023

As análises gráficas da temporada 2022-23 visam fornecer uma visão detalhada e visual do comportamento das variáveis estatísticas dos jogadores. A utilização de boxplots,

histogramas e correlogramas permite identificar padrões relevantes, como a presença de assimetrias, dispersões acentuadas e associações lineares entre os indicadores. Esses recursos gráficos reforçam as interpretações realizadas nas análises descritivas anteriores.

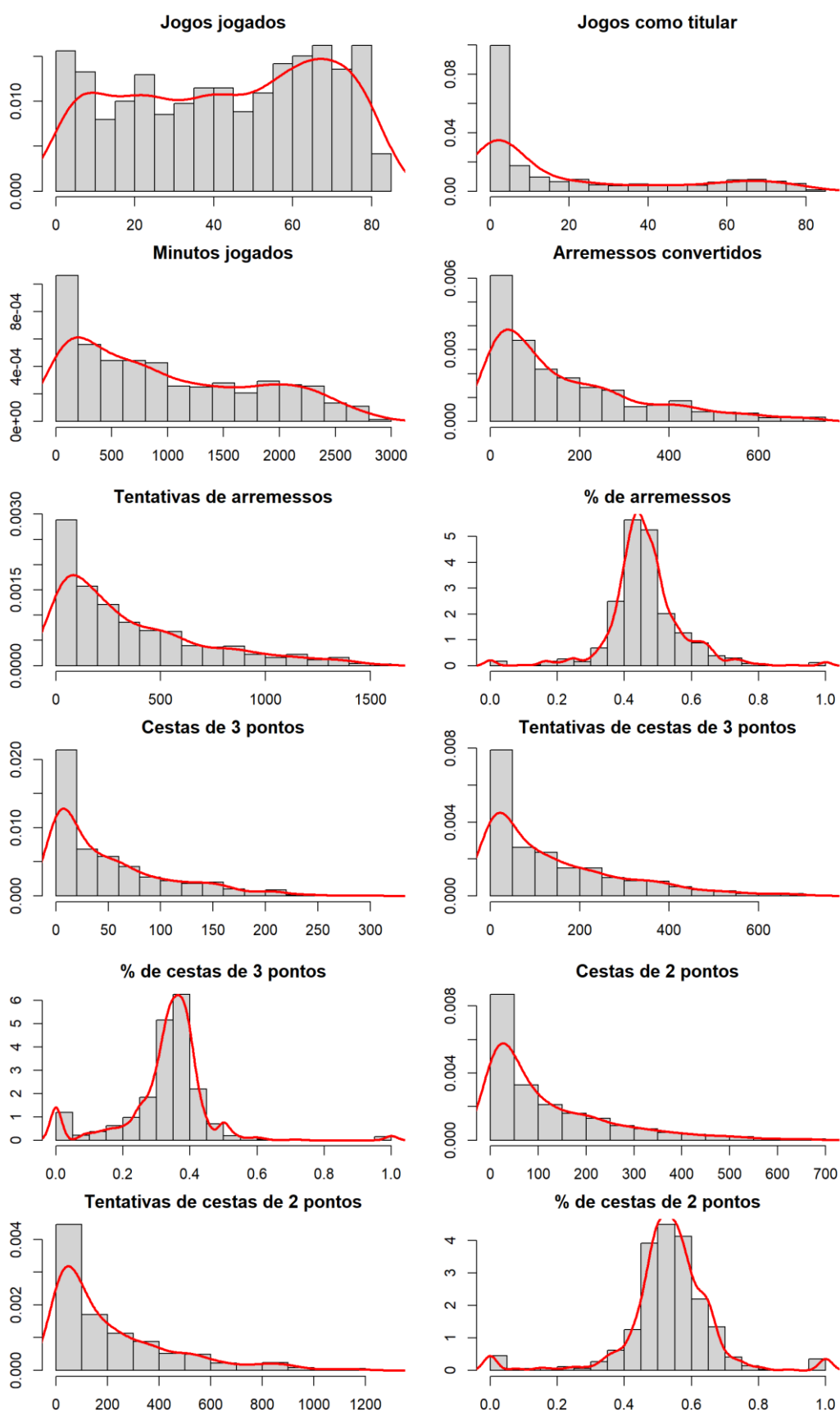
Figura 21 - Boxplots das variáveis 22-23

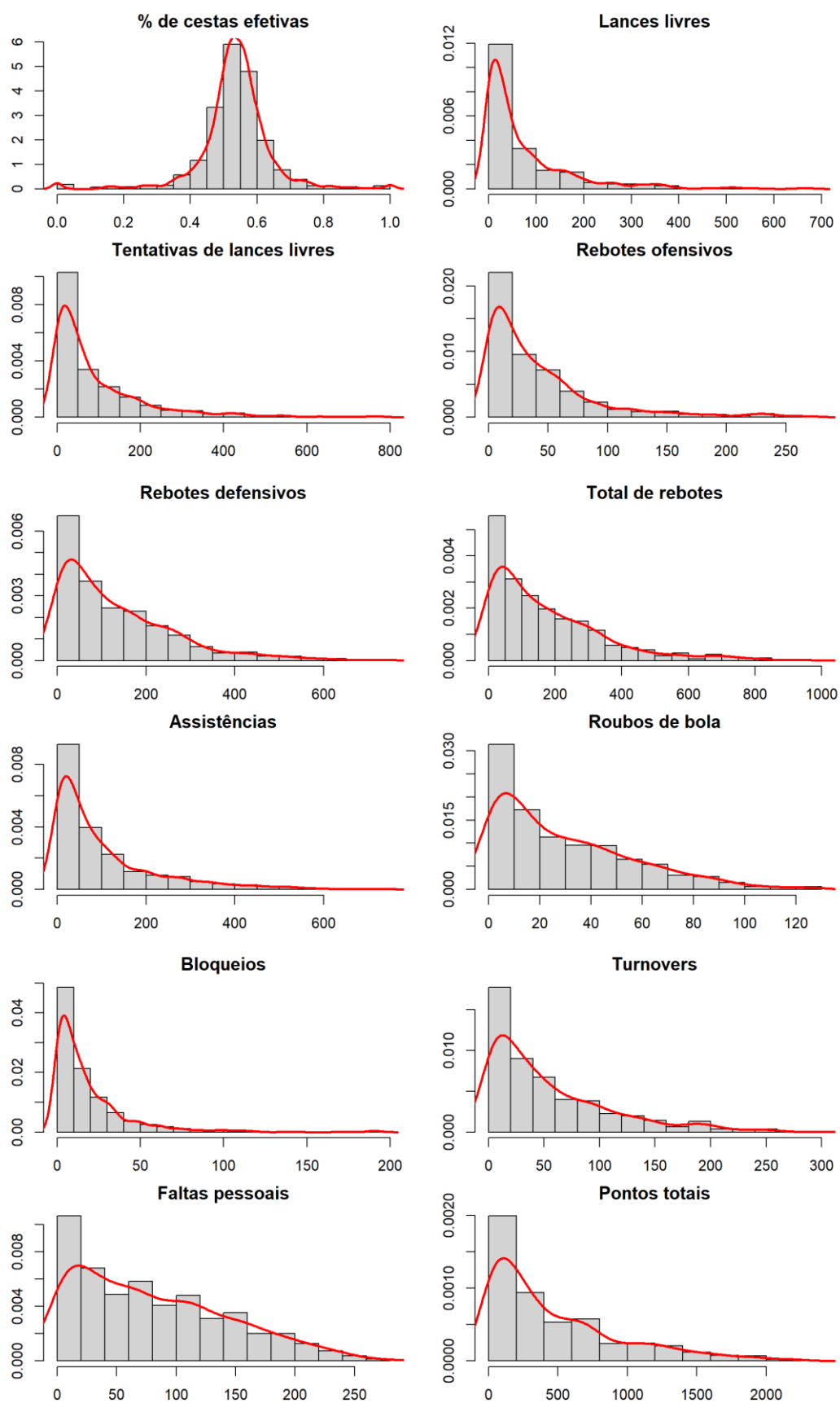




Fonte: elaborado pelo autor (2025)

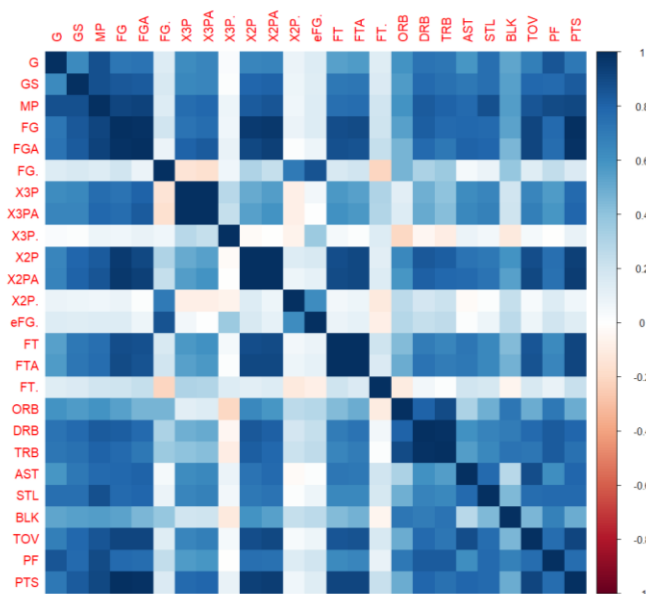
Figura 22 - Histogramas das variáveis 22-23





Fonte: elaborado pelo autor (2025)

Figura 23 - Correlograma das variáveis 22-23



Fonte: elaborado pelo autor (2025)

Através da análise dos gráficos apresentados, observa-se que as variáveis de participação em jogos revelam padrões distintos: Jogos Jogados mostra distribuição com mediana em torno de 45 jogos, enquanto Jogos como Titular apresenta forte assimetria à direita, evidenciando a hierarquia nas equipes.

Os Minutos Jogados exibem distribuição assimétrica à direita, com mediana em torno de 797 minutos e outliers ultrapassando 2900 minutos, destacando a concentração de tempo em jogadores específicos.

As variáveis de volume ofensivo (Tentativas de Arremessos, Cestas de 3 e 2 Pontos) apresentam distribuições fortemente assimétricas à direita, com poucos jogadores concentrando grande parte das ações ofensivas. Em contraste, as métricas de eficiência (Porcentagem de Arremessos, Cestas de 3 e 2 Pontos) mostram distribuições mais simétricas (médias entre 33-53%), sugerindo homogeneidade na eficiência independente da posição.

Rebotes (Ofensivos, Defensivos, Totais) revelam assimetria à direita, com medianas de 27, 94 e 126 respectivamente, e outliers chegando a 973 rebotes totais, refletindo a especialização de pivôs e alas-pivôs. Assistências e Turnovers seguem padrão similar, concentrados em jogadores com maior posse de bola, com outliers de assistências ultrapassando 740.

Estatísticas defensivas como Bloqueios e Roubos de Bola apresentam forte assimetria à direita (medianas 11 e 22), enquanto Faltas Pessoais mostram distribuição mais homogênea (mediana 68).

Em síntese, a análise da temporada 2022-2023 revela diversidade de perfis na NBA, com clara especialização em determinados fundamentos, distribuições assimétricas para variáveis de volume e mais simétricas para eficiência, fornecendo insights valiosos para a criação de rankings por posição e geral.

8.5.3 Análise de componentes principais 2022-2023

Para a temporada 2022–23, após a aplicação dos critérios mínimos de participação — pelo menos 20 jogos e 300 minutos totais — a base de dados foi composta por 485 jogadores. Esse filtro assegura que apenas jogadores com participação relevante ao longo da temporada sejam considerados, evitando que desempenhos esporádicos ou com poucos minutos impactem a análise.

A PCA (Análise de Componentes Principais) foi realizada separadamente para cada posição (PG, SG, SF, PF, C), utilizando um conjunto de variáveis relevantes para o papel de cada função em quadra. Antes de aplicar o PCA, foi feita a avaliação das medidas de adequação da matriz de correlações.

- O KMO global para a posição PG foi de 0.827, indicando boa adequação amostral, com valores de KMO para as outras posições também em níveis adequados:
 - PG: 0.827
 - SG: 0.866
 - SF: 0.837
 - PF: 0.742
 - C: 0.809
- O teste de esfericidade de Bartlett foi altamente significativo para todas as posições, com p-values abaixo de 10^{-130} , rejeitando a hipótese de matriz identidade e confirmando a existência de estrutura latente entre as variáveis.

A variância explicada pelos três primeiros componentes principais foi de 87.42% para os armadores (PG), o que indica que a redução de dimensionalidade foi muito eficaz. Isso é desejável para facilitar a visualização e análise dos dados, sem perder informações essenciais.

- Armadores (PG): 87.42%
- Alas-armadores (SG): 89.5%

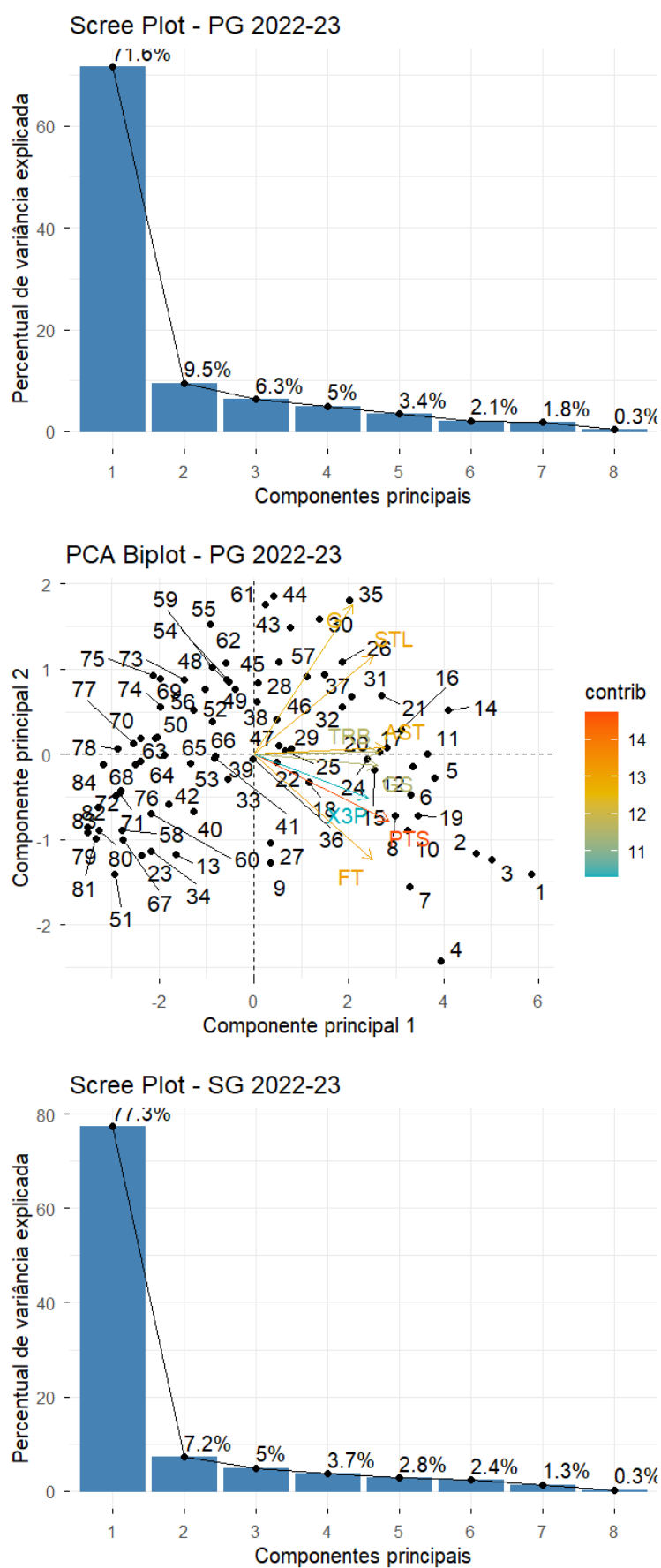
- Alas (SF): 89.3%
- Alas-pivôs (PF): 83.14%
- Pivôs (C): 89.16%

