

---

Universidade Federal da Paraíba  
Centro de Ciências Exatas e da Natureza  
Departamento de Estatística

Análise do Atletismo: Uma Investigação sobre a  
Performance de Jogadores Universitários Norte  
Americanos no *combine* ao Longo do Tempo 2000 -  
2022

Murillo Ribeiro Santos

Maio/2024

---

Murillo Ribeiro Santos

**Análise do Atleticismo: Uma Investigação sobre a  
Performance de Jogadores Universitários Norte  
Americanos no *combine* ao Longo do Tempo 2000 -  
2022**

Trabalho de Conclusão do Curso apresentado no Curso de Bacharelado em Estatística da UFPB, com vistas para a conclusão da disciplina TCC II e obtenção do grau de Bacharel em Estatística da UFPB.  
Orientador: Prof. Dr. João Agnaldo do Nascimento

**João Pessoa  
Maio de 2024**



## ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

### “Análise do Atleticismo: Uma Investigação sobre a Performance de Jogadores Universitários Norte Americanos no Combine ao Longo do Tempo 2000 - 2022”

*Murillo Ribeiro Santos*

Ao décimo dia do mês de Maio de 2024 às 14h00, de modo presencial, no Laboratório Joab Lima do Departamento de Estatística, realizou-se a sessão pública de defesa do Trabalho de Conclusão de Curso do(a) discente Murillo Ribeiro Santos, matrícula 20180011901, com a Banca Examinadora composta pelos professores: Dr. João Agnaldo do Nascimento, Presidente/Orientador (Departamento de Estatística - UFPB), Dra. Juliana Freitas Pires, Examinadora (Departamento de Estatística - UFPB), Dr. Marcelo Rodrigo Portela Ferreira, Examinador (Departamento de Estatística - UFPB) e Dr. José Carlos de Lacerda Leite, Examinador Suplente (Departamento de Estatística - UFPB). Iniciando-se os trabalhos, a presidente da Banca Examinadora cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que se fizesse, oralmente, a exposição do Trabalho de Conclusão de Curso intitulado “**Análise do Atleticismo: Uma Investigação sobre a Performance de Jogadores Universitários Norte Americanos no Combine ao Longo do Tempo 2000 - 2022**”. Concluída a apresentação, a Banca Examinadora iniciou à arguição do(a) candidato(a). Encerrados os trabalhos de arguição os examinadores reuniram-se para avaliação e deram o parecer final sobre a apresentação e defesa oral do(a) candidato(a), tendo sido atribuída à sua apresentação a nota 7,3 (Sete virgula três), na disciplina de TCC II, resultante da média aritmética das notas atribuídas pelos membros da Banca Examinadora. A aprovação do(a) discente está condicionada a entrega da versão final do Trabalho de Conclusão de Curso com a inserção da ficha catalográfica e, as alterações sugeridas pelos examinadores, à Coordenação do Curso de Bacharelado em Estatística no prazo de 15 de Maio de 2024.

Documento assinado digitalmente  
**gov.br** JOAO AGNALDO DO NASCIMENTO  
Data: 15/05/2024 16:43:10-0300  
Verifique em <https://validar.iti.gov.br>

Dr. João Agnaldo do Nascimento  
(Professor Orientador)

Documento assinado digitalmente  
**gov.br** MARCELO RODRIGO PORTELA FERREIRA  
Data: 20/05/2024 13:03:56-0300  
Verifique em <https://validar.iti.gov.br>

Dr. Marcelo Rodrigo Portela Ferreira  
(Professor Examinador)

Documento assinado digitalmente  
**gov.br** JULIANA FREITAS PIRES  
Data: 21/05/2024 10:15:17-0300  
Verifique em <https://validar.iti.gov.br>

Dra. Juliana Freitas Pires  
(Professora Examinadora)

Documento assinado digitalmente  
**gov.br** MURILLO RIBEIRO SANTOS  
Data: 22/05/2024 22:19:05-0300  
Verifique em <https://validar.iti.gov.br>

Murillo Ribeiro Santos  
(Discente)

João Pessoa, 10 de Maio de 2024.

*Dedico este trabalho a minha família e aos amigos  
que fiz durante minha trajetória acadêmica.*

## AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos a todas as pessoas que contribuíram de alguma forma para a realização deste trabalho. Em primeiro lugar, agradeço a Deus por ter me proporcionado a oportunidade maravilhosa e recheada de desafios que foi esta jornada. Em meio a todo esse processo, tenho certeza que o Murillo de 2018 estaria orgulhoso do resultado desta caminhada.

A minha família, que me apoiou nessa jornada, cada um a sua forma. Tudo que eu conquistei até hoje foi graças ao esforço e dedicação deles.

Ao meu orientador Prof. Dr. João Agnaldo do Nascimento, pela orientação dedicada, apoio constante em diversos momentos e valiosas sugestões ao longo deste projeto, e gostaria de destacar o quanto foi gratificante ter a oportunidade de aprender com um profissional tão rico em conhecimento e experiência, acumulados ao longo de sua jornada acadêmica.

As minhas amigadas construídas no curso: Borges, Carol, Amanda e Gabriel pelas loucuras compartilhadas, noites em claro, momentos de apoio mútuo, surtos e comemorações ao longo desse curso.

Especialmente Borges, com quem eu compartilho um carinho enorme, sem ele toda essa jornada não teria sido possível. Agradeço pelos momentos em que fez com que eu me sentisse em casa, mesmo estando a 2.552 KMs da minha terra.

Um agradecimento especial também a Carol, minha colega de trabalho, um pedaço de SP em João Pessoa, sua amizade foi muito importante ao longo dessa jornada.

Por fim, agradeço a todos os Docentes do Departamento de Estatística que participaram da minha trajetória, seu apoio e encorajamento foram fundamentais para alcançar este objetivo.

*"Aprender uma lição sem dor não tem significado, isso por que as pessoas não conseguem obter nada sem sacrificar alguma coisa. Mas quando elas superam as dificuldades e conseguem o que querem, conquistam um coração forte que não perde para nada. Um coração forte como aço."*

*(EDWARD ELRIC; HIROMU ARAKAWA ;  
FULLMETAL ALCHEMIST: BROTHERHOOD,  
2009)*

O futebol americano é um esporte dinâmico, altamente estratégico e fisicamente exigente. Nesse contexto, o *combine* é um evento crucial para os atletas universitários garantirem sua transição para o mundo profissional. Este estudo teve como objetivo analisar os dados históricos dos testes físicos realizados no *combine* da National Football League (NFL), para identificar possíveis tendências de aumento no nível atlético dos atletas que se preparam para entrar na liga nos anos 2000 a 2022. Esta análise é baseada em dados de 6.446 jogadores, que foram coletados durante o *combine*, entre esses achados, temos, características físicas (peso e altura) e resultados dos testes aplicados pela NFL. Foram aplicados modelos lineares generalizados (MLG) e Annual Percentage Change (APC) para avaliar as tendências de crescimento no desempenho dos atletas durante o período de tempo avaliado. Ao analisar os resultados dos MLGs, concluiu-se que, no modelo de 40 jardas, o intercepto foi de aproximadamente 5,00 segundos, com valores de P-Valor extremamente baixos ( $<2e-16$ ), o que indica uma forte ligação entre as variáveis explicativas e o desempenho no teste e o pseudo- $R^2$  foi de 82,10%, o que sugere que uma grande parte das variações nos dados foi explicada pelo modelo. No entanto, no modelo do teste do supino, o intercepto foi de cerca de 21,69, com valores de P-Valor igualmente baixos, e o pseudo- $R^2$  para este modelo foi de 42,27%. Isso indica que o modelo enfrentou dificuldades para explicar os dados. No que diz respeito ao APC, foi observado um aumento médio anual de 85,39% nos resultados dos testes físicos realizados durante o *Combine* da NFL. Dessa forma, os resultados do estudo mostram uma tendência de melhora consistente no desempenho físico, com variações significativas ao longo do tempo.

**Palavras-chaves:** Combine, Athleticismo, National Football League, Modelos lineares generalizados, Annual Percentage Change.

## ABSTRACT

American football is a dynamic, highly strategic, and physically demanding sport. In this context, the combine is a crucial event for college athletes to secure their transition into the professional world. This study aimed to analyze historical data from physical tests conducted at the National Football League (NFL) combine to identify possible trends in the athletic level of athletes preparing to enter the league from the 2000s to 2022. This analysis is based on data from 6,446 players collected during the Combine, including physical characteristics (weight and height) and results of tests administered by the NFL. Generalized linear models (GLMs) and Annual Percentage Change (APC) were applied to assess growth trends in athlete performance during the evaluated time period. Upon analyzing the results of GLMs, it was concluded that in the 40-yard dash model, the intercept was approximately 5.00 seconds, with extremely low P-Values ( $<2e-16$ ), indicating a strong link between explanatory variables and test performance, and the pseudo- $R^2$  was 82.10%, suggesting that a large portion of data variations were explained by the model. However, in the bench press test model, the intercept was around 21.69, with equally low P-Values, and the pseudo- $R^2$  for this model was 42.27%. This indicates that the model struggled to explain the data. Regarding APC, an average annual increase of 85.39% was observed in the results of physical tests conducted during the NFL Combine. Thus, the study's results show a consistent improvement trend in physical performance, with significant variations over time.

**Keywords:** Combine, Athleticism, National Football League, Generalized Linear Models, Annual Percentage Change.

<b>Lista de Figuras</b>	<b>viii</b>
<b>Lista de Tabelas</b>	<b>ix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	3
1.1.1 Principal . . . . .	3
1.1.2 Específicos . . . . .	3
<b>2 Futebol Americano</b>	<b>4</b>
2.1 Apresentação . . . . .	4
2.2 Limitações . . . . .	5
2.3 Finalidade . . . . .	5
<b>3 Metodologia</b>	<b>6</b>
3.1 Modelos Lineares Generalizados (MLG) . . . . .	6
3.1.1 Métodos considerados . . . . .	7
3.2 Avaliação de métodos de MLG . . . . .	10
3.3 Análise de Componentes Principais . . . . .	10
3.3.1 APC (Annual Percentage Change) . . . . .	11
3.3.2 Processo de Análise . . . . .	13
<b>4 Conjunto de Dados</b>	<b>16</b>
4.1 Introdução . . . . .	16
4.1.1 Variáveis Independentes . . . . .	17
4.1.2 Variáveis Dependentes . . . . .	17
4.1.3 Elenco de Variáveis . . . . .	17
4.2 Análise Exploratória . . . . .	18

4.3	Pré-processamento e outros aspectos . . . . .	26
4.3.1	Primeira etapa . . . . .	26
4.3.2	Segunda etapa . . . . .	28
<b>5</b>	<b>Experimentos Numéricos</b>	<b>30</b>
5.1	Desenho experimental . . . . .	30
5.2	Resultados . . . . .	31
<b>6</b>	<b>Conclusão</b>	<b>42</b>
6.1	Conclusão . . . . .	42
6.2	Trabalhos Futuros . . . . .	44
6.3	Anexo . . . . .	45
6.3.1	Script . . . . .	45
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>46</b>