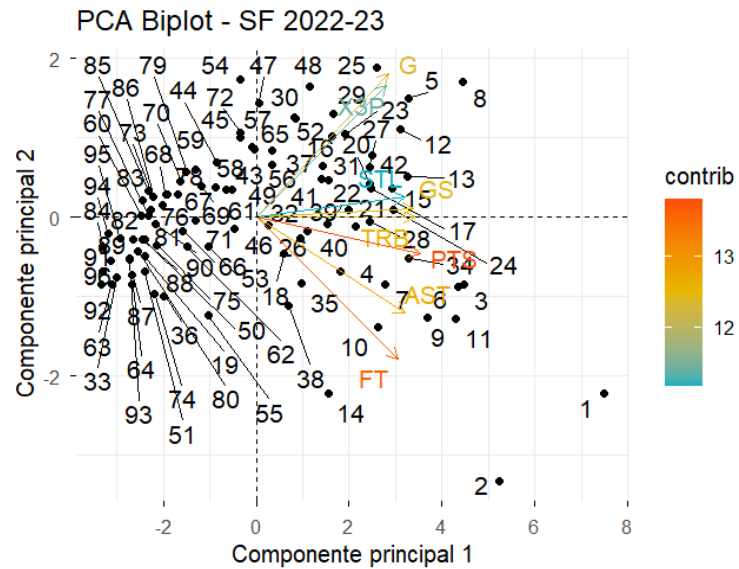
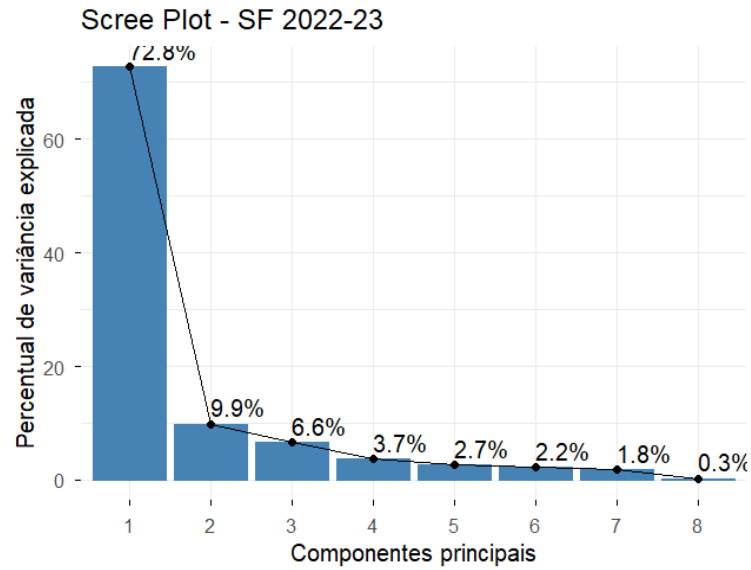
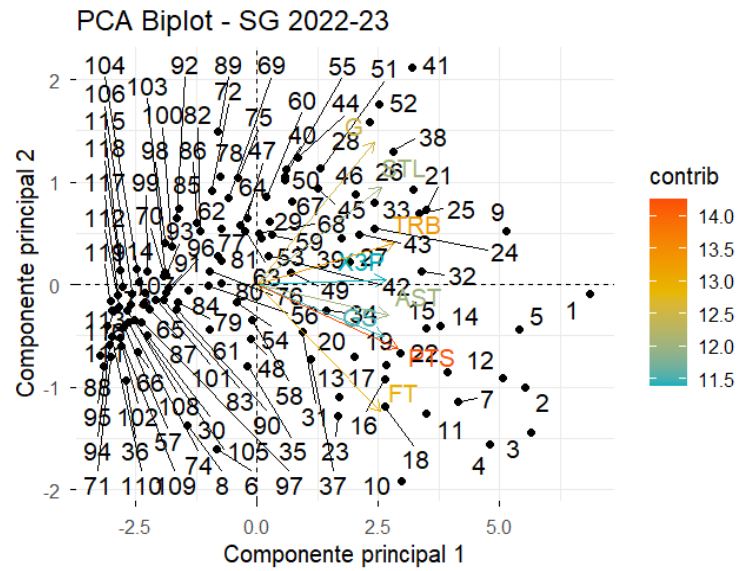
Esses percentuais indicam que, com a utilização de apenas três componentes principais, é possível capturar uma grande parte da variação nos dados, o que é ideal para fins de redução de dimensionalidade e visualização.

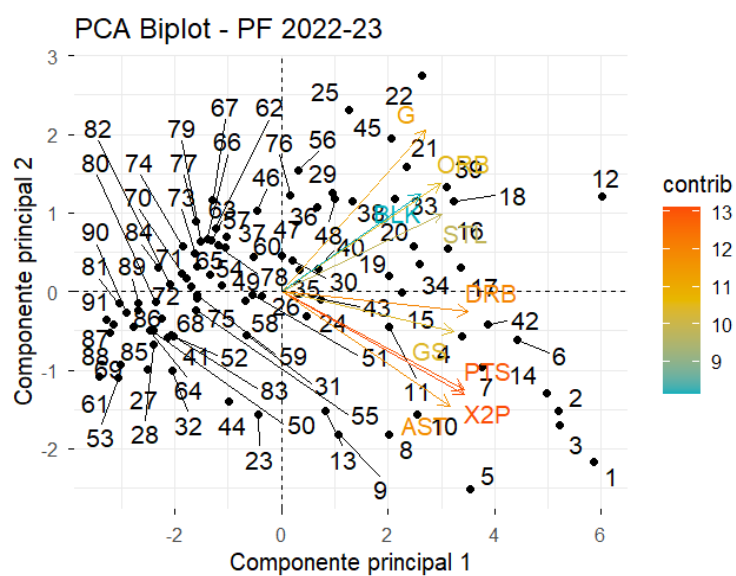
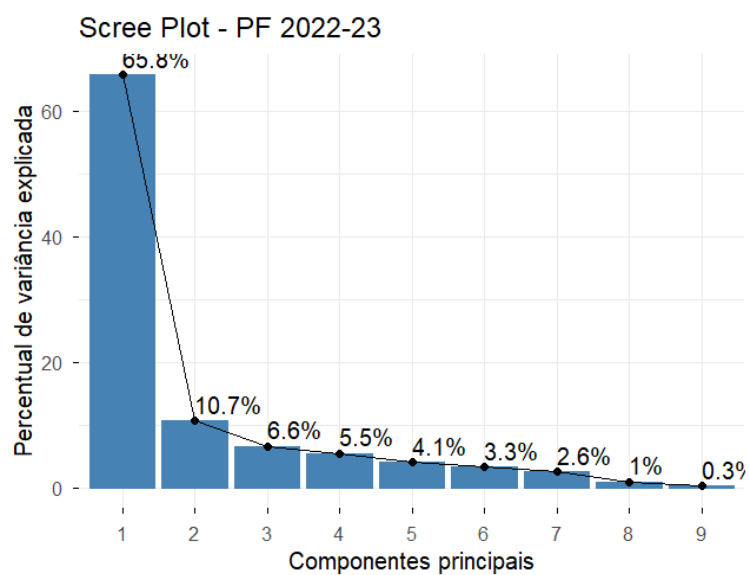
Interpretação Conceitual:

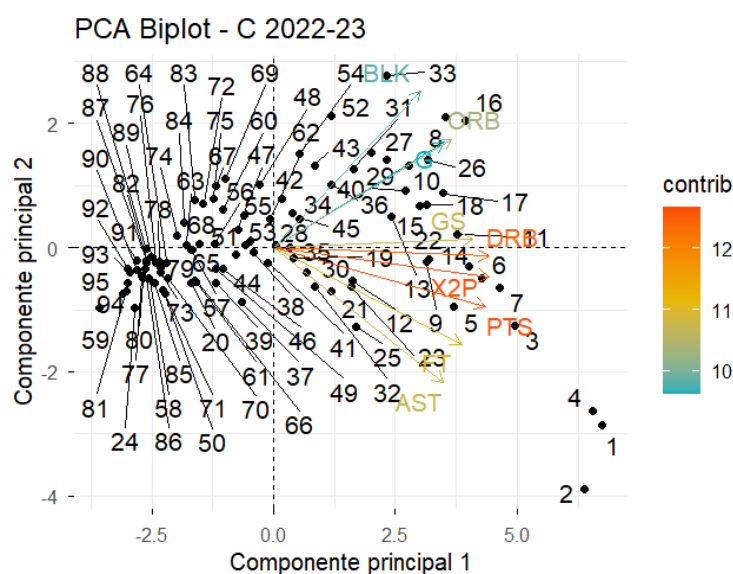
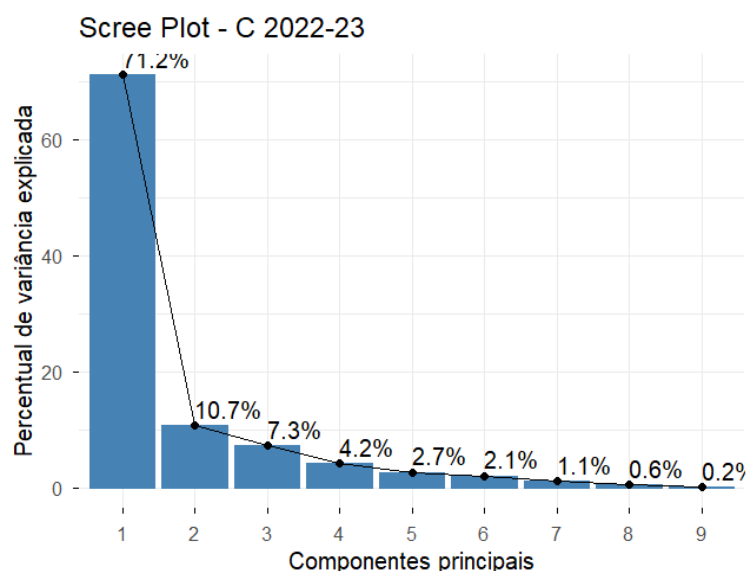
- Armadores (PG): Os componentes principais refletem a organização do jogo e volume (variáveis como G, GS, AST, TRB, STL) e a capacidade de pontuar e arremessar de fora (variáveis como PTS, X3P, FT).
- Alas-armadores (SG) e Alas (SF): Os componentes principais refletem principalmente a pontuação, eficiência de arremessos (X3P, FT) e produção geral (rebotes, assistências, roubos).
- Alas-pivôs (PF) e Pivôs (C): Para essas posições, os componentes destacam a pontuação próxima à cesta (X2P), rebotes ofensivos e defensivos (ORB, DRB), toques (BLK) e volume de participação (G, GS).

Figura 24 - Scree plots e biplots de cada posição 2022-2023









Fonte: elaborado pelo autor.

8.5.4 Clusterização por k-médias 2022-2023

Após a aplicação da PCA, o algoritmo K-means foi usado para identificar perfis de jogadores dentro de cada posição, com o número ótimo de clusters determinado pelo índice de silhueta, sugerindo $k = 2$ para todas as posições.

As silhuetas médias variaram de 0.478 (PG) a 0.544 (C), indicando um nível moderado de separação entre os clusters. Isso é esperado, dado que as características dos jogadores não são completamente distintas, mas sim agrupadas em perfis de atuação que se sobrepõem parcialmente.

- PG: $k^* = 2$, silhueta média ≈ 0.478

- SG: $k^* = 2$, silhueta média ≈ 0.55
- SF: $k^* = 2$, silhueta média ≈ 0.525
- PF: $k^* = 2$, silhueta média ≈ 0.517
- C: $k^* = 2$, silhueta média ≈ 0.544

Os clusters refletem a diferenciação entre jogadores com maior protagonismo ofensivo e maior carga de minutos jogados (Cluster 1) e jogadores mais coadjuvantes, com participações estatísticas mais modestas (Cluster 2).

8.5.5 Análise fatorial exploratória 2022-2023

Após a aplicação da PCA e da clusterização, a Análise Fatorial Exploratória (AFE) foi realizada para cada posição, a fim de identificar a estrutura latente subjacente aos dados de desempenho dos jogadores. A análise paralela sugeriu que o número ideal de fatores para a maioria das posições seria 1, exceto para os pivôs (C), onde foram extraídos 2 fatores.

Interpretação dos Fatores:

- Armadores (PG):

A AFE para os armadores (PG) sugeriu 2 fatores.

1. Fator 1: Eficiência de arremessos e pontuação. Este fator está fortemente associado a variáveis como PTS, FT, X3P, e TRB. Isso reflete a importância de pontuar e arremessar de fora para a posição de armador.
2. Fator 2: Volume de jogo e contribuição all-around. Este fator está relacionado a variáveis como AST, TRB, STL, GS, e G, que representam o envolvimento do armador no jogo, tanto em termos de assistências quanto de contribuições defensivas e minutos jogados.
3. Comunalidades (h^2) variaram de 0.317 (G) a 0.995 (PTS), indicando que, para a maioria das variáveis, o modelo fatorial é bastante eficaz em explicar a variância.

- Alas-armadores (SG) e Alas (SF):

Para os alas-armadores (SG) e alas (SF), a análise sugere 1 fator.

1. Fator 1: A única estrutura extraída reflete um desempenho geral, com forte carga em variáveis como PTS, AST, STL, GS, e FT, o que significa que as habilidades ofensivas (principalmente pontuação e arremessos) dominam o desempenho dessas posições.

2. Comunalidades para essas posições variaram de 0.450 (STL) a 0.995 (PTS), indicando que o modelo fatorial é capaz de explicar a maior parte da variância nas variáveis, especialmente em termos de eficiência ofensiva e produção geral.

- **Alas-pivôs (PF):**

Para os alas-pivôs (PF), a análise indicou 1 fator.

1. Fator 1: Este fator está fortemente relacionado a variáveis como PTS, ORB, DRB, X2P, BLK, e GS, o que reflete um desempenho defensivo e ofensivo de frontcourt, com ênfase em rebotes e eficiência ofensiva na pintura.
2. Comunalidades variaram de 0.311 (BLK) a 0.995 (PTS), mostrando que, embora algumas variáveis (como BLK) expliquem menos da variância, o modelo ainda captura bem o desempenho geral dos alas-pivôs.

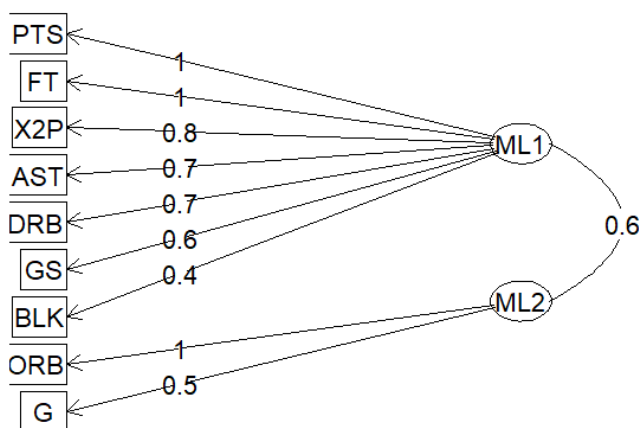
- **Pivôs (C):**

Para os pivôs (C), a análise sugeriu 2 fatores.

1. Fator 1: Pontuação e eficiência ofensiva do garrafão, com alta carga em variáveis como PTS, X2P, AST, e GS. Este fator reflete a capacidade de pontuar e organizar o jogo no garrafão.
2. Fator 2: Presença física e proteção de aro, com ênfase em BLK, ORB, e DRB, refletindo o impacto defensivo e a capacidade de proteger o aro.
3. Comunalidades variaram de 0.400 (BLK) a 0.995 (PTS), com o modelo explicando bem as variáveis relacionadas à defesa e à eficiência ofensiva.

Figura 25 - Diagramas de fatores e Análises Paralelas por posição 2022-2023

Diagrama de Fatores - C 2022-23



Análise Paralela - C 2022-23

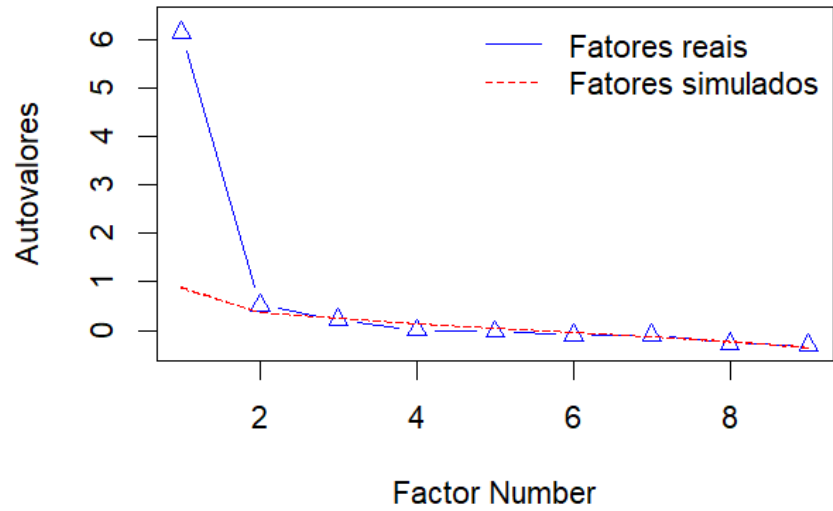
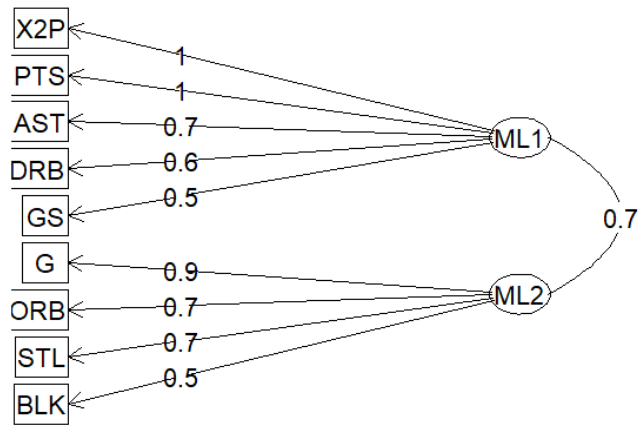


Diagrama de Fatores - PF 2022-23



Análise Paralela - PF 2022-23

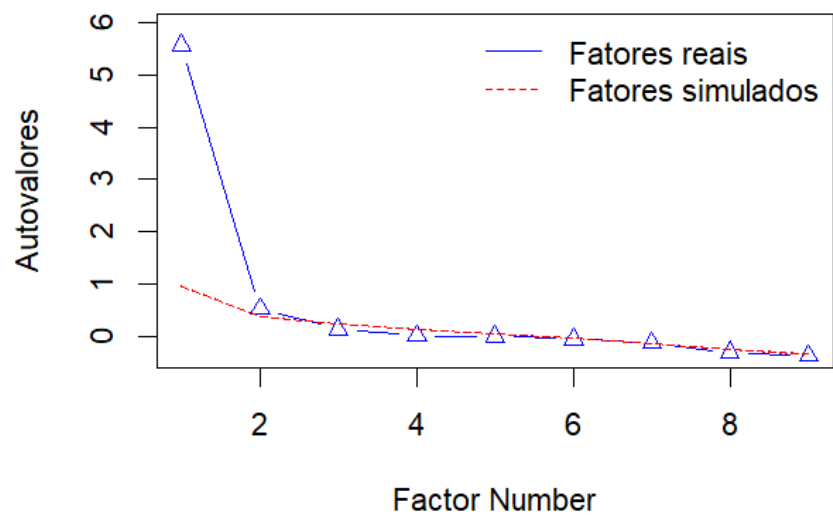
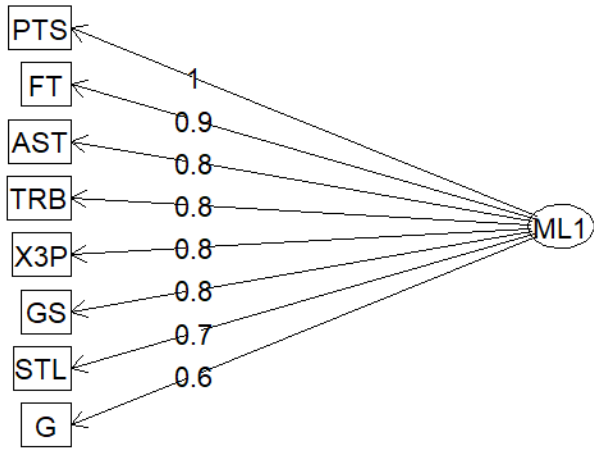


Diagrama de Fatores - PG 2022-23



Análise Paralela - PG 2022-23

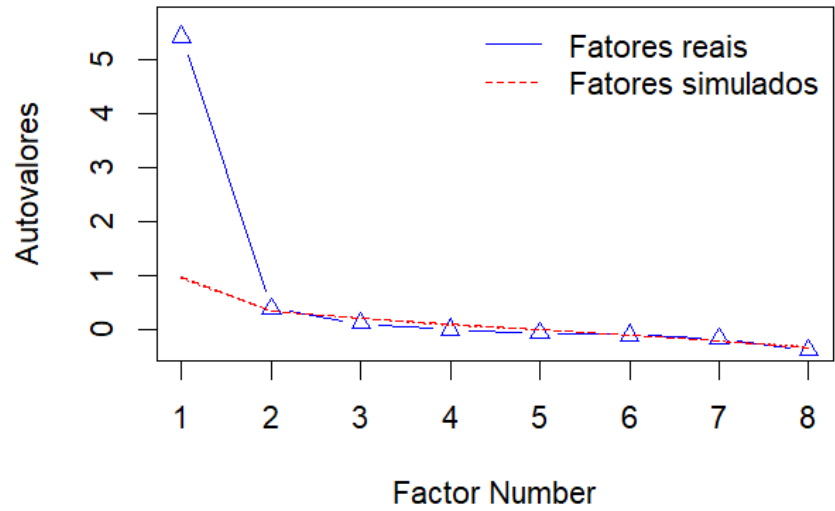
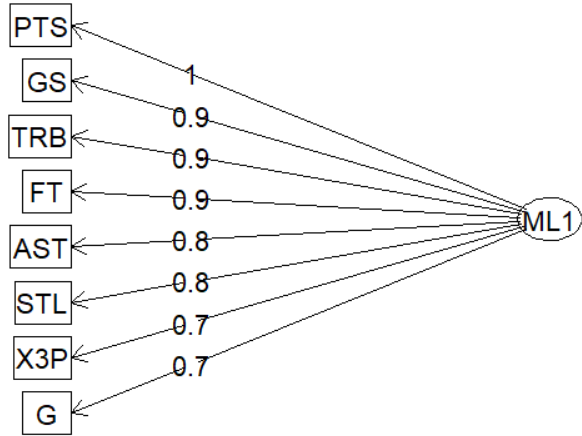


Diagrama de Fatores - SF 2022-23



Análise Paralela - SF 2022-23

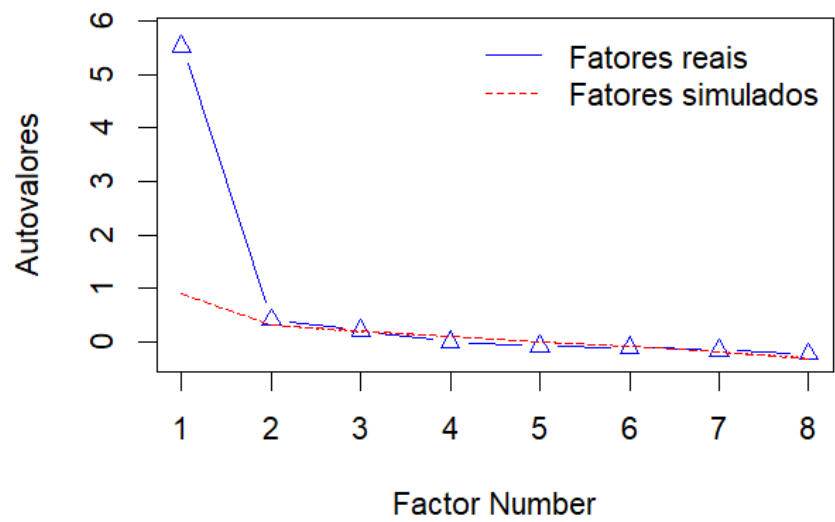
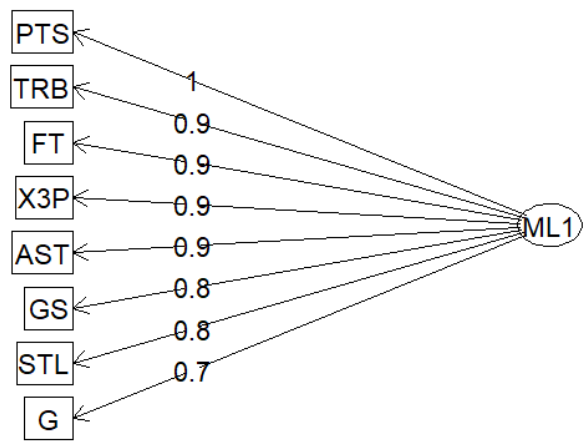
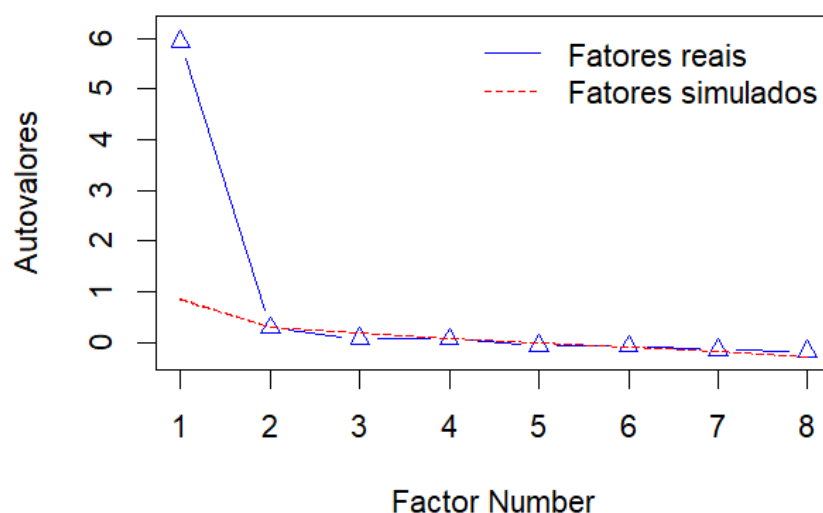


Diagrama de Fatores - SG 2022-23



Análise Paralela - SG 2022-23



Fonte: elaborado pelo autor.

8.5.6 Construção dos ranking por posição 2022-2023

A partir dos resultados da AFE por posição, os rankings foram construídos em duas etapas principais: cálculo dos escores fatoriais e combinação dos fatores utilizando pesos proporcionais à variância explicada.

As equações finais para os rankings das posições foram baseadas nas variáveis mais relevantes para cada função em quadra. Por exemplo, para os armadores (PG), o score final foi formado pelas variáveis PTS, AST, FT, e TRB.

Tabela 23 – Top 10 C 2022-23

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Joel Embiid	C	66	66	2284	33,08	10,15	4,15	100,00	1
2	Domantas Sabonis	C	79	79	2736	19,11	12,32	7,25	90,53	1
3	Nikola Jokic	C	69	69	2323	24,49	11,84	9,83	87,11	1
4	Bam Adebayo	C	75	75	2598	20,39	9,17	3,20	82,32	1
5	Anthony Davis	C	56	54	1904	25,91	12,54	2,64	79,85	1
6	Nikola Vučević	C	82	82	2746	17,65	11,01	3,23	75,29	1
7	Alperen Şengün	C	75	72	2171	14,79	9,04	3,88	71,50	1

8	Kristaps Porziņģis	C	65	65	2120	23,15	8,40	2,68	70,42	1
9	Jonas Valančiūnas	C	79	79	1968	14,11	10,18	1,77	69,90	1
10	Deandre Ayton	C	67	67	2035	17,96	9,96	1,72	67,00	1

Fonte: elaborado pelo autor.

Tabela 24 – Top 10 PF 2022-23

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Giannis Antetokounmpo	PF	63	63	2024	31,10	11,78	5,70	100,00	1
2	Evan Mobley	PF	79	79	2715	16,16	9,00	2,84	93,51	1
3	Julius Randle	PF	77	77	2737	25,14	9,96	4,10	90,13	1
4	Pascal Siakam	PF	71	71	2652	24,23	7,83	5,85	88,63	1
5	Josh Giddey	PF	76	76	2367	16,58	7,88	6,17	82,71	1
6	Jimmy Butler	PF	64	64	2138	22,91	5,86	5,31	78,31	1
7	Paolo Banchero	PF	72	72	2430	19,96	6,90	3,74	72,76	1
8	Lauri Markkanen	PF	66	66	2273	25,62	8,64	1,86	72,63	1
9	LeBron James	PF	55	54	1954	28,91	8,31	6,82	72,40	1
10	Aaron Gordon	PF	68	68	2055	16,31	6,56	2,99	67,80	1

Fonte: elaborado pelo autor.

Tabela 25 – Top 10 PG 2022-23

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Luka Dončić	PG	66	66	2391	32,39	8,62	8,02	100,00	1
2	Shai Gilgeous-Alexander	PG	68	68	2416	31,40	4,84	5,46	99,37	1
3	Trae Young	PG	73	73	2541	26,22	2,97	10,15	89,41	1
4	Damian Lillard	PG	58	58	2107	32,17	4,78	7,33	86,45	1
5	De'Aaron Fox	PG	73	73	2435	25,01	4,19	6,12	83,67	1
6	Jordan Poole	PG	82	43	2458	20,43	2,73	4,50	76,57	1

7	Stephen Curry	PG	56	56	1941	29,43	6,09	6,29	75,19	1
8	Jalen Brunson	PG	68	68	2379	24,01	3,54	6,19	74,38	1
9	Ja Morant	PG	61	59	1948	26,16	5,85	8,08	73,12	1
10	CJ McCollum	PG	75	75	2649	20,91	4,37	5,72	71,35	1

Fonte: elaborado pelo autor.

Tabela 26 – Top 10 SF 2022-23

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Jayson Tatum	SF	74	74	2732	30,07	8,77	4,62	100,00	1
2	DeMar DeRozan	SF	74	74	2682	24,54	4,64	5,09	80,69	1
3	Jaylen Brown	SF	67	67	2405	26,63	6,85	3,46	74,05	1
4	Franz Wagner	SF	80	80	2609	18,56	4,11	3,54	68,27	1
5	Scottie Barnes	SF	77	76	2678	15,31	6,65	4,82	63,96	1
6	Buddy Hield	SF	80	73	2482	16,80	5,00	2,81	63,90	1
7	Paul George	SF	56	56	1939	23,79	6,11	5,14	61,67	1
8	Klay Thompson	SF	69	69	2279	21,87	4,14	2,36	60,27	1
9	Keldon Johnson	SF	63	63	2063	21,98	5,05	2,90	57,80	1
10	OG Anunoby	SF	67	67	2386	16,78	4,96	1,96	54,49	1

Fonte: elaborado pelo autor.