## LISTA DE FIGURAS

3.1	Análise de Componentes Principais (PCA) . . . . .	11
4.1	Análise média da evolução dos desempenhos no Teste de 40 Jardas dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	22
4.2	Análise média da evolução dos desempenhos no Teste de Supino dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	23
4.3	Análise média da evolução dos desempenhos no Teste de salto em distância dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	24
4.4	Análise média da evolução dos Desempenhos no Teste dos 3 cones dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	25
4.5	Relação entre Componentes Principais e Desempenho nos Testes Físicos de Jogadores no <i>Combine</i> (2020-2022) . . . . .	27
5.1	Comparação entre a Média Anual e o Média do Valor Predito no Teste de 40 Jardas dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	36
5.2	Comparação entre a Média Anual e o Média do Valor Predito no Teste do supino dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	37
5.3	Comparação entre a Média Anual e o Média do Valor Predito no Teste de distância dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	37
5.4	Comparação entre a Média Anual e o Média do Valor Predito no Teste dos 3 cones dos Jogadores no <i>Combine</i> (2000-2022) . . . . .	38
5.5	Desenvolvimento do Índice Padrão Médio do Modelo de 40 Jardas no <i>Combine</i> (2000-2022) . . . . .	38
5.6	Desenvolvimento do Índice Padrão Médio do Modelo de supino no <i>Combine</i> (2000-2022) . . . . .	39
5.7	Desenvolvimento do Índice Padrão Médio do Modelo de salto em distância no <i>Combine</i> (2000-2022) . . . . .	39

5.8	Desenvolvimento do Índice Padrão Médio do Modelo de 3 cones no <i>Combine</i> (2000-2022) . . . . .	40
5.9	Desenvolvimento do Índice Global Padrão Médio dos Modelos MLG no <i>Combine</i> (2000-2022) . . . . .	40

## LISTA DE TABELAS

4.3	Médias das alturas e pesos dos jogadores por posição no <i>Combine</i> (2000-2022)	19
4.4	Médias dos resultados físicos dos testes dos jogadores por Posição no <i>Combine</i> (2000-2022)	21
4.5	Conversões de Unidades	26
4.6	Componentes da Altura e Peso (Sem padronização)	26
4.7	Modelos Lineares Generalizados (MLG)	28
4.8	Coefficiente dos índices na construção do índice global	29
5.1	Variáveis associadas ao desempenho no Teste de 40 Jardas dos Jogadores no <i>Combine</i> (2000-2022)	33
5.2	Variáveis associadas ao desempenho no Teste do supino dos Jogadores no <i>Combine</i> (2000-2022)	33
5.3	Variáveis associadas ao desempenho no Teste do salto em distância dos Jogadores no <i>Combine</i> (2000-2022)	34
5.4	Variáveis associadas ao desempenho no Teste do tempo dos 3 cones dos Jogadores no <i>Combine</i> (2000-2022)	35
5.5	Métricas associadas ao desempenho dos modelos no conjunto de validação	35

## LISTA DE ABREVIATURAS

**NCAA** National Collegiate Athletic Association

**NBA** National Basketball Association

**NFL** National Football League

**PCA** Análise de Componentes Principais

**MLG** Modelos Lineares Generalizados

**APC** Annual Percentage Change

**FBS** Football Bowl Subdivision

**FCS** Football Championship Subdivision

**PRF** Pro Football Reference

**QB** Quarterback

**P** Punter

**k** Kicker

**FB** Fullback

**LS** Long snapper

**C** Centers

**CB** Cornerbacks

**DB** Defensive Backs

**DE** Defensive Ends

**DL** Defensive Linemen

**DT** Defensive Tackles

**EDGE** Edge Rushers

**ILB** Inside Linebackers

**LB** Linebackers

**OG** Offensive Guards

**OL** Offensive Linemen

**OLB** Outside Linebackers

**OT** Offensive Tackles

**RB** Running Backs

**S** Safeties

**TE** Tight Ends

**WR** Wide Receivers

O futebol americano é um esporte que possui uma construção cultural diferente da que encontramos no futebol, trata-se de um esporte dinâmico, altamente estratégico e fisicamente exigente. A NFL (National Football League), a liga profissional de futebol americano dos Estados Unidos, é a referência mundial quando se fala no esporte.

Devido à natureza da NFL, onde os jogadores que participam da liga são selecionados nas universidades, é notável que a jornada de um jogador rumo à liga profissional é uma trilha árdua e altamente competitiva. A transição da universidade para a NFL é um fenômeno complexo e desafiador, que envolve não apenas o talento atlético, mas também uma série de fatores adicionais, tais como oportunidades de exposição, apoio financeiro e capacidade de gestão do tempo (STOKOWSKI; PAULE-KOBA; KAUNERT, 2019).

Para muitos jovens talentosos, o sonho de se tornar um jogador profissional é alimentado pelos exemplos de sucesso encontrados em ligas profissionais. Entretanto, o caminho para alcançar esse sonho é bastante árduo e muitas vezes imprevisível. Sendo assim, uma das principais rotas é através do sistema universitário, lugar em que os atletas têm a possibilidade de aprimorar suas habilidades enquanto recebem uma educação acadêmica (STAMBULOVA, 2009; STAMBULOVA; ALFERMANN; STATLER, 2009).

As instituições de ensino superior norte-americanas são amplamente reconhecidas por dominarem consistentemente as classificações e rankings das melhores universidades globais. Conforme destacado por Pati e Gasparini (2016), uma análise da *Times Higher Education* revelou que oito das dez principais universidades do mundo estão sediadas em solos norte-americanos.

No entanto, por trás do brilho dos holofotes, existe uma realidade financeira complexa, uma necessidade urgente de educação e uma compreensão essencial da raridade das oportunidades que o esporte profissional oferece.

Ao analisar um estudo realizado pela *National Collegiate Athletic Association* (NCAA) em 2013, intitulado "*Estimated Probability of Competing in Athletics Beyond the High School Interscholastic Level*", nota-se que, conforme os dados analisados, as chances de atletas universitários se tornarem profissionais é uma conquista rara.

Embora o estudo não forneça números exatos, a pesquisa ressalta a baixa probabilidade de tal transição ocorrer, enfatizando a importância de investir em educação e desenvolvimento pessoal durante o período universitário como preparação para um futuro além do esporte competitivo.

A etapa mais importante nesse processo de transição é o *draft*, um evento fundamental no calendário da NFL, no qual as equipes têm a oportunidade de selecionar os jogadores universitários mais promissores para integrarem suas equipes. Esta seleção é crucial para o futuro das franquias, pois ajuda a garantir uma base sólida de talentos e a construir equipes competitivas (GE, 2021).

Na esfera do futebol universitário, uma parcela significativa de jogadores da FBS (*Football Bowl Subdivision*) e FCS (*Football Championship Subdivision*) mantém a esperança de se tornarem profissionais em seu esporte, com 64% e 50%, respectivamente, expressando essa crença. No entanto, em contraste com essas expectativas, os dados de 2017 revelaram que apenas uma pequena fração, meros 1,6%, foram efetivamente recrutados por equipes da NFL durante o *draft* (NCAA, 2016; NCAA, 2018).

Deste modo, em meio as etapas realizadas para se tornar um jogador profissional, existem eventos como testes e avaliações necessárias para avaliar as capacidades dos jogadores, entre esses eventos está o ***Combine***, uma avaliação abrangente organizada pela NFL, uma vitrine crucial para os talentos emergentes demonstrarem suas habilidades físicas e cognitivas (AMORIM, 2024).

O ***Combine*** é um evento decisivo para a NFL, que acontece anualmente durante uma semana, onde os melhores estudantes universitários são submetidos a uma série de testes físicos, avaliações técnicas e entrevistas para demonstrar suas capacidades e potencial para os times da NFL. Desde o início do século XXI, o ***Combine*** se tornou uma ferramenta essencial para os atletas universitários demonstrarem suas habilidades técnicas e atléticas diante de observadores, treinadores e executivos da NFL (Pro Football Network, 2024).

Em visto disso, à medida que os jovens atletas continuam a perseguir seus sonhos de alcançar o nível profissional nos esportes, surge uma questão crucial: houve uma melhora no desempenho atlético dos jogadores nos testes físicos do *Combine* com o passar do tempo? Desta forma, analisar os dados históricos do *Combine* nos permite entender melhor a evolução do desempenho atlético e identificar tendências que podem influenciar as decisões de recrutamento, oferecendo oportunidades para aprimorar ainda mais o desenvolvimento dos futuros profissionais nos campos.

## 1.1 Objetivos

### 1.1.1 Principal

Analisar os dados históricos dos testes físicos do *Combine* da NFL para identificar se há uma tendência de aumento do nível atlético dos atletas que estão se preparando para ingressar na liga, de 2000 a 2022.

### 1.1.2 Específicos

- Realizar uma análise comparativa entre os dados atuais e históricos para identificar possíveis tendências ou mudanças no desempenho atlético ao longo do tempo;

## 2.1 Apresentação

O futebol americano atrai milhões de fãs em todo o mundo, originado nos Estados Unidos e conhecido como *football*, teve suas raízes estabelecidas em solo americano no final do século XIX, passando por várias evoluções ao longo do tempo, tornando-se um dos esportes mais populares e lucrativos do país. De acordo com os registros do Livro de Recordes e Fatos da National Football League (NFL, 2020, p. 344), o primeiro embate documentado desse esporte remonta a 1869, quando as equipes das prestigiosas universidades de Princeton e Rutgers se enfrentaram.

Ao examinar a popularidade do futebol americano, é evidente que ele desfruta de um status proeminente nos Estados Unidos. Conforme revelado por pesquisas recentes, em 2017, o futebol americano emergiu como o esporte favorito de 37% dos participantes, superando o basquete com 11%, o beisebol com 9% e o futebol com 7%. Além disso, o estudo destacou que o futebol americano não só lidera em termos de popularidade geral, mas também se sobressai em segmentos demográficos, como idade e gênero. Por exemplo, entre os entrevistados do sexo masculino, 42% expressaram que o futebol americano é sua modalidade esportiva preferida, enquanto 10% mostraram preferência pelo beisebol e 9% pelo basquete (GALLUP, 2018).

Um dos aspectos mais marcantes do futebol americano é sua natureza altamente física, sendo o atletismo fundamental para o sucesso dos jogadores. O jogo é conhecido por sua intensidade, estratégia e espetáculo, envolvendo uma combinação única de inteligência,

força, velocidade, habilidade e agilidade.

A notável popularidade do futebol americano se traduz em uma liga de grande relevância competitiva e impacto de marca significativo. A NFL, por exemplo, gera anualmente uma receita impressionante de USD 10 bilhões, sendo a maioria proveniente de contratos televisivos, totalizando USD 6 bilhões. Embora já tenha um poder considerável nos Estados Unidos, a NFL busca expandir seus horizontes, visando capitalizar o mercado global em crescimento para os direitos de transmissão esportiva. Essa estratégia visa não apenas aumentar sua base de fãs internacionalmente, mas também impulsionar suas receitas anuais (RUGG, 2016).

Ao explorar as nuances do futebol americano, é interessante traçar paralelos com outros esportes, como o futebol tradicional. Embora diferentes em muitos aspectos, ambos os jogos compartilham a importância do trabalho em equipe, estratégia e habilidade individual para alcançar o sucesso.

## 2.2 Limitações

Os estudos realizados enfrentam algumas restrições que devem ser levadas em conta. Primeiramente, a avaliação depende da disponibilidade e qualidade dos registros históricos dos testes físicos do *Combine* da NFL. Sendo assim, possíveis erros ou inconsistências nos dados podem comprometer a precisão das conclusões.

Adicionalmente, o período de análise deste atual estudo que se estende de 2000 a 2022, pode não abranger todas as mudanças significativas ocorridas no cenário da NFL, o que pode limitar a abrangência das conclusões. Além disso, fatores externos não contemplados neste estudo, como lesões de jogadores-chave, mudanças nas regras do jogo ou eventos históricos, podem interferir nos resultados da análise de forma não previsível.

## 2.3 Finalidade

Por sua vez, é necessário entender cada vez mais e compreender as tendências no desempenho físico dos atletas de futebol americanos ao longo das últimas duas décadas. Além disso, o objetivo é fornecer informações valiosas para os especialistas em esportes, treinadores, atletas e pesquisadores interessados em aprimorar o desempenho do atletismo no cenário de competição profissional.

### 3.1 Modelos Lineares Generalizados (MLG)

Os Modelos Lineares Generalizados (MLGs), também conhecidos como modelos exponenciais lineares, foram criados por Nelder e Wedderburn em 1972, com a família exponencial uniparamétrica (NELDER; WEDDERBUN, 1972). Esses modelos são uma abordagem estatística que combina elementos da teoria de regressão linear e da teoria de modelos lineares generalizados, fornecendo uma estrutura flexível para a análise de dados.

Um modelo linear é caracterizado por uma distribuição de probabilidade para a variável resposta, pertencente à família exponencial de distribuições, que inclui um conjunto de variáveis independentes que descrevem a estrutura linear do modelo.

Adicionalmente, os modelos de MLGs possuem uma ligação que conecta a média da variável de resposta à estrutura linear. Essa função de ligação é crucial na modelagem, pois relaciona a escala em que os preditores operam à escala da variável resposta, permitindo uma interpretação adequada dos efeitos das variáveis independentes sobre a variável resposta.

Os MLGs se tornaram uma ferramenta indispensável em diversas áreas, como epidemiologia, ciências sociais e biomédica, onde a compreensão das conexões entre variáveis é crucial para inferências precisas. Uma fonte relevante para o entendimento dos MLGs é o livro "Generalized Linear Models", de McCullagh e Nelder (1989), que apresenta uma ampla cobertura teórica e prática desses modelos.

### 3.1.1 Métodos considerados

No âmbito dos métodos estatísticos, os Modelos Lineares Generalizados (MLGs) desempenham um papel crucial na análise de dados devido à sua flexibilidade e aplicabilidade em diversas situações. Similar à diversidade de técnicas encontradas em métodos de aprendizado de máquina, os MLGs abrangem uma variedade de modelos, cada um adaptado para lidar com diferentes distribuições de dados e estruturas de relacionamento entre as variáveis. Dentre os métodos considerados, pode-se citar os modelos baseados em Família Gama, Família Gaussiana, Família Quasipoisson e Família Gaussiana Inversa, destacando suas formulações matemáticas, funcionalidades e aplicações em análise estatística.

A Família Gama é bastante empregada em Modelos Lineares Generalizados (MLGs) devido à sua flexibilidade e capacidade de modelar uma variedade de distribuições de dados. A formulação matemática de um modelo da Família Gama é determinada pela distribuição gama como uma função de verossimilhança, sendo especialmente útil para variáveis de resposta positivas e assimétricas. A função de ligação identidade é empregada com modelos da Família Gama para manter a interpretação direta dos coeficientes (MCCULLAGH; NELDER, 1989).

A função de densidade de probabilidade da distribuição gama para uma variável aleatória  $Y$  é dada por:

$$f(y; \mu, \phi) = \frac{1}{\Gamma(\phi)\mu^\phi} y^{\phi-1} e^{-\frac{y}{\mu}}, \quad y > 0, \mu > 0, \phi > 0$$

onde:

- $\mu$  é a média da distribuição,
- $\phi$  é o parâmetro de dispersão (conhecido como parâmetro de forma na distribuição gama),
- $\Gamma(\cdot)$  é a função gama,
- $e$  é a base do logaritmo natural.

A função de ligação identidade é comumente utilizada com a Família Gama, e é dada por:

$$g(\mu) = \mu$$

A Família Gaussiana, também conhecida como distribuição normal, é bastante utilizada em Modelos Lineares Generalizados (MLGs) devido às suas propriedades bem conhecidas e à sua aplicabilidade em uma variedade de situações (DOBSON; BARNETT, 2008). A fórmula matemática da Família Gaussiana é determinada pela função de densidade de probabilidade da distribuição normal:

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty < y < \infty,$$

Onde:

- $\mu$  é a média da distribuição,
- $\sigma^2$  é a variância da distribuição,
- $\pi$  é uma constante matemática,
- $e$  é a base do logaritmo natural.

Sendo assim, usada com frequência em situações em que os dados são simétricos e são distribuídos de forma normal. Ela é particularmente útil para modelos de regressão quando a variável resposta é constante e tem uma distribuição próxima da normalidade.

A Família Quasipoisson é uma extensão do Modelo de Poisson e é frequentemente empregada em Modelos Lineares Generalizados (MLGs) quando os dados apresentam superdispersão, ou seja, quando a variação excede a expectativa do modelo de Poisson clássico. A formulação matemática da Família Quasipoisson é semelhante à do modelo de Poisson, mas agrega um fator adicional para inibir a superdispersão (ZEILEIS; KLEIBER; JACKMAN, 2008).

A função de densidade de probabilidade da distribuição Quasipoisson é dada por:

$$f(y; \mu, \theta) = \frac{e^{-\mu} \mu^y}{y!} \frac{1}{1 + \frac{\theta \mu}{1 - e^{-\theta \mu}}}, \quad y = 0, 1, 2, \dots$$

onde:

- $\mu$  é a média da distribuição,
- $\theta$  é o parâmetro adicional que controla a superdispersão,

- $e$  é a base do logaritmo natural,
- $y!$  é o fatorial de  $y$ .

Deste modo, a Família Quasipoisson é particularmente vantajosa em circunstâncias em que os dados apresentam uma discrepância excessiva em relação ao modelo de Poisson usual, como é frequente em dados de contagem em que há sobredispersão.

A Família Gaussiana Inversa é uma das distribuições disponíveis nos Modelos Lineares Generalizados (MLGs) e é especialmente útil para lidar com dados que mostram assimetria e valores extremos. A formulação matemática da Família Gaussiana Inversa é baseada na distribuição gaussiana padrão, porém com a inversão dos valores (SCHRODINGER, 1915).

A função de densidade de probabilidade da distribuição Gaussiana Inversa é dada por:

$$f(y; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} e^{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}}, \quad y > 0,$$

Onde:

- $\mu$  é a média da distribuição,
- $\lambda$  é o parâmetro de forma da distribuição,
- $\pi$  é uma constante matemática,
- $e$  é a base do logaritmo natural.

Essa distribuição é usada com frequência em modelagem estatística quando os dados têm uma cauda longa e valores extremos, o que pode ser ignorado por distribuições simétricas, como a gaussiana padrão.

Sendo assim, empregada em diversas áreas, como finanças, biologia e avaliação de riscos. Em particular, ela é útil em situações em que é necessário modelar dados com uma grande variação e com valores extremos.

## 3.2 Avaliação de métodos de MLG

Avaliar a eficiência e a adequação dos métodos de MLG é crucial para obter dados precisos e confiáveis em análises estatísticas. Nesse contexto, foram usadas diversas técnicas de modelagem, cada uma caracterizada por diferentes famílias de distribuições e funções de ligação. Os modelos foram previamente validados a partir da função desvio ( $D^*$ ), onde  $D^* \sim \chi^2_{(n-p, \alpha)}$ , se  $D^*$  for menor que o valor crítico, então o modelo é estatisticamente significativo, posteriormente os modelos foram avaliados utilizando o cálculo e análise da raiz do erro quadrático médio, juntamente com o Pseudo-R<sup>2</sup>.

## 3.3 Análise de Componentes Principais

A Análise de Componentes Principais (PCA) é uma técnica estatística usada para simplificar conjuntos de dados complexos, mantendo o máximo possível de informação. Ela busca identificar os padrões mais importantes nos dados, conhecidos como "componentes principais". Esses componentes representam as direções de maior variação nos dados. Em termos mais simples, a PCA ajuda a encontrar as características mais significativas dos dados, permitindo uma compreensão mais clara do conjunto de dados como um todo (Abdi; Williams, 2010).

### Fórmulas Principais:

- Matriz de Covariância Amostral

Antes de calcular os componentes principais, é necessário calcular a matriz de covariância amostral dos dados originais. A covariância amostral entre duas variáveis  $X_i$  e  $X_j$  é dada por:

$$\text{Cov}(X_i, X_j) = \frac{1}{n-1} \sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)$$

onde  $n$  é o número de observações e  $\bar{X}_i$  e  $\bar{X}_j$  são as médias das variáveis  $X_i$  e  $X_j$ , respectivamente.

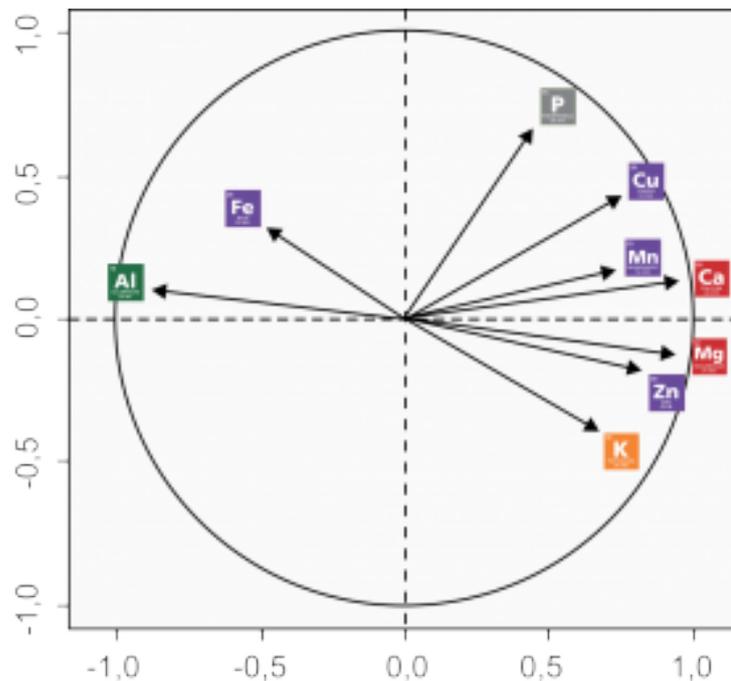
- Autovalores e Autovetores

Após a matriz de covariância, os autovalores e autovetores são calculados. Os autovetores representam as direções dos componentes principais e os autovalores são a quantidade de variação explicada por cada componente.

- Projeção dos Dados

Os dados originais são projetados nos elementos principais, resultando em uma representação dos dados em um espaço de menor dimensionalidade.

Figura 3.1: Análise de Componentes Principais (PCA)



Fonte: *Statplace*

### 3.3.1 APC (Annual Percentage Change)

Annual Percentage Change (APC) é uma medida resumo empregada para avaliar a relevância das variações ao longo do tempo em uma série temporal, expressando essas mudanças como uma porcentagem anual. Este teste é frequentemente empregado em análises de tendências e variações anuais em dados temporais, permitindo identificar padrões e comportamentos ao longo do tempo (ANTUNES; CARDOSO, 2015).