Tabela 27 – Top 10 SG 2022-23

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Anthony Edwards	SG	79	79	2842	24,63	5,80	4,43	100,00	2
2	Zach LaVine	SG	77	77	2768	24,84	4,48	4,25	94,63	2
3	Donovan Mitchell	SG	68	68	2432	28,26	4,25	4,43	93,71	2
4	Mikal Bridges	SG	83	83	2963	20,13	4,39	3,29	85,08	2
5	Jalen Green	SG	76	76	2602	22,14	3,74	3,70	83,71	2
6	Kyrie Irving	SG	60	60	2241	27,05	5,07	5,52	79,20	2
7	Dejounte Murray	SG	74	74	2693	20,47	5,26	6,05	78,39	2

8	Spencer Dinwiddie	SG	79	79	2725	17,33	3,42	6,52	74,38	2
9	RJ Barrett	SG	73	73	2475	19,60	5,04	2,75	71,00	2
10	Devin Booker	SG	53	53	1835	27,75	4,53	5,53	69,88	2

Fonte: elaborado pelo autor.

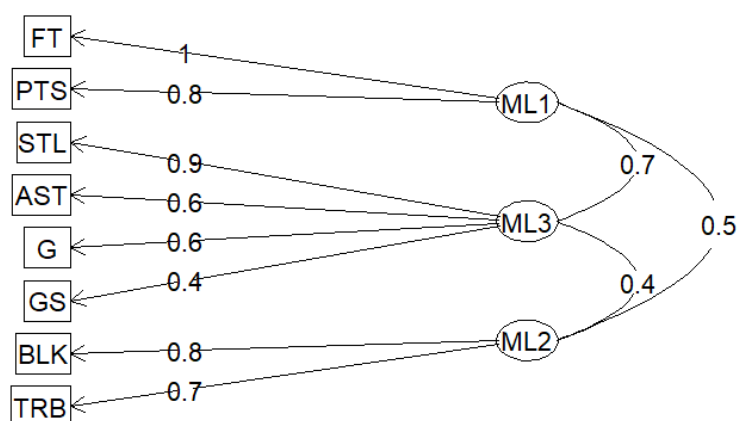
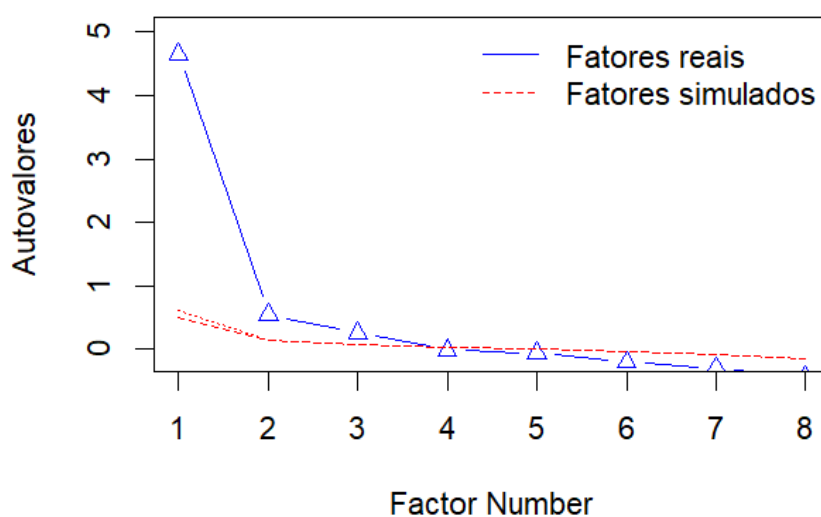
8.5.7 Ranking geral 2022-2023

Após os rankings por posição, foi gerado um ranking geral, combinando todas as posições. A variância explicada cumulativa pelos três fatores foi de 76%, o que indica que o modelo é eficiente para resumir as diferenças entre os jogadores.

A equação final do índice geral incorporou variáveis como PTS, GS, AST, TRB, STL, BLK, e FT. A visualização do ranking geral pode ser feita com gráficos de Top 10, mostrando os jogadores mais impactantes da temporada.

O MVP da temporada 2022-2023 foi Joel Embiid, que ocupou a 1ª posição no ranking geral, com score normalizado de 100. Isso valida a eficácia do modelo de análise, já que o jogador mais reconhecido oficialmente aparece no topo do ranking, reforçando a robustez da metodologia estatística utilizada.

Figura 26 - Diagrama de fatores, Análise paralela e Top10 geral 2022-2023

Fatores - Ranking Geral 2022-23**Análise paralela - Ranking geral 2022-23**

Fonte: elaborado pelo autor.

Tabela 28 – Top 10 Geral 2022-23

rank_geral	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_geral_norm	cluster
1	Joel Embiid	C	66	66	2284	33,08	10,15	4,15	100,00	1
2	Jayson Tatum	SF	74	74	2732	30,07	8,77	4,62	98,32	1
3	Shai Gilgeous-Alexander	PG	68	68	2416	31,40	4,84	5,46	95,60	1

4	Luka Dončić	PG	66	66	2391	32,39	8,62	8,02	94,97	1
5	Anthony Edwards	SG	79	79	2842	24,63	5,80	4,43	93,71	2
6	Nikola Jokic	C	69	69	2323	24,49	11,84	9,83	91,56	1
7	Giannis Antetokounmpo	PF	63	63	2024	31,10	11,78	5,70	88,25	1
8	Domantas Sabonis	C	79	79	2736	19,11	12,32	7,25	88,12	1
9	Julius Randle	PF	77	77	2737	25,14	9,96	4,10	86,79	1
10	Bam Adebayo	C	75	75	2598	20,39	9,17	3,20	82,77	1

Fonte: elaborado pelo autor.

8.6 ANÁLISE EXPLORATÓRIA DOS DADOS DA TEMPORADA 2023-2024

Encerrando a etapa de exploração descritiva, esta subseção examina os dados da temporada 2023-24. A análise contempla a distribuição das variáveis associadas ao desempenho dos atletas, permitindo observar diferenças de atuação entre posições, níveis de protagonismo ofensivo e impacto defensivo. A partir de medidas de tendência central e dispersão, são destacadas tendências relevantes e jogadores com perfis estatísticos atípicos, oferecendo uma visão geral consistente para as etapas analíticas seguintes.

8.6.1 Estatísticas das variáveis da temporada 2023-2024

A análise descritiva das variáveis da temporada 2023-24 da NBA, apresentada na tabela abaixo, contempla estatísticas fundamentais que permitem avaliar a distribuição e variação dos indicadores de desempenho dos atletas. Esses dados oferecem um panorama quantitativo inicial indispensável para a interpretação dos resultados obtidos nas etapas seguintes da pesquisa.

Tabela 29 - Estatística das variáveis 23-24

VARIÁVEL	MÉDIA	D.PADRÃO	MÍN	Q1	MEDIANA	Q3	MAX
G	40,8802	25,5812	1	17	41	64,5	84
GS	18,4408	25,0241	0	0	5	29	82
MP	907,7578	809,4986	1	160	680	1527,5	2989

FG	156,1523	173,9499	0	19	88	235,5	837
FGA	330,6965	358,1608	0	47,5	201	496,5	1652
FG.	0,4475	0,1106	0	0,4	0,4465	0,5	0,8
X3P	48,0136	59,1260	0	2	22	77	357
X3PA	131,8244	154,3069	0	10	69	209,5	876
X3P.	0,3169	0,1371	0	0,2802	0,346	0,385	1
X2P	108,1387	133,6266	0	13	55	149,5	803
X2PA	198,8721	240,1321	0	26	107	271,5	1245
X2P.	0,5197	0,1367	0	0,4785	0,53	0,581	1
eFG.	0,5157	0,1146	0	0,483	0,5295	0,575	0,917
FT	62,5850	88,2262	0	6	28	79	567
FTA	79,9170	109,3416	0	9	39	104	782
FT.	0,7497	0,1599	0	0,686	0,778	0,84	1
ORB	39,3986	47,7896	0	7	23	54	335
DRB	122,6272	132,1360	0	19	82	183,5	826
TRB	162,0258	174,1016	0	28	113	241,5	1120
AST	99,8653	124,6541	0	13	54	130	752
STL	28,1673	27,0037	0	5	20	44	150
BLK	19,0870	26,7057	0	3	9	27	254
TOV	48,0190	53,4040	0	7	30	69	282
PF	70,5836	61,7233	0	15	56	112,5	254
PTS	422,9034	475,8143	0	51,5	244	624,5	2370

Fonte: elaborado pelo autor (2025)

Observando a tabela de estatísticas das variáveis, diversas informações relevantes podem ser extraídas, oferecendo uma visão sobre as características dos jogadores da NBA na temporada 2023-2024:

a) Jogos e Participação: O número de jogos disputados apresenta média de 40,88 jogos por jogador (mínimo 1, máximo 84). Quanto aos jogos como titular, a mediana de apenas 5 jogos, contrastando com o máximo de 82, revela clara distinção entre titulares fixos e reservas, evidenciando a hierarquia nas equipes.

b) Minutos Jogados: Média de 907,76 minutos por temporada, variando de 1 a 2.989 minutos. A mediana de 680 minutos indica que metade dos jogadores tem participação moderada, enquanto um grupo seleto acumula tempo significativamente maior.

c) Eficiência Ofensiva: A porcentagem de arremessos de quadra tem média de 44,76% com mediana de 44,65%, sugerindo distribuição relativamente simétrica. Nas cestas de três pontos, a média de 31,70% com mediana de 34,6% demonstra eficiência considerável dos especialistas em arremessos de longa distância.

d) Volume Ofensivo: Os arremessos convertidos apresentam média de 156,15 cestas por temporada (máximo 837), enquanto as tentativas têm média de 330,70 (máximo 1.652). Nos pontos totais, a média de 422,90 contrasta com o máximo de 2.370, evidenciando o papel dominante dos principais pontuadores.

e) Rebotes: Média de 162,03 rebotes totais por temporada (máximo 1.120), com predominância dos defensivos (média 122,63) sobre os ofensivos (média 39,40). A mediana de rebotes totais (113) significativamente abaixo da média indica concentração em jogadores específicos, tipicamente pivôs e alas-pivôs.

f) Estatísticas Defensivas e de Criação: As assistências apresentam média de 99,87 (máximo 752), evidenciando especialização dos armadores. Roubo de bola (média 28,17) e bloqueios (média 19,09) mostram medianas abaixo das médias, indicando que poucos jogadores se destacam nestas estatísticas defensivas.

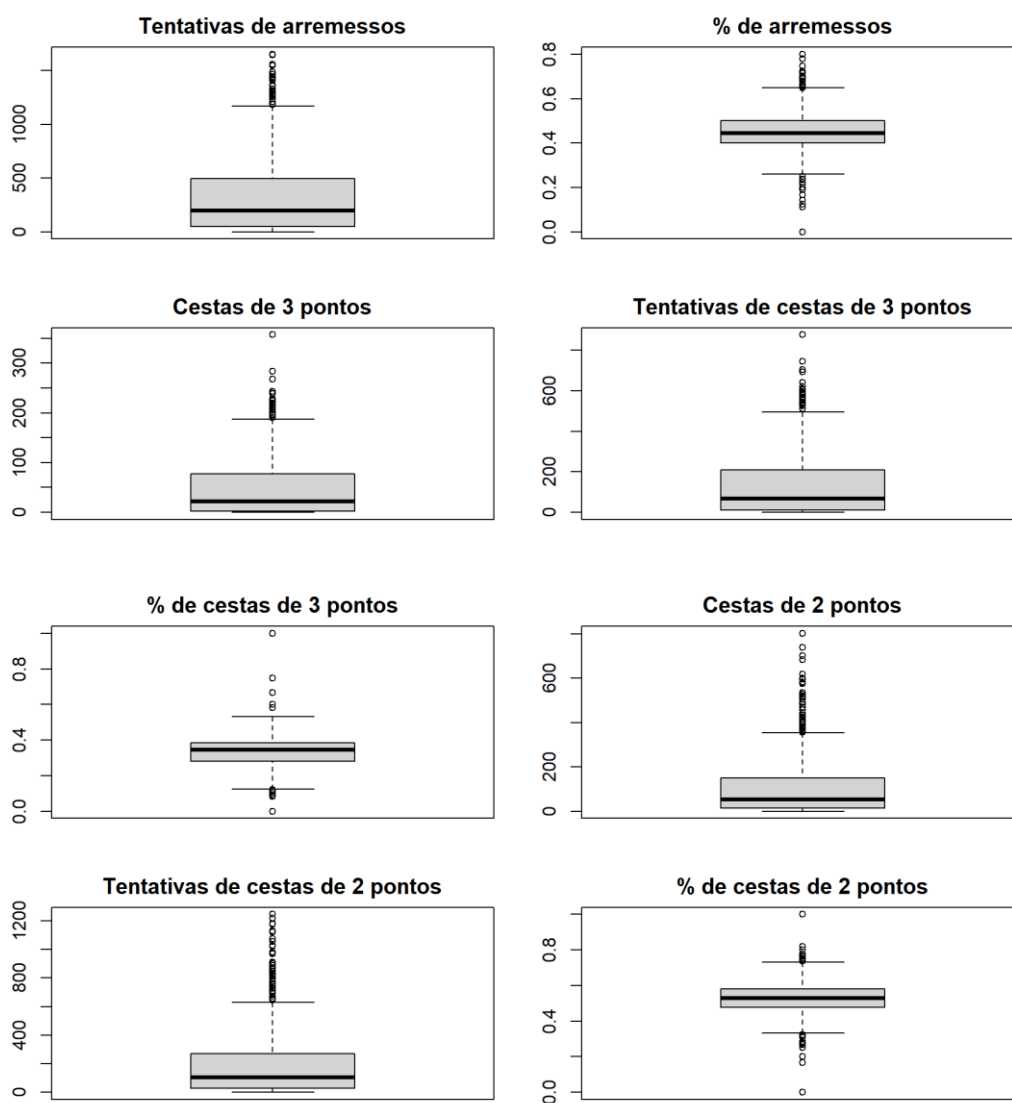
Esta análise revela a diversidade de perfis na NBA, com clara especialização em determinados fundamentos e distribuição assimétrica para variáveis de volume, contrastando com distribuições mais equilibradas para variáveis de eficiência.

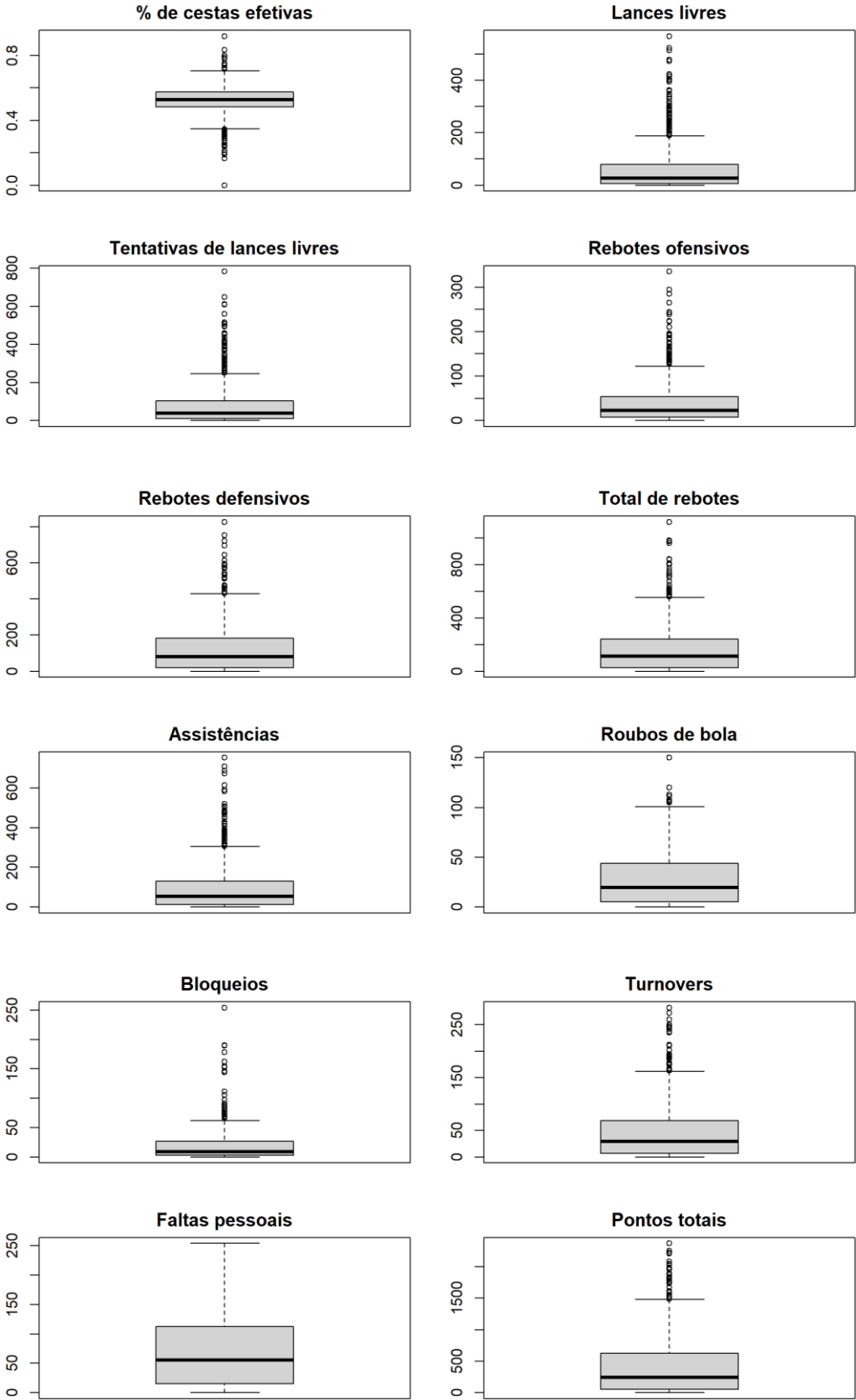
8.6.2 Análise gráfica das variáveis da temporada 2023-2024

Por fim, esta subseção apresenta os gráficos correspondentes à temporada 2023-24 da NBA, utilizando boxplots, histogramas e correlogramas como ferramentas de análise visual.

Tais representações permitem observar a distribuição estatística das variáveis, detectar valores extremos e compreender as relações entre os diferentes aspectos do desempenho dos jogadores. Essa etapa é fundamental para consolidar a leitura exploratória dos dados antes da aplicação dos métodos multivariados.

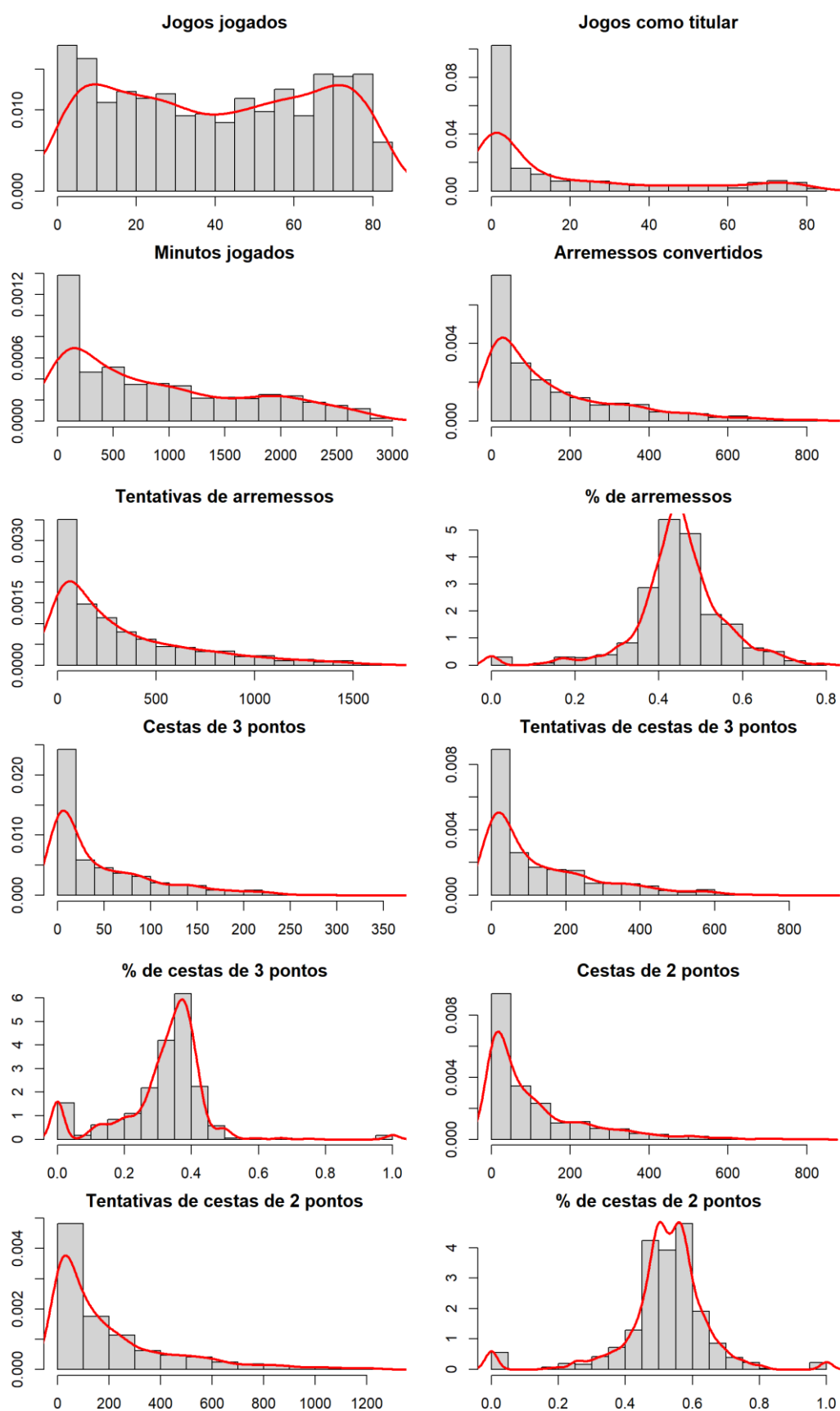
Figura 27 - Boxplots das variáveis 23-24

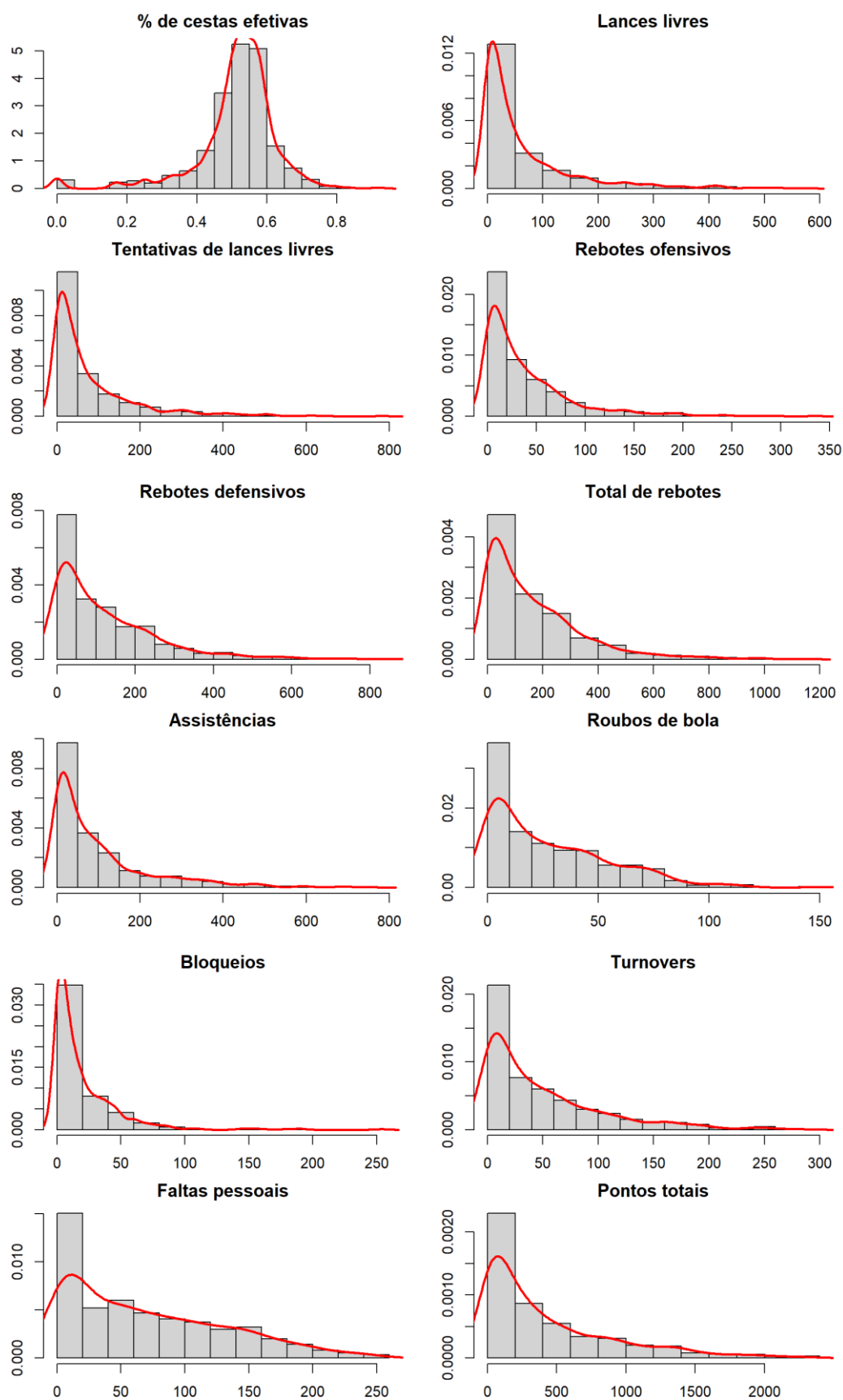




Fonte: elaborado pelo autor (2025)

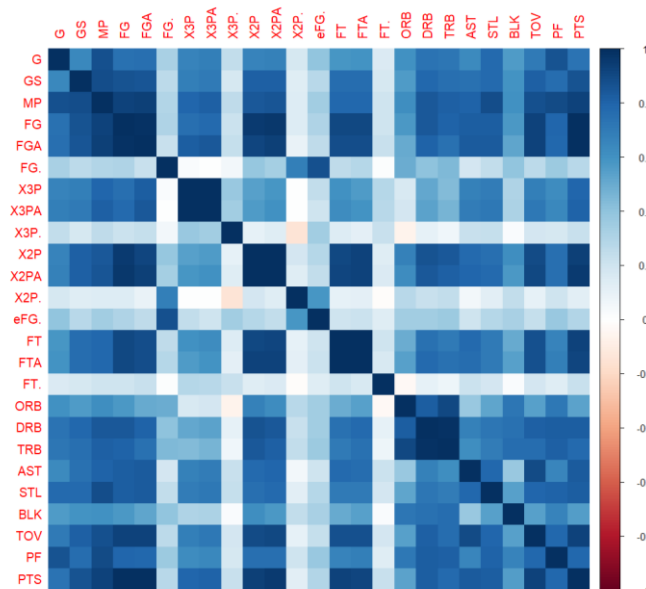
Figura 28 - Histogramas das variáveis 23-24





Fonte: elaborado pelo autor (2025)

Figura 29 - Correlograma das variáveis 23-24



Fonte: elaborado pelo autor (2025)

Os histogramas e boxplots dessas variáveis evidenciam forte assimetria à direita. A maioria dos atletas atuou em menos da metade da temporada, com concentração de jogos abaixo dos 40. A variável MP (minutos jogados) confirma essa tendência, com valores concentrados abaixo dos 1000 minutos. Os outliers visíveis representam atletas com alto volume de quadra e centralidade tática nas equipes.

As variáveis associadas à produção ofensiva apresentam distribuição altamente assimétrica. Os histogramas de FG (cestas feitas), FGA (tentativas) e PTS (pontos totais) revelam grande número de jogadores com baixa participação ofensiva, com caudas longas atingindo atletas que ultrapassam 2000 pontos. Os boxplots reforçam a existência de outliers marcantes, indicando o domínio ofensivo de poucos jogadores no cenário da liga.

As distribuições dessas variáveis também se mostram assimétricas, com destaque para a grande dispersão em 3PA (tentativas de 3 pontos), revelando uma diversidade tática no uso do perímetro entre os jogadores. O volume de tentativas é geralmente baixo, com poucos atletas superando marcas de 500 tentativas. As distribuições de 2P e 2PA seguem padrão semelhante, porém com menor dispersão.

As métricas de eficiência, ao contrário das anteriores, apresentam distribuições simétricas e compactas, com boxplots sem grande presença de outliers. As médias giram em torno de 44% a 52% em arremessos gerais (FG%), com 3P% e FT% mantendo padrão próximo

de normalidade. Esse comportamento denota consistência técnica entre os jogadores que atuaram por tempo relevante, mesmo com variação no volume de arremessos.

As distribuições de ORB (rebotes ofensivos), DRB (defensivos) e TRB (totais) evidenciam ampla dispersão, com assimetrias acentuadas. Os gráficos mostram que apenas um subconjunto reduzido de atletas domina os rebotes. O mesmo se aplica à variável AST, com distribuição concentrada nos níveis baixos e presença clara de outliers — representando os principais armadores e distribuidores de bola da liga.

As variáveis defensivas e de controle de posse revelam comportamento esperado: STL (roubos) e BLK (tocos) possuem baixa média e alta assimetria. TOV (erros) segue padrão semelhante, com jogadores de protagonismo ofensivo concentrando os valores mais elevados. PF (faltas pessoais) apresenta distribuição mais simétrica, sugerindo um padrão defensivo moderadamente uniforme entre os atletas.

O mapa de calor confirma fortes correlações entre variáveis de volume, como FG, FGA, PTS, MP e AST. Tais correlações refletem o fato de que tempo em quadra e protagonismo ofensivo estão intrinsecamente ligados. Por outro lado, as métricas de eficiência (FG%, FT%, eFG%) apresentam correlações menores com variáveis de volume, validando a ideia de que volume e eficiência são dimensões estatísticas distintas e complementares.

8.6.3 Análise de componentes principais 2023-2024

Para a temporada 2023–24, após a aplicação dos critérios mínimos de participação — pelo menos 20 jogos e 300 minutos totais — a base de dados ficou composta por 493 jogadores. Esse filtro assegura que apenas jogadores com participação significativa ao longo da temporada sejam analisados, evitando distorções causadas por desempenhos isolados com baixa amostragem.

A PCA (Análise de Componentes Principais) foi realizada separadamente para cada posição (PG, SG, SF, PF, C), utilizando variáveis relevantes para o desempenho de cada função em quadra. Antes da aplicação da PCA, foi feita a avaliação das medidas de adequação da matriz de correlação, e os resultados foram extremamente satisfatórios.

- O KMO global para a posição PG foi de 0.875, indicando boa adequação amostral. Os valores de KMO por variável para as outras posições também foram elevados:
 - PG: 0.875
 - SG: 0.829
 - SF: 0.804

- PF: 0.79
- C: 0.834
- O teste de esfericidade de Bartlett foi altamente significativo ($p < 10^{-174}$) para todas as posições, confirmando que as variáveis são correlacionadas e que há uma estrutura latente nas variáveis analisadas.

A variância explicada pelos três primeiros componentes principais foi de 89.9% para os armadores (PG), o que demonstra que a redução de dimensionalidade foi eficaz, preservando uma grande parte da informação original.