A análise de séries temporais pode revelar tendências crescentes, decrescentes ou estacionárias. Para estimar essas tendências, é comum ajustar funções matemáticas aos pontos observados, utilizando a regressão linear como uma ferramenta fundamental.

A regressão linear é uma técnica estatística utilizada para modelar a relação entre uma variável dependente  $Y$  e uma ou mais variáveis independentes  $X$ . O objetivo é encontrar a reta de melhor ajuste que minimize a distância entre os pontos observados e os valores previstos. Essa reta é representada pela equação:

$$Y = b_0 + b_1X$$

Onde:

- $Y$  é a variável dependente que queremos prever,
- $X$  é a variável independente que usamos para prever  $Y$ ,
- $b_0$  é o intercepto da reta, representando o valor de  $Y$  quando  $X$  é zero,
- $b_1$  é o coeficiente angular da reta, representando a mudança média em  $Y$  para uma unidade de mudança em  $X$ .

Para quantificar a variação ao longo do tempo, é preferível estimar a taxa percentual de mudança. Para isso, aplicamos a transformação logarítmica aos valores da série temporal, o que não apenas facilita a comparação com diferentes escalas, mas também reduz a heterogeneidade de variância dos resíduos da análise de regressão.

Ao aplicar a transformação logarítmica, podemos expressar a relação entre os valores consecutivos da série temporal da seguinte forma:

$$\log(Y_i) = b_0 + b_1X_i$$

$$\log(Y_{i+1}) = b_0 + b_1X_{i+1}$$

Diferenciando essas duas equações, obtemos:

$$\log(Y_{i+1}) - \log(Y_i) = b_1(X_{i+1} - X_i)$$

Por propriedades dos logaritmos, podemos reescrever esta equação como:

$$\log\left(\frac{Y_{i+1}}{Y_i}\right) = b_1$$

O que implica:

$$\frac{Y_{i+1}}{Y_i} = 10^{b_1}$$

Portanto, a taxa de mudança pode ser estimada como:

$$\text{APC} = (10^{b_1} - 1) \times 100\%$$

$$\text{IC95\%} = [(10^{b_{1\min}} - 1) \times 100\%, (10^{b_{1\max}} - 1) \times 100\%]$$

Essa expressão nos fornece a taxa de mudança anual (ou outra unidade de tempo relevante), e o intervalo de confiança pode ser calculado aplicando o intervalo de confiança do coeficiente  $b_1$ .

O APC calculado é usado para testar se há uma mudança significativa na variável ao longo do tempo. Isso é feito comparando o APC observado com um valor de referência, como zero (sem mudança). Se o APC for significativamente diferente de zero, isso indica que há uma mudança significativa na variável ao longo do tempo (ANTUNES; CARDOSO, 2015).

### 3.3.2 Processo de Análise

Durante o processo de análise, os índices dos modelos foram calculados e, em seguida, padronizados em uma escala de 0 a 100. Essa padronização foi realizada visando facilitar a compreensão e análise dos resultados. O índice indicará o desempenho do jogador em relação ao esperado, onde números próximos a 100 indicam um desempenho muito acima do esperado, enquanto resultados próximos de 0 sugerem um desempenho muito abaixo do esperado.

As etapas realizadas foram as seguintes:

- Cálculo dos índices por modelo

$$R_i = (y_i - \hat{u}_i)$$

onde:

- $y_i$  é o valor da  $i$ -ésima observação
- $\hat{u}_i$  é o valor da  $i$ -ésima observação estimado pelo modelo

$$\text{Índice}_i = \frac{R_i - \min(R)}{\max(R) - \min(R)}$$

- Avaliação dos índices ao longo do tempo

Para avaliar a evolução dos índices ao longo do tempo, foi calculado, a média anual dos índices e analisado o comportamento ao longo dos anos. Deste modo, essa análise permitiu identificar tendências gerais e/ou padrões específicos em determinados períodos.

- Criação do índice Global

Para criação do índice global, foi realizado uma análise dos resultados de todos os modelos, sendo assim, algumas etapas foram realizadas, sendo elas: aplicação da técnica de PCA nos índices dos modelos, seguida pela padronização dos resultados obtidos na etapa anterior. O índice global é a padronização da primeira componente gerada na análise, sendo assim, uma combinação linear dos quatro índices anteriores.

- Avaliação do índice global ao longo do tempo

Ao avaliar o comportamento do índice global ao longo do tempo, foi calculado a média anual do índice e analisado sua evolução ao longo dos anos. Desta forma, a fim de garantir a obtenção de resultados confiáveis e de qualidade, foi utilizado a medida resumo do APC para determinar a tendência da série temporal do índice global.

Em suma, o processo de análise envolveu as etapas descritas anteriormente, com objetivo de avaliar o desempenho dos modelos em relação aos dados coletados. Inicialmente, foram calculados e estabelecidos padrões para simplificar a compreensão e comparação dos resultados, logo em seguida, foram examinadas as tendências ao longo do tempo, tanto dos índices de cada modelo quanto do índice global, a fim de identificar padrões e comportamentos específicos, fornecendo assim, uma visão ampla do desempenho geral dos modelos, garantindo a obtenção de resultados confiáveis e de qualidade.

## 4.1 Introdução

Neste presente estudo, a população foi composta por jogadores maiores de 18 anos que concluíram o ensino médio após no mínimo 3 anos e receberam o convite pelo comitê de seleção do evento, composto por membros da equipe da NFL e por representantes das equipes ( $n = 6.446$ ), os quais estavam incluídos na base de dados do *Pro Football Reference* (PFR), referentes ao período de 2000 a 2022. Compuseram a amostra jogadores que atenderam os critérios de elegibilidade, sendo os critérios de inclusão referentes a: jogadores do sexo masculino, no período de 2000 a 2022. Foram adotados como critérios de exclusão: jogadores listados nas seguintes posições: "QB"(quarterback), "P"(punter), "K"(kicker), "FB"(fullback) e "LS"(long snapper), jogadores que não tiveram suas alturas medidas durante o processo de coleta de dados e os jogadores que participaram do *Combine* em 2021 devido à influência potencial da pandemia de COVID-19 nos resultados do teste.

Os dados coletados durante o *Combine*, que incluem dados físicos como altura, peso, velocidade, agilidade e força, fornecem uma valiosa perspectiva sobre o potencial atlético e a adequação de um jogador para as demandas do futebol americano profissional. Os dados são frequentemente examinados por equipes da NFL e especialistas em recrutamento para auxiliar na tomada de decisões no *Draft* da NFL, quando os times selecionam novos jogadores para suas equipes.

Dessa forma, os dados coletados do site *Pro Football Reference*, por meio do *Kaggle*, oferecem uma grande variedade de dados sobre os atletas que participaram do *Combine*

entre 2000 e 2022. Esses dados são fundamentais para pesquisadores, analistas e amantes do futebol americano, pois dão *insights* valiosos sobre as tendências atléticas, o desempenho dos jogadores e as mudanças ao longo do tempo no perfil dos atletas que desejam uma carreira na NFL.

Esta análise exploratória dos dados do *Combine* da NFL pode oferecer uma compreensão mais aprofundada das características físicas e competências dos jogadores, bem como informações sobre como esses atributos podem influenciar o êxito de um atleta ao nível profissional. Ao analisar esses dados, podemos identificar padrões, tendências e correlações que ajudam a orientar as estratégias de seleção das equipes da NFL e a compreensão geral do panorama do futebol americano profissional.

#### 4.1.1 Variáveis Independentes

- a) Altura: medida em polegadas e convertida para centímetros (cm).
- b) Peso: medida em libras e convertida para quilogramas (kg).
- c) Posição: posição do participante.

#### 4.1.2 Variáveis Dependentes

- a) Resultado do teste de 40 jardas: tempo de corrida, medido em segundos.
- b) Resultado do teste de salto em distância: medida originalmente em polegadas e convertida para centímetros (cm).
- c) Resultado do teste do supino: quantidade de repetições.
- d) Resultado do teste dos 3 cones: teste de agilidade, medido em segundos.

#### 4.1.3 Elenco de Variáveis

As informações utilizadas neste estudo foram extraídas do *Pro Football Reference*, acessadas através do portal *Kaggle*. Com base nessas informações, foram selecionadas as variáveis que permitiram analisar características dos jogadores do *Combine*. Estas variáveis incluíram: altura (medida em polegadas e convertida para centímetros (cm)),

peso (medida em libras e convertida para quilogramas (kg)), posição (posição do atleta), conforme visualizado no Quadro 1.

Quadro 1 – Detalhamento das variáveis independentes - (*Pro Football Reference*).

Variáveis do estudo	Classificação
Altura	Númerica
Peso	Númerica
Posição	Nominal

**Fonte:** *Elaborado pelo autor.*

As variáveis dependentes representam resultados importantes dos jogadores, utilizadas para analisar seu desempenho. Estas variáveis incluem o resultado do teste de 40 jardas (corrida) medido em segundos, resultado do teste de salto em distância medido em centímetros, resultado do teste do supino medido em pela quantidade de repetições e resultado do teste dos 3 cones (agilidade) medido em segundos, conforme visualizado no Quadro 2

Quadro 2 – Detalhamento das variáveis dependentes - (*Pro Football Reference*).

Variáveis do estudo	Classificação
Resultado do teste de 40 jardas	Númerica
Resultado do teste de salto em distância	Númerica
Resultado do teste do supino	Númerica
Resultado do teste dos 3 cones	Númerica

**Fonte:** *Elaborado pelo autor.*

## 4.2 Análise Exploratória

Durante as últimas duas décadas, o *Combine* da NFL tem sido um marco crucial na avaliação do potencial atlético de aspirantes a jogadores profissionais de futebol americano. Ao analisar os dados históricos dos testes físicos realizados entre 2000 e 2022, podemos verificar se houve uma tendência discernível em relação ao nível atlético dos participantes ao longo do tempo.

Esta análise exploratória visa identificar possíveis padrões ou mudanças no desempenho físico dos atletas, oferecendo informações valiosas sobre o cenário evolutivo dos talentos emergentes na NFL. Ao aprofundarmos nossa compreensão dessas tendências,

podemos contribuir para uma visão mais aprofundada sobre a preparação e as exigências físicas dos atletas que almejam o sucesso na liga profissional de futebol dos Estados Unidos.

A partir dos dados analisados e aplicação dos critérios de elegibilidade e exclusão nos dados do Combine da NFL entre 2000 e 2022, foi obtido um total de 6.446 atletas. Observa-se que, em média, esses atletas apresentaram um peso de 111,41 kg ( $\pm 21,20$ ) e uma altura de 187,49 cm ( $\pm 6,82$ ). Ao analisar por posição, destaca-se os *Offensive Tackles* (OTs) com um peso médio de 143,01 kg ( $\pm 6,09$ ) e uma altura média de 197,14 cm ( $\pm 3,11$ ). Já os *Cornerbacks* (CBs) apresentaram um peso médio de 87,55 kg ( $\pm 4,09$ ) e uma altura média de 180,95 cm ( $\pm 4,03$ ).

Nos testes físicos realizados durante o *Combine*, as médias foram de 4,77 segundos para o tiro de 40 jardas, 20 repetições no supino, 291,47 cm no salto em distância e 7,30 segundos para o teste dos 3 cones. Esses dados proporcionam uma visão abrangente do desempenho físico dos atletas que participaram durante o período avaliado, permitindo-nos identificar possíveis tendências ou mudanças no perfil atlético dos atletas da NFL.