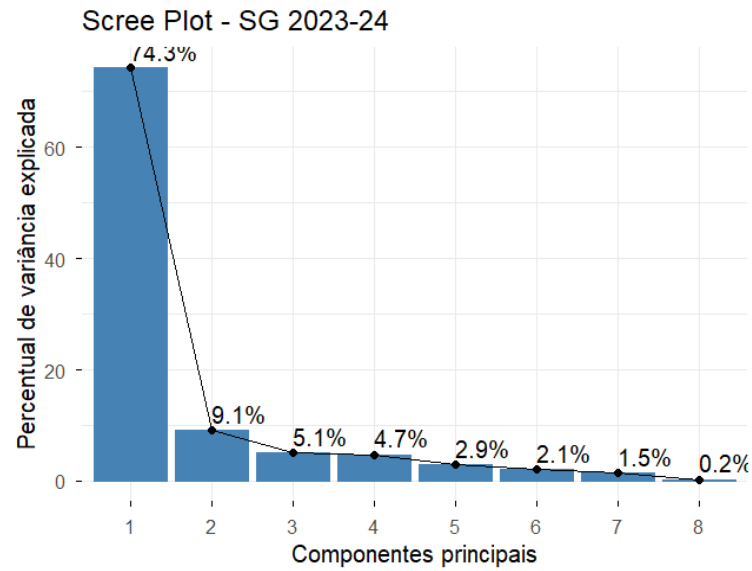
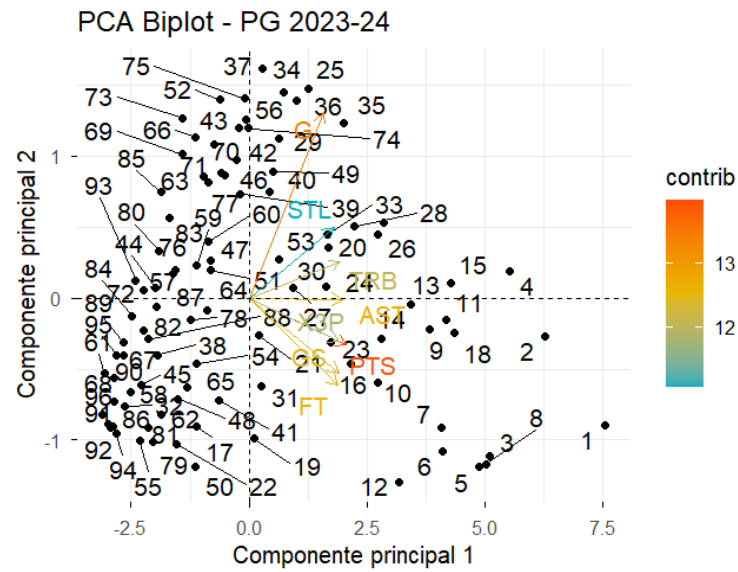
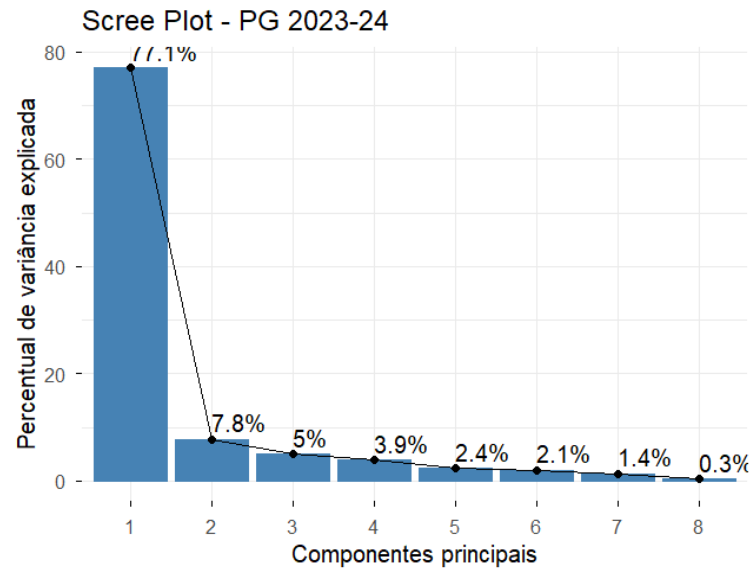
- PG: 89.9%
- SG: 88.49%
- SF: 86.95%
- PF: 87.43%
- C: 89.88%

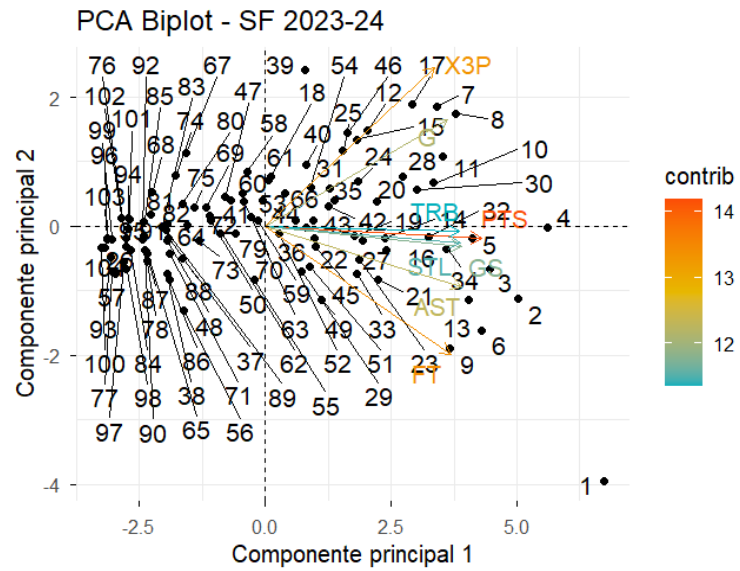
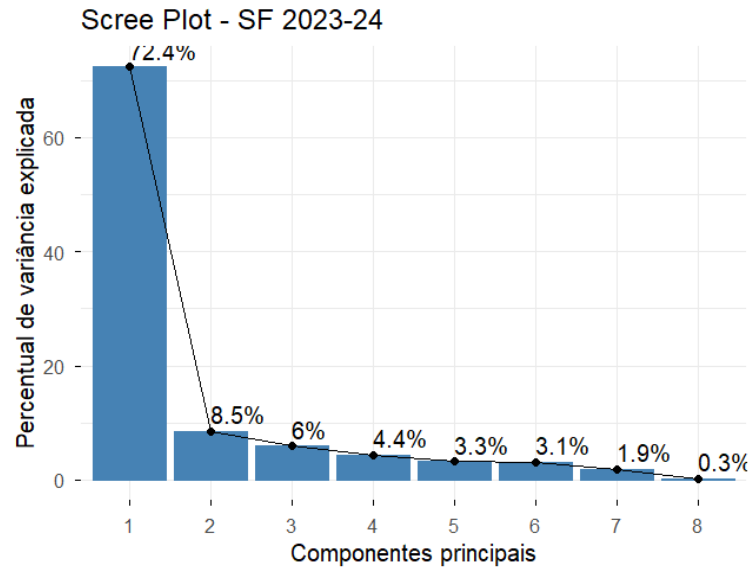
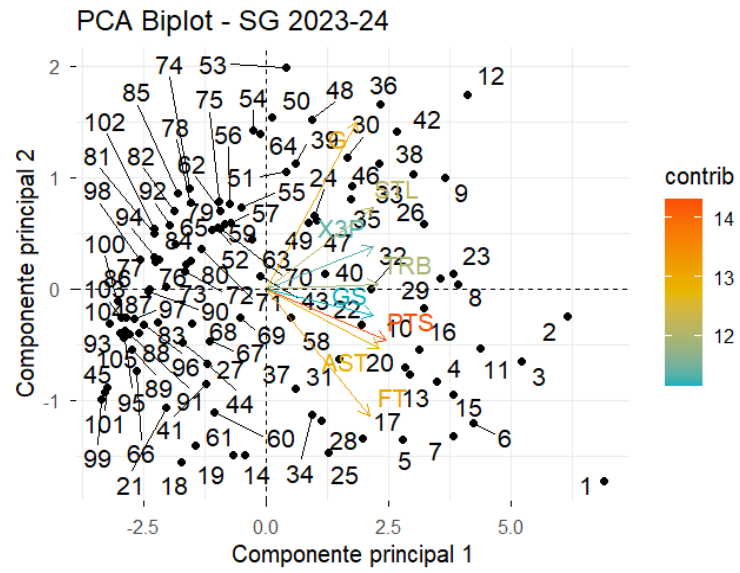
Esses percentuais elevados indicam que apenas os três primeiros componentes principais são suficientes para explicar a maior parte da variação nos dados, o que é ideal para fins de redução de dimensionalidade e visualização.

Interpretação Conceitual:

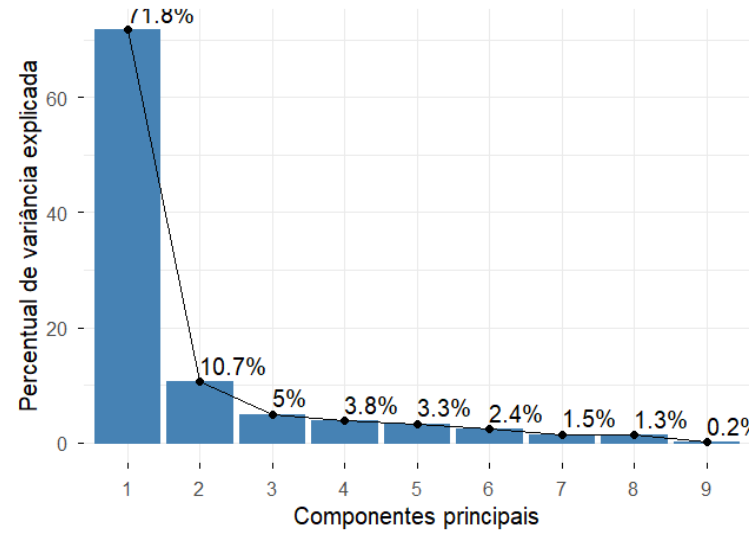
- Armadores (PG): Os componentes principais refletem a organização do jogo e volume (variáveis como G, GS, AST, TRB, STL) e a capacidade de pontuar e arremessar de fora (variáveis como PTS, X3P, FT).
- Alas-armadores (SG) e Alas (SF): Os componentes principais refletem a pontuação, eficiência de arremessos (X3P, FT) e produção geral (rebotes, assistências, roubos).
- Alas-pivôs (PF) e Pivôs (C): Para essas posições, os componentes refletem a pontuação próxima à cesta (X2P), rebotes ofensivos e defensivos (ORB, DRB), toques (BLK) e participação nos jogos (G, GS).

Figura 30 - Scree plots e biplots de cada posição 2023-2024

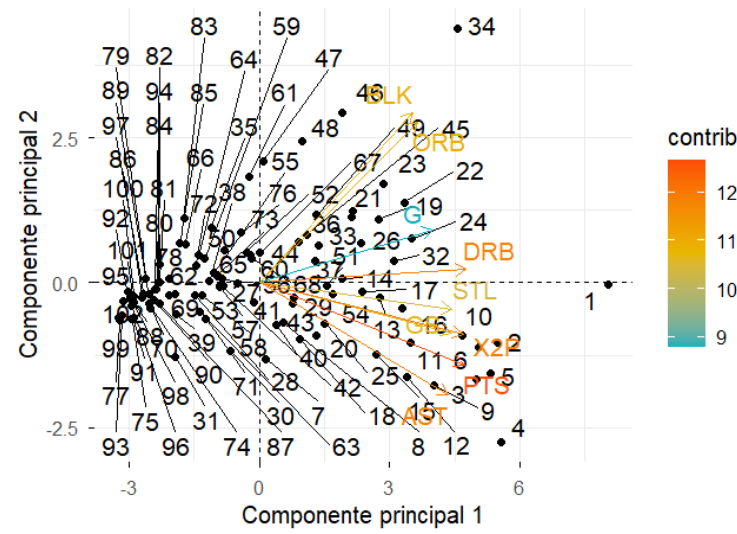


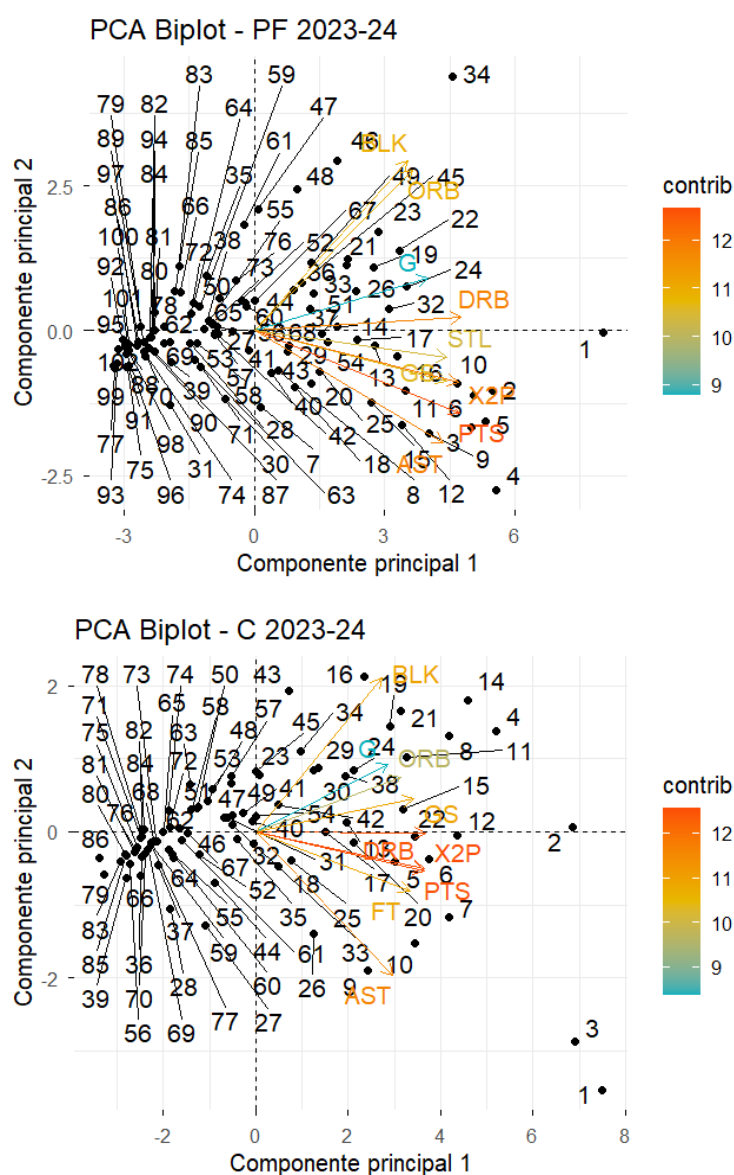


Scree Plot - PF 2023-24



PCA Biplot - PF 2023-24





Fonte: elaborado pelo autor.

8.6.4 Clusterização por k-médias 2023-2024

Com os escores dos três primeiros componentes principais, foi aplicado o algoritmo K-means para identificar perfis de jogadores dentro de cada posição. O número ótimo de clusters foi sugerido pelo índice de silhueta, que indicou $k = 2$ para todas as posições.

As silhuetas médias variaram de 0.478 (PG) a 0.567 (C), indicando um nível moderado de separação entre os clusters. Este valor está dentro do esperado para dados esportivos, onde as distâncias entre os clusters podem ser mais sutis.

- PG: $k^* = 2$, silhueta média ≈ 0.57
- SG: $k^* = 2$, silhueta média ≈ 0.53

- SF: $k^* = 2$, silhueta média ≈ 0.53
- PF: $k^* = 2$, silhueta média ≈ 0.53
- C: $k^* = 2$, silhueta média ≈ 0.56

O K-means resultou em dois clusters, refletindo a diferenciação entre jogadores com maior protagonismo ofensivo e maior carga de minutos jogados (Cluster 1), e jogadores com contribuições mais modestas (Cluster 2).

8.6.5 Análise fatorial exploratória por posição 2023-2024

Após a aplicação da PCA e da clusterização dos jogadores, foi realizada a Análise Fatorial Exploratória (AFE) para cada posição. O objetivo da AFE foi identificar a estrutura latente que subjaz os dados e agrupar as variáveis correlacionadas, reduzindo a dimensionalidade e facilitando a interpretação dos fatores que mais influenciam o desempenho dos jogadores.