Ao analisar a Tabela 4.3, é possível observar que a média das alturas e pesos dos jogadores por posição no período de 2000 a 2022, evidenciam que as posições de Offensive Linemen (OL) e Offensive Tackles (OT) apresentam as maiores médias de altura, com 194,06 cm e 197,14 cm, respectivamente. Por outro lado, a posição de Cornerbacks (CB) possui a menor média de altura, com 180,95 cm. Quanto ao peso, Offensive Guards (OG) e Offensive Tackles (OT) apresentam as maiores médias, com 142,67 kg e 143,01 kg, respectivamente, enquanto Cornerbacks (CB) e Wide Receivers (WR) apresentam as menores médias de peso, com 87,55 kg e 91,56 kg, respectivamente.

Tabela 4.3: Médias das alturas e pesos dos jogadores por posição no *Combine* (2000-2022)

Posição	Peso (kg)	Altura (cm)
Centers (C)	137,37	191,19
Cornerbacks (CB)	87,55	180,95
Defensive Backs (DB)	92,08	186,69
Defensive Ends (DE)	121,99	192,55
Defensive Linemen (DL)	131,98	191,18
Defensive Tackles (DT)	139,19	190,32
Edge Rushers (EDGE)	114,37	191,83
Inside Linebackers (ILB)	109,54	186,16
Linebackers (LB)	107,62	187,16

Offensive Guards (OG)	142,67	193,07
Offensive Linemen (OL)	142,3	194,06
Outside Linebackers (OLB)	108,95	187,41
Offensive Tackles (OT)	143,01	197,14
Running Backs (RB)	97,1	179,08
Safeties (S)	93,93	183,5
Tight Ends (TE)	115,13	193,8
Wide Receivers (WR)	91,56	184,82

**Fonte:** *Elaborado pelo autor*

Com base nas descobertas do estudo de Robbins (2011), que analisou os requisitos físicos para ser convocado para a National Football League (NFL), fica claro que as diferenças nas características físicas dos jogadores por posição na NFL são fundamentais para o desempenho atlético e a seleção de talentos. Robbins quantificou dados do Combine da NFL entre 2005 e 2009, examinando nove avaliações de desempenho físico para 15 posições distintas. Os resultados mostraram diferenças significativas nas avaliações de desempenho entre as posições, com posições ofensivas e defensivas com características físicas distintas.

As posições de ataque apresentaram desempenho inferior em habilidades como sprint, salto e mudança de direção, no entanto, foram superiores em termos de força na parte superior do corpo. No entanto, as posições defensivas nas costas tiveram o pior resultado no teste de força da parte superior do corpo. Essas descobertas corroboram a relevância de levar em conta não somente as médias de altura e peso, mas também as capacidades físicas específicas de cada posição para avaliar o desempenho e selecionar os jogadores para a NFL realizadas neste presente estudo (ROBBINS, 2011).

Já para a Tabela 4.4, que apresenta os resultados Físicos dos Jogadores por Posição, observa-se que diferentes posições demonstram desempenhos variados em diferentes testes físicos realizados durante o período de 2000 a 2022. Por exemplo, Defensive Tackles (DT) apresentam a maior média no teste de Supino, com 27,6, enquanto Running Backs (RB) apresentam a menor média nesse teste, com 19,43. Além disso, Offensive Tackles (OT) mostram a maior média no teste de 40 jardas, com 5,24, enquanto Defensive Backs (DB) demonstram a menor média nesse mesmo teste, com 4,46. Essas variações destacam a diversidade de habilidades físicas exigidas em diferentes posições no futebol americano.

Tabela 4.4: Médias dos resultados físicos dos testes dos jogadores por Posição no *Combine* (2000-2022)

<b>Posição</b>	<b>40 jardas</b>	<b>Supino</b>	<b>Salto</b>	<b>3 cones</b>
Centers (C)	5,23	26,09	258,57	7,72
Cornerbacks (CB)	4,49	14,49	309,79	6,94
Defensive Backs (DB)	4,46	12,5	316,23	6,72
Defensive Ends (DE)	4,83	23,83	291,23	7,33
Defensive Linemen (DL)	4,98	25,42	285,21	7,58
Defensive Tackles (DT)	5,1	27,6	266,49	7,7
Edge Rushers (EDGE)	4,69	22,9	303,78	7,15
Inside Linebackers (ILB)	4,75	22,37	291,68	7,17
Linebackers (LB)	4,64	20,25	308,21	7,08
Offensive Guards (OG)	5,3	25,58	255,66	7,92
Offensive Linemen (OL)	5,18	25,76	270,29	7,85
Outside Linebackers (OLB)	4,68	22,55	298,69	7,13
Offensive Tackles (OT)	5,24	24,38	260,56	7,86
Running Backs (RB)	4,55	19,43	300,72	7,07
Safeties (S)	4,56	16,95	306,53	7
Tight Ends (TE)	4,77	20,18	292,18	7,19
Wide Receivers (WR)	4,51	14,57	306,75	6,98

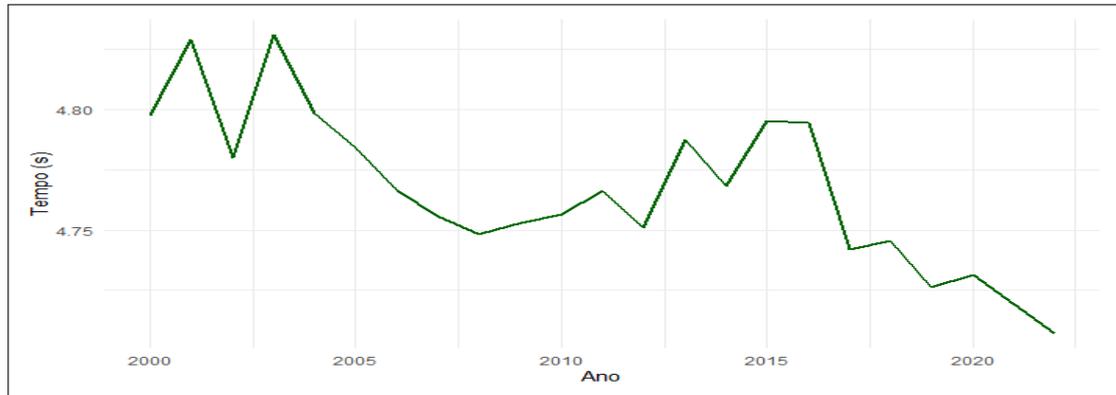
**Fonte:** *Elaborado pelo autor*

Com base nos dados da Tabela 4.4, que apresenta os desempenhos físicos dos jogadores em cada posição, fica claro que diferentes posições apresentam variações significativas em seus desempenhos em diversos testes físicos, o que reflete a variedade de habilidades requeridas em diferentes papéis no futebol americano. Essa conclusão é sustentada pelo estudo de Gabbett et al. (2016), que analisou os testes físicos que melhor predizem a capacidade de realizar esforços de alta intensidade repetitivos em jogadores profissionais de rugby league. Apesar do foco do estudo ser o rugby, suas descobertas sobre a relação entre testes físicos e desempenho em campo podem ser aplicadas ao contexto do futebol americano.

Ao analisar a Figura 4.1, percebe-se que, nos anos iniciais, foram registradas as médias mais altas nos testes de 40 jardas dos jogadores. Ao longo do tempo, houve uma queda na média, que voltou a crescer significativamente nos anos de 2011 e 2016. No entanto, ao longo dos anos seguintes, houve uma nova queda. Essas flutuações podem ser atribuídas a uma variedade de fatores, como mudanças na composição dos jogadores selecionados, evolução dos métodos de treinamento e condicionamento físico, investimentos em tecno-

logia e ciência do esporte, mudanças nas regras do jogo e fatores externos, como lesões ou condições meteorológicas durante os testes.

Figura 4.1: Análise média da evolução dos desempenhos no Teste de 40 Jardas dos Jogadores no *Combine* (2000-2022)



**Fonte:** *Elaborado pelo autor*

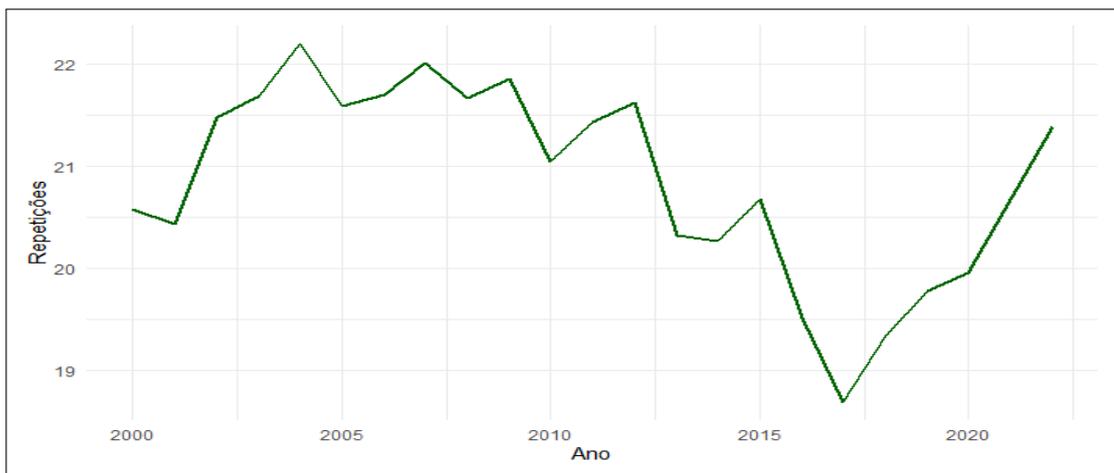
Essas variações podem ser atribuídas a uma série de fatores, como discutido por Dompier et al. (2015) em seu estudo sobre as tendências de lesões no futebol americano universitário. Os autores enfatizam que mudanças na formação dos jogadores, na forma como os métodos de treinamento e condicionamento físico são desenvolvidos, bem como avanços na tecnologia e ciência do esporte, podem ter um impacto direto no desempenho dos atletas ao longo do tempo.

Além disso, mudanças nas regras do jogo, como ajustes nos procedimentos de avaliação durante o Combine da NFL, também podem ter um impacto relevante nessas flutuações. Sendo assim, ao analisar as variações nos resultados dos testes ao longo dos anos, é crucial considerar a complexidade dos fatores envolvidos e sua influência cumulativa no desempenho dos jogadores ao longo do tempo (DOMPIER et al., 2015)

Ao analisar os dados apresentados na Figura 4.2 referentes à média da evolução dos desempenhos no Teste de Supino dos jogadores, é possível notar um aumento significativo nos valores logo depois dos anos iniciais. Essas primeiras melhorias podem ser atribuídas ao aperfeiçoamento técnico dos atletas, assim como aos investimentos em programas de condicionamento físico e treinamento específico para o teste de supino. Ao longo dos anos, houve variações, com períodos de baixas e altas, provavelmente influenciados por fatores como mudanças na equipe técnica, variações nos métodos de treinamento e até mesmo lesões individuais.

No entanto, a partir de 2012, percebe-se uma queda no desempenho, o que pode ser resultado de mudanças na forma de treinar, lesões mais frequentes ou até mesmo uma mudança na ênfase dada ao teste de supino dentro do contexto do jogo. Entretanto, é interessante notar que a partir de 2017, os valores voltam a crescer de forma exponencial, sugerindo possíveis ajustes nas estratégias de treinamento ou uma nova ênfase no aprimoramento dessa habilidade específica.