A análise paralela sugeriu que o número ideal de fatores seria 1 para a maioria das posições, exceto para os pivôs (C), onde 2 fatores foram extraídos. A interpretação dos fatores extraídos e a análise das comunalidades (h^2) indicam quão bem cada variável é representada pelos fatores.

Interpretação dos Fatores:

- Armadores (PG):

A análise fatorial para os armadores (PG) resultou em 1 fator.

1. Fator único: Este fator reflete um desempenho global dos armadores, com forte ênfase em pontuação e eficiência de arremessos (como PTS, X3P, FT) e também em contribuições all-around, como assistências (AST), rebotes (TRB), e roubos de bola (STL). Este fator abrange as principais responsabilidades dos armadores em termos de organização do jogo e produção ofensiva.
2. Comunalidades (h^2) variaram de 0.415 (para G) a 0.989 (para PTS), o que indica que a maior parte da variância dos dados pode ser explicada por esse único fator, especialmente para variáveis como PTS, AST, e TRB.

- Alas-armadores (SG) e Alas (SF):

Para os alas-armadores (SG) e alas (SF), a análise fatorial também extraiu 1 fator.

1. Fator único: O fator extraído reflete um desempenho global, com forte ênfase em pontuação, assistências e eficiência de arremessos. Para os alas-armadores (SG), o desempenho é caracterizado pela combinação de variáveis como PTS,

AST, STL, e X3P, que refletem tanto a produção ofensiva quanto a contribuição defensiva e de perímetro.

2. Comunalidades para essas posições foram altas, especialmente para PTS (com uma comunalidade de 0.995), o que indica que o fator extraído representa bem o desempenho global dos jogadores.

Para os alas (SF), a versatilidade do jogador é refletida no fator, com PTS, GS, AST, e STL como as variáveis mais relevantes. Essas variáveis indicam a importância do impacto geral dos alas no jogo, tanto em termos de ofensiva quanto de contribuição defensiva.

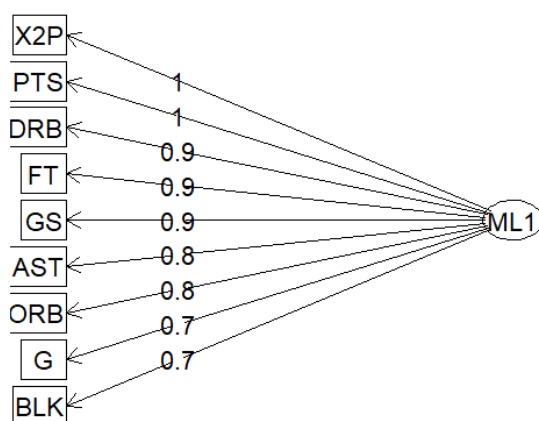
- Alas-pivôs (PF) e Pivôs (C):

Para os pivôs (C), a análise fatorial sugeriu 2 fatores.

1. Fator 1: Este fator reflete a eficiência ofensiva e participação no garrafão, com forte carga em variáveis como PTS, X2P, e AST. Essas variáveis refletem o impacto dos pivôs na pontuação próxima à cesta e na organização do jogo no garrafão.
2. Fator 2: Este fator reflete a presença defensiva e proteção de aro, com ORB, DRB, BLK e G como as variáveis mais relevantes. Esse fator evidencia a importância da defesa, com ênfase no controle de rebotes e proteção de aro.

Figuras 31 - Diagrama de Fatores e Análise paralela por posição

Diagrama de Fatores - C 2023-24



Análise Paralela - C 2023-24

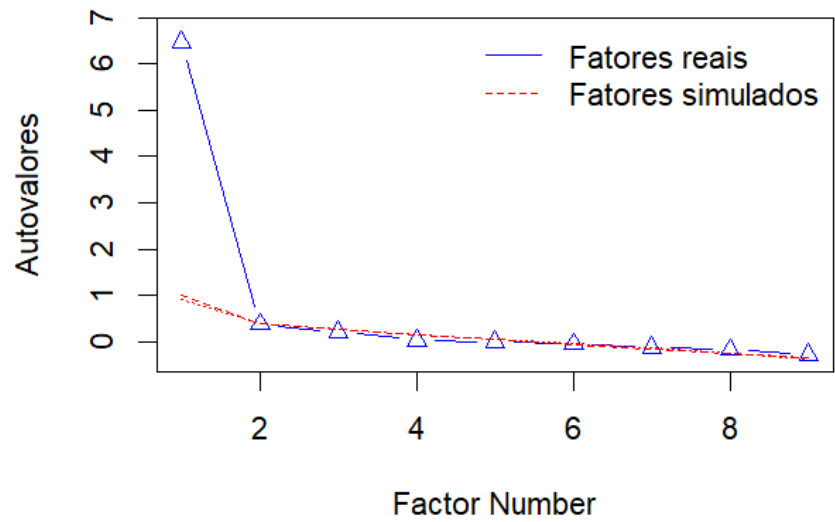
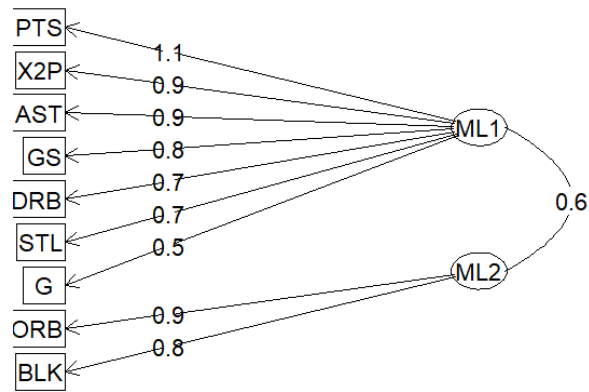


Diagrama de Fatores - PF 2023-24



Análise Paralela - PF 2023-24

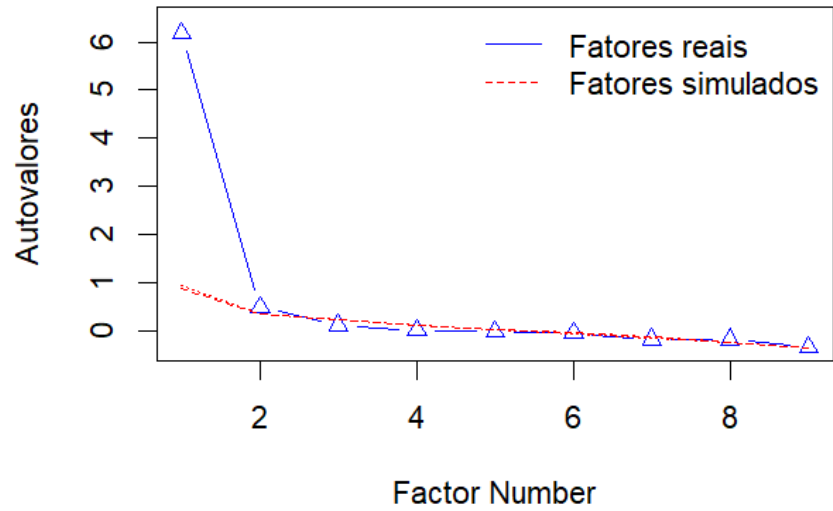
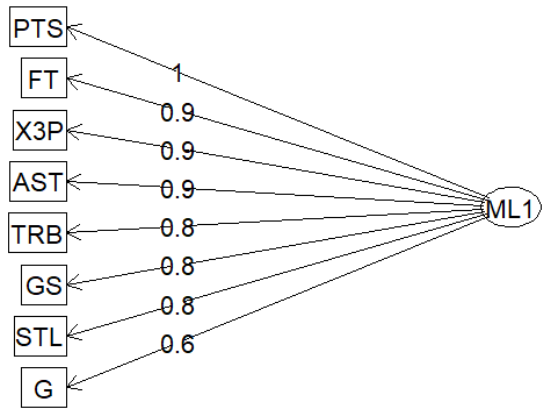


Diagrama de Fatores - PG 2023-24



Análise Paralela - PG 2023-24

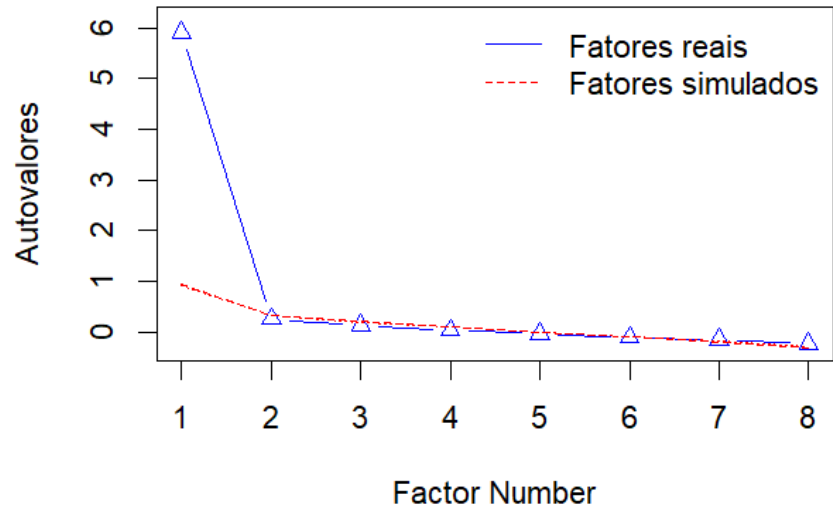
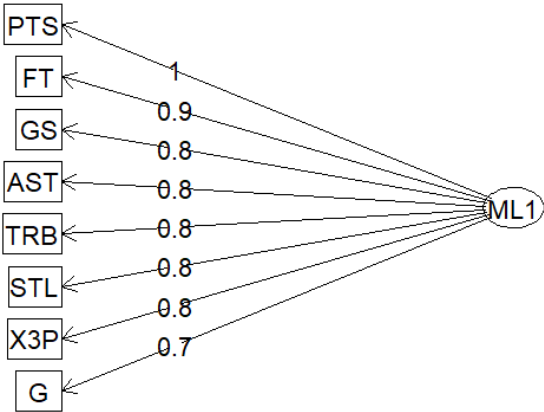


Diagrama de Fatores - SF 2023-24



Análise Paralela - SF 2023-24

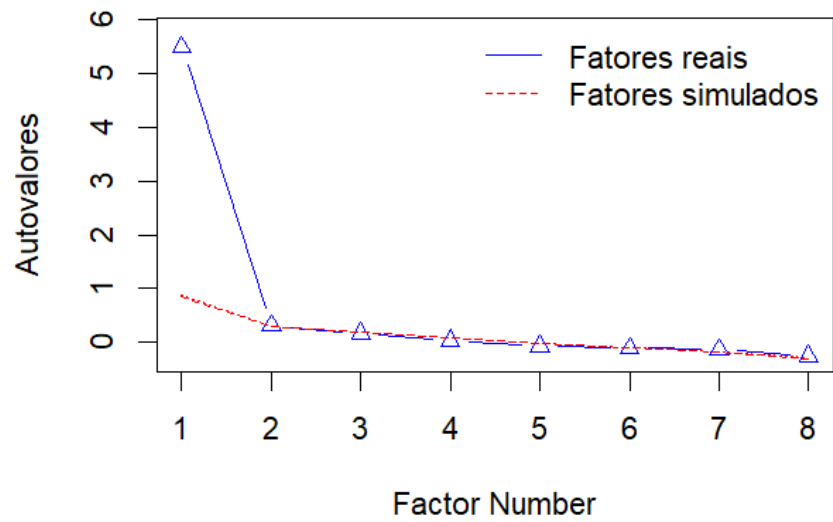
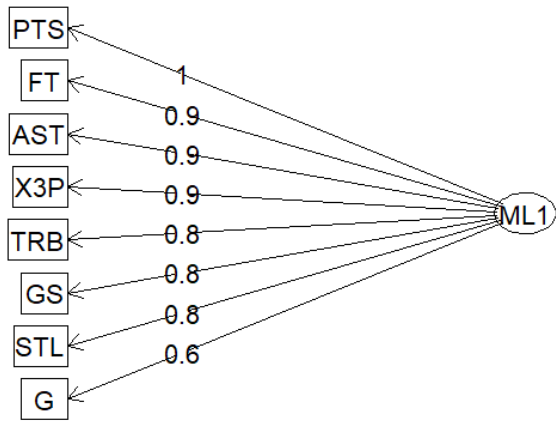
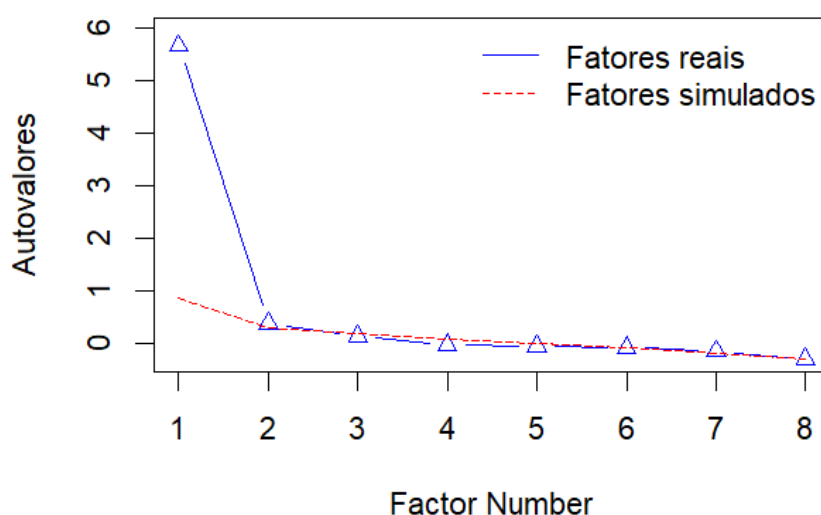


Diagrama de Fatores - SG 2023-24



Análise Paralela - SG 2023-24



Fonte: elaborado pelo autor.

8.6.6 Construção de rankings por posição 2023-2024

A partir dos resultados da AFE por posição, os rankings foram criados em duas etapas principais: cálculo dos escores fatoriais e combinação dos fatores utilizando pesos proporcionais à variância explicada.

As equações finais para os rankings de PG, SG, SF, PF e C foram formuladas com base nas variáveis mais relevantes para o desempenho de cada posição. A equação para os armadores (PG), por exemplo, foi formada pelas variáveis PTS, AST, FT, e TRB.

Tabela 30 – Top 10 C 2023-24

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Nikola Jokic	C	79	79	2737	26,39	12,35	8,96	100,00	1
2	Anthony Davis	C	76	76	2700	24,68	12,64	3,50	92,66	1
3	Domantas Sabonis	C	82	82	2928	19,43	13,66	8,21	85,04	1
4	Bam Adebayo	C	71	71	2416	19,25	10,38	3,92	69,03	1
5	Victor Wembanyama	C	71	71	2106	21,44	10,63	3,86	67,80	1
6	Jarrett Allen	C	77	77	2442	16,51	10,53	2,73	67,61	1

7	Nikola Vučević	C	76	74	2610	18,03	10,54	3,34	65,40	1
8	Alperen Şengün	C	63	63	2046	21,13	9,35	4,95	63,75	1
9	Rudy Gobert	C	76	76	2593	13,96	12,92	1,34	60,40	1
10	Chet Holmgren	C	82	82	2413	16,55	7,90	2,44	59,77	1

Fonte: elaborado pelo autor.

Tabela 31 – Top 10 PF 2023-24

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Giannis Antetokounmpo	PF	73	73	2567	30,44	11,52	6,52	100,00	1
2	Kevin Durant	PF	75	75	2791	27,09	6,60	5,04	75,23	1
3	Jayson Tatum	PF	74	74	2645	26,85	8,12	4,92	72,94	1
4	Pascal Siakam	PF	80	80	2657	21,71	7,06	4,28	71,35	1
5	Paolo Banchero	PF	80	80	2799	22,55	6,92	5,39	70,62	1
6	LeBron James	PF	71	71	2504	25,66	7,30	8,30	68,55	1
7	Zion Williamson	PF	70	70	2207	22,87	5,80	5,03	65,10	1
8	Daniel Gafford	PF	74	66	1815	11,00	7,59	1,55	62,10	1
9	Kyle Kuzma	PF	70	70	2281	22,19	6,63	4,23	58,79	1
10	Paul George	PF	74	74	2502	22,58	5,24	3,54	58,29	1

Fonte: elaborado pelo autor.