Figura 4.2: Análise média da evolução dos desempenhos no Teste de Supino dos Jogadores no *Combine* (2000-2022)

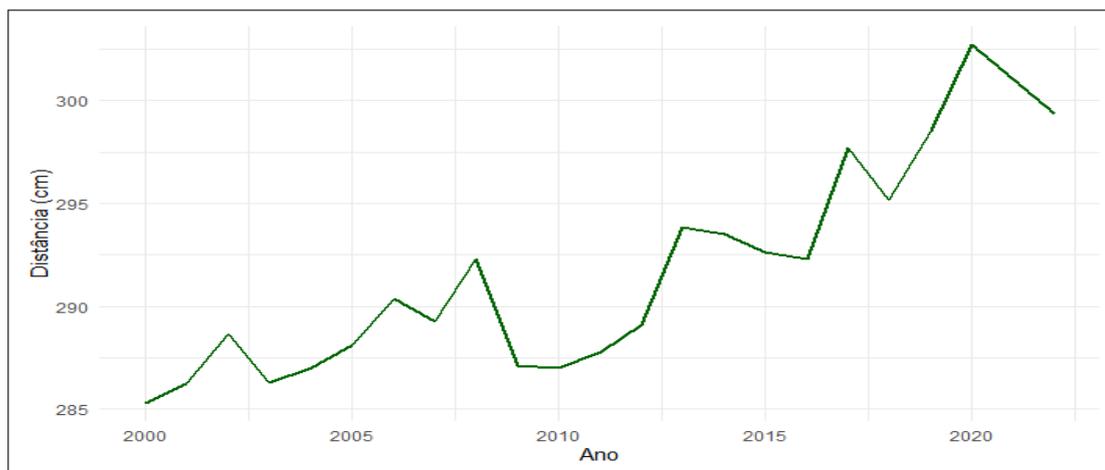


Fonte: Elaborado pelo autor

Essas observações estão em linha com os resultados de um estudo realizado por Suchomel et al. (2019), que investigou os efeitos de diferentes métodos de treinamento de força no desempenho do supino. Os autores destacaram que melhorias iniciais no desempenho podem ser atribuídas ao aperfeiçoamento técnico dos atletas, bem como a investimentos em programas de condicionamento físico e treinamento específico para o teste de supino.

Ao examinar a Figura 4.3, é possível notar uma tendência interessante nos resultados ao longo dos anos. Nos primeiros anos, houve avanços nos valores alcançados, sugerindo melhorias nos desempenhos dos jogadores. Vários fatores podem ser responsáveis por essas melhorias, tais como aprimorar as técnicas de treinamento, aprimorar as táticas e até mesmo o aprimoramento individual dos atletas. No entanto, a partir do ano de 2008 até 2012, os resultados apresentam um *gap* significativo. Essa queda pode ser causada por mudanças na equipe técnica, lesões importantes de jogadores-chave, ou até mesmo ajustes nas políticas de recrutamento da equipe. Após esse período de queda, há oscilações nos resultados, indicando possíveis obstáculos de consistência ou adaptação no ambiente esportivo. Entretanto, não há uma tendência clara de evolução nos anos finais, sugerindo que outros fatores podem ter influenciado o desempenho dos jogadores nesse período.

Figura 4.3: Análise média da evolução dos desempenhos no Teste de salto em distância dos Jogadores no *Combine* (2000-2022)



**Fonte:** Elaborado pelo autor

Sendo assim, ao examinar as variações no desempenho dos atletas ao longo do tempo, é crucial levar em conta não somente os aspectos relacionados ao treinamento e às habilidades individuais, mas também os fatores externos, como alterações nas regras do jogo, avanços na tecnologia de equipamentos esportivos e fatores socioeconômicos, que podem ter um impacto significativo na performance dos atletas (HOOTMAN et al., 2007)

Ao examinar o último teste relacionado ao teste dos 3 cones dos jogadores, apresentado na Figura 4.4, é possível notar variações nos resultados nos anos iniciais. No entanto, a partir de 2004, percebe-se uma queda significativa nos valores alcançados nos anos seguintes. Essa queda pode ser atribuída a uma série de fatores, como mudanças na forma de treinar, lesões recorrentes de jogadores importantes, ou até mesmo uma transição de gerações de atletas. A partir de 2013, nota-se uma melhoria nos resultados, sugerindo a necessidade de ajustes nas táticas de treinamento ou recrutamento. No entanto, logo após 2016, há uma nova queda nos valores, que voltam a aumentar a partir de 2018. Essas variações podem indicar a complexidade do cenário esportivo, que envolve aspectos como mudanças na liderança técnica, adaptação às novas regras ou até mesmo influências externas, como o contexto econômico ou político.

Ao analisar o estudo realizado por Machado et al. (2013), que discute a relação entre eficácia ofensiva e variabilidade de padrões de jogo em futebol, destaca-se a importância da capacidade de adaptação às diferentes situações do jogo para o sucesso das equipes. Este estudo, conduzido em 2013, sugere que as variações nos resultados do teste dos 3 cones

Figura 4.4: Análise média da evolução dos Desempenhos no Teste dos 3 cones dos Jogadores no *Combine* (2000-2022)



**Fonte:** *Elaborado pelo autor*

ao longo dos anos podem ser atribuídas a uma série de fatores dinâmicos e interconectados, como mudanças nas abordagens de treinamento, lesões de jogadores importantes e influências externas, o que ressalta a complexidade do cenário esportivo.

Em síntese, a análise dos quatro testes mostrados nas Figuras 4.1 a 4.4 revela uma variedade de padrões e tendências ao longo dos anos. Nos testes de 40 jardas (Figura 4.1), percebeu-se uma variação nas médias ao longo do tempo, com picos e quedas que podem indicar mudanças na composição dos jogadores, métodos de treinamento e outros fatores. Da mesma forma, nos testes de supino (Figura 4.2), notou-se um aumento inicial seguido de oscilações e uma queda acentuada em um determinado período, provavelmente causada por mudanças nas abordagens de treinamento. O teste do salto em distância apresentado na Figura 4.3 mostrou uma tendência de melhoria inicial seguida por uma queda significativa em um determinado período de tempo, sugerindo possíveis desafios enfrentados pela equipe durante esse período. Finalmente, o teste dos 3 cones (Figura 4.4) apresentou variações nos valores ao longo dos anos, com momentos de melhora e queda que podem indicar diversos fatores complexos no ambiente esportivo. Esta análise enfatiza a relevância de levar em conta diversas variáveis, como estratégias de treinamento, composição da equipe e influências externas, ao analisar as tendências dos resultados dos testes ao longo do tempo.

## 4.3 Pré-processamento e outros aspectos

### 4.3.1 Primeira etapa

Para facilitar a compreensão dos resultados registrados no banco de dados, foram realizadas adaptações por meio das seguintes conversões e essas adaptações foram aplicadas para garantir uma melhor interpretação dos dados coletados. Logo abaixo, segue as conversões realizadas na Tabela 4.5.

Tabela 4.5: Conversões de Unidades

Conversão	Valor Convertido
* 1 Polegada	2,54 cm
** 5' 10"	$(12 \times 5 + 10) \times 2,54$ cm
*** 1 Libra	0,453592 g

**Fonte:** *Elaborado pelo autor*

Inicialmente, os dados foram submetidos a uma análise usando a técnica de Análise de Componentes Principais (PCA) sem padronização, com foco nas variáveis peso e altura. Após a utilização do PCA, verificou-se que a primeira componente principal é responsável por 95,96% da variabilidade total dos dados. Esta componente principal representa uma combinação linear das variáveis originais e aponta em uma direção específica no espaço de características. É evidente que ambas as variáveis, peso e altura, apresentam coeficientes negativos nesta parte, indicando que elas tendem em direção oposta à medida que a principal aumenta. Ademais, o coeficiente associado ao peso é menor do que o coeficiente da altura, sugerindo que o peso exerce uma influência mais significativa na direção da componente principal, conforme é observado na tabela 4.6 logo abaixo.

Tabela 4.6: Componentes da Altura e Peso (Sem padronização)

Variável	Coefficiente
Altura	-0,2411
Peso	-0,9705

**Fonte:** *Elaborado pelo autor*

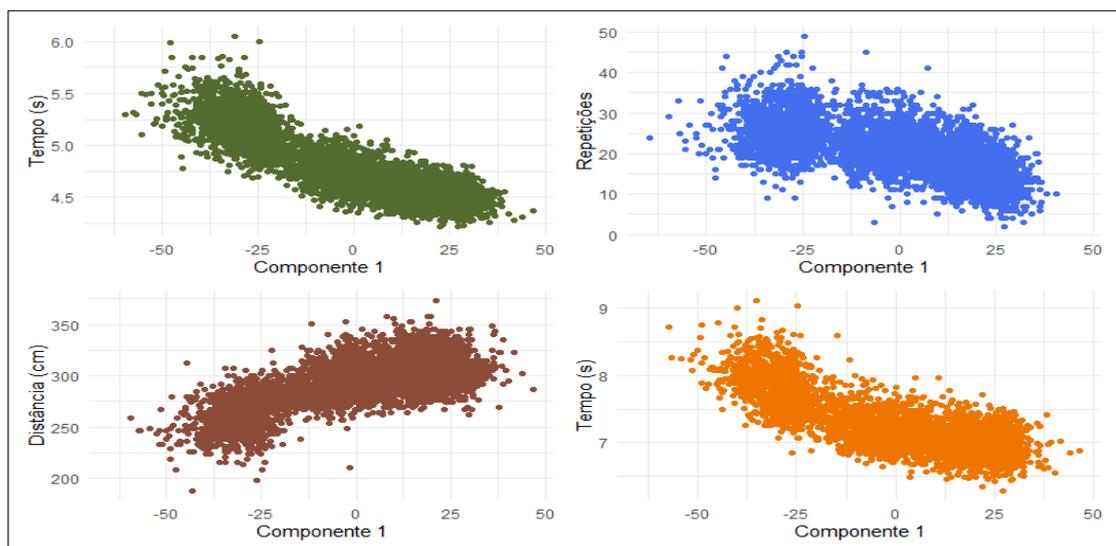
Com base nesses resultados, optou-se por utilizar somente a primeira componente principal para representar os dados, uma vez que ela é capaz de explicar a maioria da variabilidade observada. Posteriormente, foram elaborados gráficos para cada grupo de testes físicos, tais como o teste de 40 jardas, o teste de supino, o salto a distância e o teste dos 3 cones, utilizando a variável principal como referência. Esse procedimento é a primeira etapa do processamento dos dados, fornecendo uma base sólida para análises futuras.

A análise da relação entre as componentes principais e o desempenho nos testes físicos de jogadores no período de 2020 a 2022 oferece *insights* valiosos sobre a interação entre variáveis antropométricas e habilidades físicas.

Os quatro gráficos apresentados na Figura 4.5 destacam essa relação ao exibir a distribuição dos resultados dos testes físicos em relação à primeira componente principal. No eixo x, temos a representação da componente principal, enquanto no eixo y são apresentados os resultados de cada teste físico específico, incluindo o teste de 40 jardas, teste de supino, teste de salto a distância e teste dos 3 cones.

A disposição dos gráficos permite uma análise comparativa do impacto da primeira componente principal em diferentes aspectos do desempenho físico dos jogadores. Esse contexto é fundamental para compreender como variáveis antropométricas podem influenciar diretamente as capacidades físicas dos atletas, fornecendo uma base sólida para futuras análises e estratégias de treinamento personalizadas.

Figura 4.5: Relação entre Componentes Principais e Desempenho nos Testes Físicos de Jogadores no *Combine* (2020-2022)



Fonte: Elaborado pelo autor

### 4.3.2 Segunda etapa

Nesta etapa, foi selecionada uma amostra aleatória para validação de 20% para cada teste físico, enquanto os participantes não testados no teste físico em questão foram excluídos da análise. Em seguida, foram gerados histogramas para os testes de 40 jardas, Salto em distância e 3 cones, visando analisar as distribuições das variáveis principais e facilitar a compreensão dos resultados obtidos no banco de dados. Como os dados referentes ao modelo do Supino são discretos, não foi necessário construir um histograma para essa variável.

Deste modo, após a escolha dos melhores modelos para cada teste físico a partir da função desvio, temos os quatro modelos detalhados na Tabela 4.7 abaixo, foi observado que cada modelo foi adaptado às características específicas das variáveis analisadas.

Vale ressaltar que os modelos foram desenvolvidos e validados com conjuntos de dados de tamanhos diferentes: 40 Jardas (n= 4.898 para desenvolvimento e n= 1.225 para validação), Salto em distância (n= 4.020 para desenvolvimento e n= 1.005 para validação), Supino (n= 3.690 para desenvolvimento e n= 922 para validação), e 3 cones (n= 3.231 para desenvolvimento e n= 808 para validação)

Tabela 4.7: Modelos Lineares Generalizados (MLG)

<b>Modelo</b>	<b>Família</b>	<b>Função de Ligação</b>
40 jardas	Gama	Identidade
Supino	Quasipoisson	Identidade
Salto em distância	Gaussiana	Inversa
3 cones	Gaussiana Inversa	Identidade

**Fonte:** *Elaborado pelo autor*

Logo após a construção dos quatro modelos MLG, foram obtidas métricas para a avaliação do desempenho de cada modelo. Em seguida, os erros foram calculados e analisados ao longo do período de 2000 a 2022. Sendo assim, a partir dos erros calculados, foram criados os índices. Em seguida, os dados foram avaliados ano a ano e, a partir dos índices dos quatro modelos desenvolvidos, foi construído o índice global, representando a Análise de Componentes Principais (PCA) dos quatro índices individuais obtidos.

O primeiro componente principal dos índices representa uma variabilidade de 97,33%. Além disso, os coeficientes de cada índice individual no primeiro componente são apresentados na Tabela 4.8, logo abaixo.

Tabela 4.8: Coeficiente dos índices na construção do índice global

<b>Modelo</b>	<b>Coeficiente</b>
40 Jardas	-0.5859289
Supino	-0.3916531
Salto em distância	-0.4508889
3 cones	-0.5477175

**Fonte:** *Elaborado pelo autor*

Assim, após a obtenção do índice global, ele é padronizado e multiplicado por (-1) devido ao seu valor negativo, visando facilitar a interpretação. Em seguida, é calculada a média ao longo dos anos para utilizar a medida resumo Anual Percentage Change (APC). Essa etapa é de extrema importância para avaliar o objetivo principal do presente estudo, afim proporcionar *insights* significativos sobre os dados analisados.

## 5.1 Desenho experimental

Para aprimorar a compreensão e avaliação dos resultados dos testes físicos realizados durante o *Combine* da NFL, os atletas foram agrupados de acordo com suas posições em grupos que apresentaram resultados semelhantes em cada teste específico. Esses agrupamentos facilitam a identificação de padrões e tendências nas diferentes posições e ajudam a contextualizar os dados de maneira mais precisa.

Os atletas foram agrupados conforme os resultados dos testes físicos: 40 jardas, salto em distância, supino e 3 cones. Assim, as posições foram divididas em três grupos distintos dentro de cada teste físico, de acordo com os padrões de desempenho similares observados anteriormente na tabela 4.4.

Quadro 3 - Grupos e subgrupos de posições dos testes no *Combine* da NFL

<b>Teste de 40 jardas</b>	Subgrupo 1	OL, C, OG, OT, DT
	Subgrupo 2	ILB, LB, OLB, DL, DE, EDGE, TE
	Subgrupo 3	S, RB, WR, CB, DB
<b>Teste do supino</b>	Subgrupo 4	DT, C, OL, OG, DL, OT
	Subgrupo 5	DE, EDGE, OLB, ILB, LB, TE
	Subgrupo 6	RB, S, WR, CB, DB
<b>Teste de salto em distância</b>	Subgrupo 7	S, LB, WR, CB, DB
	Subgrupo 8	EDGE, RB, OLB, TE, ILB, DE, DL
	Subgrupo 9	OL, C, OG, OT, DT
<b>Teste do tempo dos 3 cones</b>	Subgrupo 10	OG, OT, OL, C, DT, DL
	Subgrupo 11	DE, TE, ILB, EDGE, OLB, LB, RB
	Subgrupo 12	S, WR, CB, DB

**Fonte:** *Elaborado pelo autor.*

O Grupo do teste de 40 jardas englobou posições como *Offensive Linemen* (OL), *Centers* (C), *Offensive Guards* (OG), *Offensive Tackles* (OT), e *Defensive Tackles* (DT) no Subgrupo 1; *Inside Linebackers* (ILB), *Linebackers* (LB), *Outside Linebackers* (OLB), *Defensive Linemen* (DL), *Defensive Ends* (DE), *Edge Rushers* (EDGE), e *Tight Ends* (TE) no Subgrupo 2; e *Safeties* (S), *Running Backs* (RB), *Wide Receivers* (WR), *Cornerbacks* (CB), e *Defensive Backs* (DB) no Subgrupo 3.

O Grupo do teste de supino englobou posições como *Defensive Tackles* (DT), *Centers* (C), *Offensive Linemen* (OL), *Offensive Guards* (OG), *Defensive Linemen* (DL), e *Offensive Tackles* (OT) no Subgrupo 4; *Defensive Ends* (DE), *Edge Rushers* (EDGE), *Outside Linebackers* (OLB), *Inside Linebackers* (ILB), *Linebackers* (LB), e *Tight Ends* (TE) no Subgrupo 5; e *Running Backs* (RB), *Safeties* (S), *Wide Receivers* (WR), *Cornerbacks* (CB), e *Defensive Backs* (DB) no Subgrupo 6.

O Grupo do teste de salto em distância englobou posições como *Safeties* (S), *Linebackers* (LB), *Wide Receivers* (WR), *Cornerbacks* (CB), e *Defensive Backs* (DB) no Subgrupo 7; *Edge Rushers* (EDGE), *Running Backs* (RB), *Outside Linebackers* (OLB), *Tight Ends* (TE), *Inside Linebackers* (ILB), *Defensive Ends* (DE), e *Defensive Linemen* (DL) no Subgrupo 8; e *Offensive Linemen* (OL), *Centers* (C), *Offensive Guards* (OG), *Offensive Tackles* (OT), e *Defensive Tackles* (DT) no Subgrupo 9.

O Grupo do teste de tempo do 3 cone englobou posições como *Offensive Guards* (OG), *Offensive Tackles* (OT), *Offensive Linemen* (OL), *Centers* (C), *Defensive Tackles* (DT), e *Defensive Linemen* (DL) no Subgrupo 10; *Defensive Ends* (DE), *Tight Ends* (TE), *Inside Linebackers* (ILB), *Edge Rushers* (EDGE), *Outside Linebackers* (OLB), *Linebackers* (LB), e *Running Backs* (RB) no Subgrupo 11; e *Safeties* (S), *Wide Receivers* (WR), *Cornerbacks* (CB), e *Defensive Backs* (DB) no Subgrupo 12.

Essa estrutura de agrupamento permitiu uma análise mais aprofundada e específica do desempenho físico dos jogadores em diferentes testes, permitindo a identificação de padrões e tendências em cada categoria de posição. A análise do potencial e da adequação dos prospectos para o jogo profissional é crucial para equipes e analistas da NFL.

## 5.2 Resultados

Ao analisar os resultados, pode-se afirmar que os modelos lineares generalizados permitiram estimar os resultados médios dos testes físicos realizados com precisão, levando em consideração as variáveis do estudo associadas ao desfecho, tais como peso, altura e

subgrupos.

Dessa forma, em relação às variáveis associadas ao desfecho, foi calculado os índices para cada teste físico. A partir dessas descobertas, buscou-se avaliar a evolução ao longo dos anos dos atletas universitários e identificar se houve uma melhora significativa nos resultados dos testes realizados.

Para o teste do tiro de 40 jardas, foram ajustados 3 modelos na família gama com 3 funções de ligação distintas (identidade, inversa e logarítmica). Após avaliar os três modelos através da função desvio e realizar o teste de bondade de ajuste, foi determinado que o modelo adequado, que possuiu o menor valor de função desvio (4877,53), foi o modelo com a ligação identidade (Qui-quadrado de referência: 5057,86).

Com relação ao teste do Salto em distância, foram ajustados 7 modelos. Destes, 4 pertenciam à família da gaussiana inversa, com funções de ligação identidade, inversa, logarítmica e  $\frac{1}{\mu^2}$ , e 3 modelos pertenciam à família gaussiana, com funções de ligação identidade, inversa e logarítmica. Após avaliar os sete modelos através da função desvio e realizar o teste de bondade de ajuste, determinou-se que o modelo adequado, com a menor função desvio (4016,00), foi o da família gaussiana com ligação inversa (Qui-quadrado de referência: 4164,54).

Conforme o teste do Supino, foram ajustados 4 modelos na família da Quasipoisson, com funções de ligação identidade, inversa, raiz quadrada e logarítmica. Após avaliar os quatro modelos através da função desvio e realizar o teste de bondade de ajuste, determinou-se que o modelo adequado, com a menor função desvio (3720,60), foi o da família Quasipoisson com ligação identidade (Qui-quadrado de referência: 3828,36).

De acordo o teste dos 3 cones, foram ajustados 5 modelos. Destes, 3 pertenciam à família da gama, com funções de ligação identidade, inversa, log e inversa, e 2 modelos pertenciam à família gaussiana inversa, com funções de ligação identidade e inversa. Após avaliar os cinco modelos através da função desvio e realizar o teste de bondade de ajuste, determinou-se que o modelo adequado, com a menor função desvio (3198,21), foi o da família gaussiana inversa com ligação identidade (Qui-quadrado de referência: 3360,27).

Ao analisar o modelo de 40 jardas com a família gama e função de ligação identidade, apresentado na Tabela 5.1, podemos constatar que o intercepto do modelo é de aproximadamente 5,00, indicando o valor esperado do desempenho no teste quando todas as outras variáveis são zero. Além disso, o modelo considera uma componente principal e a inclusão de subgrupos específicos (Subgrupo 2 e Subgrupo 3) como variáveis explicativas. As estimativas dos parâmetros para essas variáveis indicam suas respectivas contribuições

para o desempenho no teste, enquanto os valores de S.E (Erro Padrão) e P-Valor fornecem informações sobre a precisão das estimativas e a significância estatística das variáveis no modelo, respectivamente. Os valores extremamente baixos de P-Valor ( $<2e-16$ ) sugerem uma forte associação entre as variáveis explicativas e o desempenho no Teste de 40 Jardas dos jogadores.