Tabela 32 – Top 10 PG 2023-24

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score norm	cluster
1	Luka Dončić	PG	70	70	2624	33,86	9,24	9,80	100,00	1
2	Shai Gilgeous-Alexander	PG	75	75	2553	30,05	5,53	6,20	93,33	1
3	Jalen Brunson	PG	77	77	2726	28,73	3,61	6,74	89,73	1
4	De'Aaron Fox	PG	74	74	2659	26,57	4,59	5,65	80,54	1
5	Stephen Curry	PG	74	74	2421	26,43	4,46	5,12	80,11	1

6	Devin Booker	PG	68	68	2447	27,07	4,53	6,94	75,09	1
7	Damian Lillard	PG	73	73	2579	24,32	4,38	6,96	74,91	1
8	Tyrese Maxey	PG	70	70	2260	25,94	3,69	6,19	73,65	1
9	Coby White	PG	79	78	2881	19,10	4,53	5,13	61,80	1
10	Tyrese Haliburton	PG	69	68	2224	20,13	3,93	10,90	58,09	1

Fonte: elaborado pelo autor.

Tabela 33 – Top 10 SF 2023-24

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	DeMar DeRozan	SF	79	79	2989	24,01	4,30	5,33	100,00	1
2	Mikal Bridges	SF	82	82	2854	19,59	4,54	3,65	83,70	1
3	Kawhi Leonard	SF	68	68	2330	23,72	6,12	3,59	81,99	1
4	Jaylen Brown	SF	70	70	2343	23,00	5,53	3,56	79,92	1
5	Franz Wagner	SF	72	72	2337	19,74	5,29	3,74	73,50	1
6	Miles Bridges	SF	69	67	2581	21,00	7,32	3,35	72,91	1
7	Brandon Ingram	SF	64	64	2103	20,81	5,06	5,67	68,57	1
8	Michael Porter Jr.	SF	81	81	2565	16,74	6,98	1,51	67,71	1
9	Klay Thompson	SF	77	63	2284	17,86	3,30	2,29	67,34	1
10	Brandon Miller	SF	74	68	2383	17,28	4,26	2,36	64,12	1

Fonte: elaborado pelo autor.

Tabela 34 – Top 10 SG 2023-24

rank	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_norm	cluster
1	Anthony Edwards	SG	79	78	2770	25,94	5,44	5,13	100,00	1
2	Dejounte Murray	SG	78	78	2783	22,47	5,33	6,44	85,10	1
3	Jalen Green	SG	82	82	2601	19,63	5,16	3,55	77,84	1
4	Kyrie Irving	SG	58	58	2030	25,64	5,00	5,16	70,71	1
5	Donovan Mitchell	SG	55	55	1942	26,60	5,09	6,05	70,39	1
6	Cam Thomas	SG	66	51	2075	22,45	3,24	2,91	70,14	1

7	Collin Sexton	SG	78	51	2075	18,69	2,63	4,91	69,97	1
8	Jordan Poole	SG	78	66	2346	17,38	2,72	4,40	64,83	1
9	Bogdan Bogdanović	SG	79	33	2401	16,87	3,44	3,13	63,37	1
10	Austin Reaves	SG	82	57	2629	15,85	4,29	5,49	62,80	1

Fonte: elaborado pelo autor.

8.6.7 Ranking geral 2023-2024

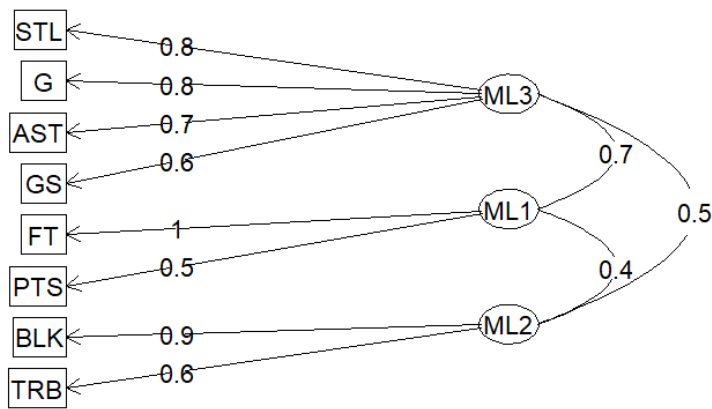
Após a construção dos rankings por posição, foi gerado um ranking geral combinando todas as posições. A variância explicada cumulativa pelos três fatores foi de 76%, o que significa que a estrutura latente explicita bem as diferenças entre os jogadores.

A equação final do índice geral foi construída a partir de variáveis como PTS, GS, AST, TRB, STL, BLK, FT.

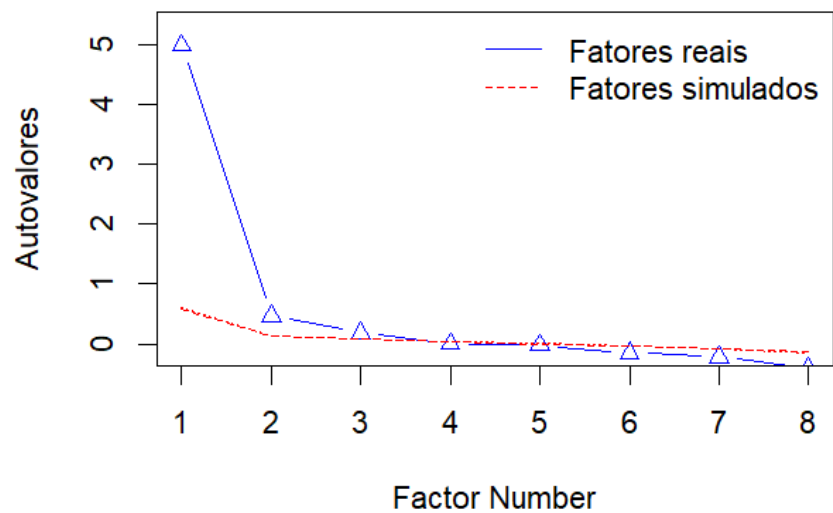
O MVP da temporada 2023-2024 foi Nikola Jokic, que ocupou a 4ª posição no ranking geral, com um score normalizado de 97.54. Esse resultado valida a eficácia do modelo de análise estatística, já que o jogador mais reconhecido oficialmente aparece no topo do ranking, confirmando a precisão do índice criado.

Figuras 32 - Diagrama de fatores, análise paralela e Top10 geral 2023-2024

Fatores - Ranking Geral 2023-24



Análise paralela - Ranking geral 2023-24



Fonte: elaborado pelo autor.

Tabela 35 – Top 10 Geral 2023-24

rank_geral	Player	Pos	G	GS	MP	PTS/g	TRB/g	AST/g	score_geral_norm	cluster
1	Anthony Davis	C	76	76	2700	24,68	12,64	3,50	100,00	1
2	Shai Gilgeous-Alexander	PG	75	75	2553	30,05	5,53	6,20	99,00	1
3	Giannis Antetokoumpo	PF	73	73	2567	30,44	11,52	6,52	98,75	1

4	Nikola Jokic	C	79	79	2737	26,39	12,35	8,96	97,54	1
5	Luka Dončić	PG	70	70	2624	33,86	9,24	9,80	95,37	1
6	Victor Wembanyama	C	71	71	2106	21,44	10,63	3,86	92,21	1
7	Domantas Sabonis	C	82	82	2928	19,43	13,66	8,21	84,52	1
8	Anthony Edwards	SG	79	78	2770	25,94	5,44	5,13	83,51	1
9	Kevin Durant	PF	75	75	2791	27,09	6,60	5,04	82,92	1
10	DeMar DeRozan	SF	79	79	2989	24,01	4,30	5,33	82,78	1

Fonte: elaborado pelo autor.

9 CONCLUSÕES

A análise estatística desenvolvida permitiu construir e avaliar um sistema de rankings de desempenho para jogadores da NBA ao longo das temporadas de 2019-20 a 2023-24, com base exclusivamente em métricas quantitativas de produção em quadra (pontos, rebotes, assistências, roubos, tocos e indicadores de eficiência de arremessos). A combinação de Análise de Componentes Principais, clusterização K-means e Análise Fatorial Exploratória forneceu uma estrutura multivariada capaz de sintetizar a informação de um grande conjunto de variáveis em escores fatoriais padronizados, utilizados tanto em rankings por posição quanto em um índice geral de desempenho.

Os modelos fatoriais gerais estimados para cada temporada apresentaram índices de ajuste compatíveis com soluções parcimoniosas, ainda que com sinais de misspecification moderada, especialmente em termos de RMSEA. O quadro a seguir resume os principais indicadores de ajuste para o modelo fatorial do ranking geral em cada ano:

Tabela 36 – Parâmetros de ajuste por temporada

Temporada	RMSEA	BIC	Tucker-Lewis (TLI)
2019-20	0,143	25,74	0,905
2020-21	0,178	69,31	0,862
2021-22	0,175	69,76	0,873
2022-23	0,162	52,42	0,890
2023-24	0,161	53,13	0,900

Fonte: elaborado pelo autor.

Em todas as temporadas, os valores de TLI situaram-se próximos ou acima de 0,86, sugerindo boa confiabilidade da solução fatorial, enquanto o RMSEA, embora acima do patamar usual de “excelente ajuste”, permaneceu em níveis aceitáveis para modelos aplicados com grande número de indicadores correlacionados e amostras reais complexas. Os valores de BIC indicam que, apesar da penalização pela complexidade, os modelos com 2 ou 3 fatores oferecem compromisso adequado entre ajuste e parcimônia, justificando o uso dos escores fatoriais como base para o índice de desempenho.

A validação externa do índice foi realizada por meio da comparação entre o ranking geral obtido e os prêmios oficiais de Most Valuable Player (MVP) de cada temporada. Na tabela a seguir, apresentam-se, para cada temporada, o MVP oficial, a posição e o score normalizado desse jogador no ranking construído, bem como o jogador que ocupa a primeira posição (“MVP do modelo”):

Tabela 37 – MVP's de cada temporada

Temporada	MVP oficial	Score	MVP ranking	Score
2019-20	Giannis Antetokounmpo	82,34 (2º)	James Harden	100,00
2020-21	Nikola Jokic	100,00 (1º)	Nikola Jokic	100,00
2021-22	Nikola Jokic	96,83 (2º)	Joel Embiid	100,00
2022-23	Joel Embiid	100,00 (1º)	Joel Embiid	100,00
2023-24	Nikola Jokic	97,54 (4º)	Anthony Davis	100,00

Fonte: elaborado pelo autor.

Observa-se elevada concordância entre o índice proposto e a avaliação subjetiva da liga: em duas temporadas (2020-21 e 2022-23) o MVP oficial também é o líder do ranking, em duas outras (2019-20 e 2021-22) aparece em 2º lugar com scores superiores a 80 em escala 0–100, e em 2023-24 permanece no Top 4 com score superior a 95. Essa proximidade sistemática indica que o modelo captura adequadamente o desempenho global dos atletas, ao mesmo tempo em que evidencia casos em que a premiação oficial privilegia narrativas contextuais (por exemplo, impacto em franquia específica) que não são diretamente contempladas pelo índice puramente estatístico.

Do ponto de vista estatístico, os resultados mostram que a maior parte da variância conjunta das principais estatísticas de boxscore pode ser sintetizada em um pequeno número de fatores latentes, frequentemente associados a dimensões interpretáveis como “produção ofensiva”, “contribuição geral” e “presença física/defensiva”. A combinação desses fatores em um único escore padronizado, ponderado pela variância explicada, produz rankings estáveis por posição e gerais, que permitem comparar jogadores que atuam em contextos distintos sem recorrer a critérios subjetivos de avaliação.

Embora os índices de ajuste indiquem espaço para refinamento – em especial na redução do RMSEA e na incorporação de variáveis contextuais e métricas avançadas adicionais –, os resultados empíricos obtidos ao longo de cinco temporadas sugerem que o modelo proposto constitui uma ferramenta estatisticamente consistente e operacionalmente simples para análise de desempenho na NBA.

REFERÊNCIAS

ALMAS, Saulo Peters. **Análise das estatísticas relacionadas ao jogo que discriminam as equipes vencedoras das perdedoras no basquetebol profissional brasileiro**. Researchgate, 2015. Disponível em: https://www.researchgate.net/publication/286510868_Analise_das_estatisticas_relacionadas_ao_jogo_que_discriminam_as_equipes_vencedoras_das_perdedoras_no_basquetebol_profissional_brasileiro. Acesso em: 27 set. 2024.

ALVES, Tiago Gonçalves De Sampaio. **Estatística no basquete**: análise de desempenho dos jogadores da Liga Nacional de Basquetebol dos Estados Unidos. UnB, 2023. Disponível em: <https://bdm.unb.br/handle/10483/37105>. Acesso em: 15 set. 2024.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 5. ed. [S.l.]: Saraiva, 2003.

CABRAL, Fernando Ramos. **A matemática por trás do jogo**: o impacto e a importância dos dados estatísticos no basquetebol. RIIFAP, 2022. Disponível em: <https://repositorio.ifap.edu.br/jspui/handle/prefix/654>. Acesso em: 27 set. 2024.

CANAN, Felipe; MENDES, José Carlos; SILVA, Rogério Vaz Da. **Análise estatística no basquetebol de base**: perfil do Campeonato Paranaense de Basquetebol masculino Sub-17. SciELO, 2015. Disponível em: <https://www.scielo.br/j/rbefe/a/JvtNP3h7YZ3qfTrH3wvzsHc/#>. Acesso em: 15 set. 2024.

CHUA, Garvyn Jay. **Statistical analysis and predictive modeling in basketball**: unveiling key variables for championship success. eScholarship, 2023. Dissertação (Mestrado em Applied Statistics & Data Science), University of California, Los Angeles, 2023. Disponível em: <https://escholarship.org/uc/item/6m2663c3>. Acesso em: 27 set. 2024.

GALERANI, Fernanda de Paula. **Montando o time ideal uma análise multivariada da NBA**. Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente, 2024. Disponível em: https://unesp.primo.exlibrisgroup.com/discovery/search?vid=55UNESP_INST:UNESP. Acesso em: 6 dez. 2024.

HAIR, J. F. et al. **Análise Multivariada de Dados**. [S.l.]: Bookman, 2009.

JOHNSON, Richard Allen; WICHERN, Dean W. **Applied multivariate statistical analysis**. 6th. ed. Delhi: Pearson Education, 2007. 773p.

JOVANOVIĆ, Mladen. **Statistical modelling for sports scientists: practical introduction using R**. Researchgate, 2019. Disponível em: https://www.researchgate.net/publication/336093323_Statistical_Modelling_for_Sports_Scientists_Practical_Introduction_Using_R_Part_1. Acesso em: 25 set. 2024.

JUNIOR, Dante De Rose. **Análise estatística de jogos de basquetebol: o fator “mando de jogo”**. Lecturas: Educación Física y Deportes, 2002. Disponível em: <https://www.efdeportes.com/efd54/basq.htm>. Acesso em: 27 set. 2024.

TIQUINHO. **A importância das estatísticas para analisar os esportes americanos**. Disponível em: <https://www.queroapostar.com/artigos/importancia-das-estatisticas-para-analisar-os-esportes-americanos/>. Acesso em: 20 out. 2024

NAÇÕES Unidas Brasil: **Objetivos de Desenvolvimento Sustentável**. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 08 out. 2024.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada uma abordagem aplicada**. [S.l.] Belo Horizonte: Editora Ufmg, 2005.