Tabela 5.1: Variáveis associadas ao desempenho no Teste de 40 Jardas dos Jogadores no *Combine* (2000-2022)

Variável	Estimativa do parâmetro	S.E	P-Valor
Intercepto	4,9972565	0,0084145	$<2e-16$
Componente 1	-0,0070218	0,0002449	$<2e-16$
Subgrupo 2	-0,2637795	0,0082599	$<2e-16$
Subgrupo 3	-0,3312353	0,0132197	$<2e-16$

**Fonte:** *Elaborado pelo autor*

Conforme o modelo do teste do supino com a família Quasipoisson e função de ligação identidade, apresentado na Tabela 5.2, podemos constatar que o intercepto do modelo é de aproximadamente 21,69, sugerindo uma relação positiva com o desempenho no teste. As demais estimativas de -0,14 e -2,68 indicam contribuições negativas para o desempenho. Além disso, o modelo considera um componente principal e a inclusão de um subgrupo específico (Subgrupo 6) como variáveis explicativas, uma vez que o (Subgrupo 7) não foi estatisticamente significativo. As estimativas dos parâmetros para essas variáveis indicam suas respectivas contribuições para o desempenho no teste, enquanto os valores de S.E (Erro Padrão) e P-Valor fornecem informações sobre a precisão das estimativas e a significância estatística das variáveis no modelo, respectivamente. Os valores extremamente baixos de P-Valor ( $<2e-16$ ) sugerem uma forte associação entre as variáveis explicativas e o desempenho no teste de salto em distância dos jogadores.

Tabela 5.2: Variáveis associadas ao desempenho no Teste do supino dos Jogadores no *Combine* (2000-2022)

Variável	Estimativa do parâmetro	S.E	P-Valor
Intercepto	21,68796	0,14662	$<2e-16$
Componente 1	-0,13994	0,00646	$<2e-16$
Subgrupo 6	-2,68011	0,27317	$<2e-16$

**Fonte:** *Elaborado pelo autor*

Com relação ao modelo do teste de salto em distância com a família gaussiana e função de ligação inversa, apresentado na Tabela 5.3, podemos observar que o intercepto do modelo é aproximadamente  $3,32e-03$ , enquanto o Componente 1 é  $-3,29e-06$ , sugerindo uma relação negativa com o desempenho no teste. Por outro lado, as demais estimativas de  $6,95e-05$  e  $4,12e-04$  indicam contribuições positivas para o desempenho. Além disso, o modelo considera um componente principal e a inclusão de subgrupos específicos (Subgrupo 8 e Subgrupo 9) como variáveis explicativas. As estimativas dos parâmetros para essas variáveis indicam suas respectivas contribuições para o desempenho no teste. Enquanto isso, os valores de S.E (Erro Padrão) e P-Valor fornecem informações sobre a precisão das estimativas e a significância estatística das variáveis no modelo, respectivamente. Os valores extremamente baixos de P-Valor ( $<2e-16$ ) sugerem uma forte associação entre as variáveis explicativas e o desempenho no teste de salto em distância dos jogadores.

Tabela 5.3: Variáveis associadas ao desempenho no Teste do salto em distância dos Jogadores no *Combine* (2000-2022)

Variável	Estimativa do parâmetro	S.E	P-Valor
Intercepto	3,319e-03	7,181e-06	$<2e-16$
Componente 1	-3,293e-06	2,897e-07	$<2e-16$
Subgrupo 8	6,955e-05	7,991e-06	$<2e-16$
Subgrupo 9	4,116e-04	1,670e-05	$<2e-16$

**Fonte:** *Elaborado pelo autor*

Dado o modelo do teste dos 3 cones com a família gaussiana inversa e função de ligação identidade, apresentado na Tabela 5.4, observamos que o intercepto do modelo é aproximadamente 7,51, sugerindo uma relação positiva com o desempenho no teste. As demais estimativas indicam o valor esperado do desempenho no teste quando todas as outras variáveis são zero. Além disso, o modelo considera um componente principal e a inclusão de subgrupos específicos (Subgrupo 11 e Subgrupo 12) como variáveis explicativas. As estimativas dos parâmetros para essas variáveis indicam suas respectivas contribuições para o desempenho no teste. Os valores de S.E (Erro Padrão) e P-Valor fornecem informações sobre a precisão das estimativas e a significância estatística das variáveis no modelo, respectivamente, sendo todos os p-valores extremamente baixos ( $<2e-16$ ), o que sugere uma forte associação entre as variáveis explicativas e o desempenho no teste de tempo dos 3 cones dos jogadores.

Tabela 5.4: Variáveis associadas ao desempenho no Teste do tempo dos 3 cones dos Jogadores no *Combine* (2000-2022)

Variável	Estimativa do parâmetro	S.E	P-Valor
Intercepto	7,5104217	0,0159655	<2e-16
Componente 1	-0,0096419	0,0004645	<2e-16
Subgrupo 11	-0,3317645	0,0176464	<2e-16
Subgrupo 12	-0,3375333	0,0256878	<2e-16

**Fonte:** *Elaborado pelo autor*

Em síntese, os modelos fornecem uma compreensão mais profunda de como diversos fatores impactam o desempenho dos jogadores nesses testes específicos ao longo do período analisado. No teste de 40 jardas e no teste de 3 cones, a estimativa negativa dos parâmetros indica um favorecimento do resultado do jogador no teste realizado. Já no teste de supino e no salto em distância, a estimativa negativa dos parâmetros indica um desfavorecimento do resultado do jogador no teste realizado.

Com base nos resultados dos modelos gerados, foi conduzida uma previsão utilizando os dados de teste, com o objetivo de avaliar o desempenho do modelo em relação a novos dados. Essa análise é refletida na Tabela 5.5.

Tabela 5.5: Métricas associadas ao desempenho dos modelos no conjunto de validação

Modelo	Raiz do erro quadrático médio	Pseudo-R <sup>2</sup>
40 jardas	0,13	82,10
Supino	4,83	42,27
Salto em distância	16,22	53,82
3 cones	0,24	67,46

**Fonte:** *Elaborado pelo autor*

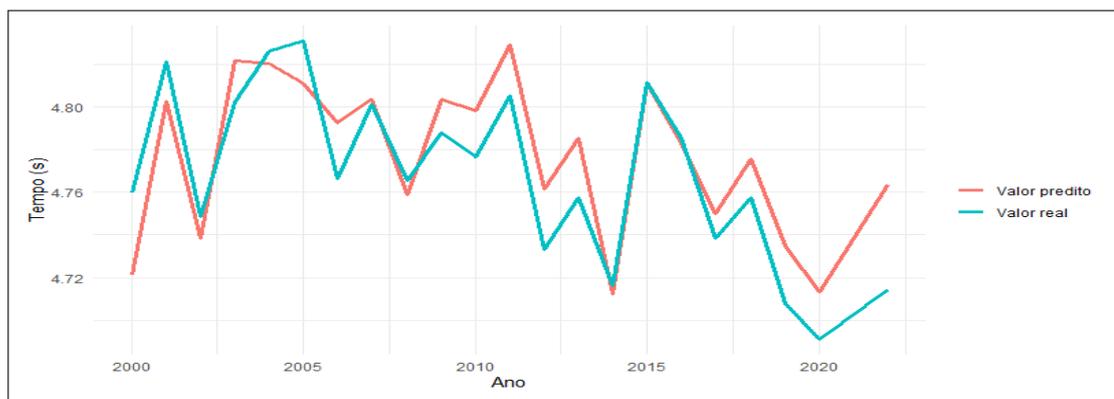
Desta forma, com relação aos resultados dos modelos analisados é possível observar que, as duas métricas principais utilizadas foram: raiz do erro quadrático médio e pseudo-R<sup>2</sup>. A raiz do erro quadrático médio (RMSE), fornece uma medida da dispersão dos valores previstos em relação aos valores reais, sendo assim, ao obter valores menores, indica um melhor ajuste do modelo aos dados. Nesse contexto, o modelo de "40 jardas" apresenta a menor raiz do erro quadrático médio (0,13 segundos), sugerindo uma boa precisão em suas previsões. Em contrapartida, o modelo de "Salto em distância" exibe a maior raiz do erro quadrático médio (16,22 centímetros), indicando uma maior variabilidade nas previsões em relação aos dados reais.

Com relação ao pseudo- $R^2$ , uma medida de quão bem o modelo se ajusta aos dados em comparação com um modelo de referência simples, revela a capacidade de explicação do modelo em termos de variabilidade total dos dados. Aqui, observamos que o modelo "40 jardas" também apresenta o maior pseudo- $R^2$  (82,10%), sugerindo que uma grande parte da variabilidade nos dados é explicada pelo modelo. Por outro lado, o modelo "Supino" possui um menor pseudo- $R^2$  (42,27%), indicando uma explicação relativamente inferior da variabilidade dos dados em comparação com os outros modelos. Essas métricas combinadas fornecem uma visão abrangente do desempenho e da capacidade explicativa de cada modelo em relação aos dados observados.

Sendo assim, com base nos dados observados e nas previsões dos modelos, logo abaixo é apresentado uma comparação entre os valores médios dos testes realizados pelos jogadores e os valores médios previstos por cada modelo construído.

Ao comparar os valores médios no teste de 40 jardas e os valores médios previstos na figura 5.1, é possível observar que o modelo demonstra um bom desempenho, estando melhor ajustado aos resultados ao longo dos anos. Isso evidencia os achados mencionados anteriormente, corroborados pelas métricas obtidas do modelo.

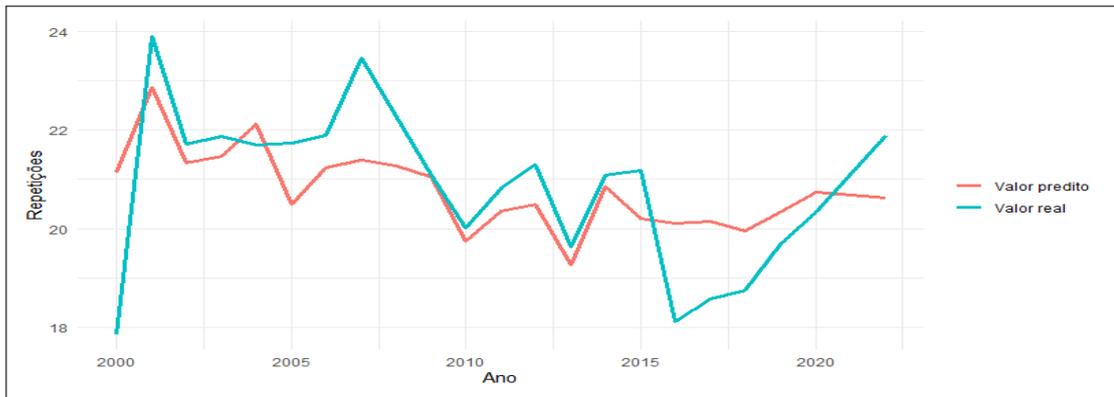
Figura 5.1: Comparação entre a Média Anual e o Média do Valor Predito no Teste de 40 Jardas dos Jogadores no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

Com relação aos valores médios no teste do supino e os valores médios previstos na figura 5.2, percebe-se que o modelo não consegue capturar as variações que ocorrem ao longo dos anos. Em alguns anos, é possível obter uma boa previsão, enquanto em outros anos, nota-se um desempenho inferior ao comparar os valores. Isso reforça os achados mencionados anteriormente pelas métricas avaliadas.

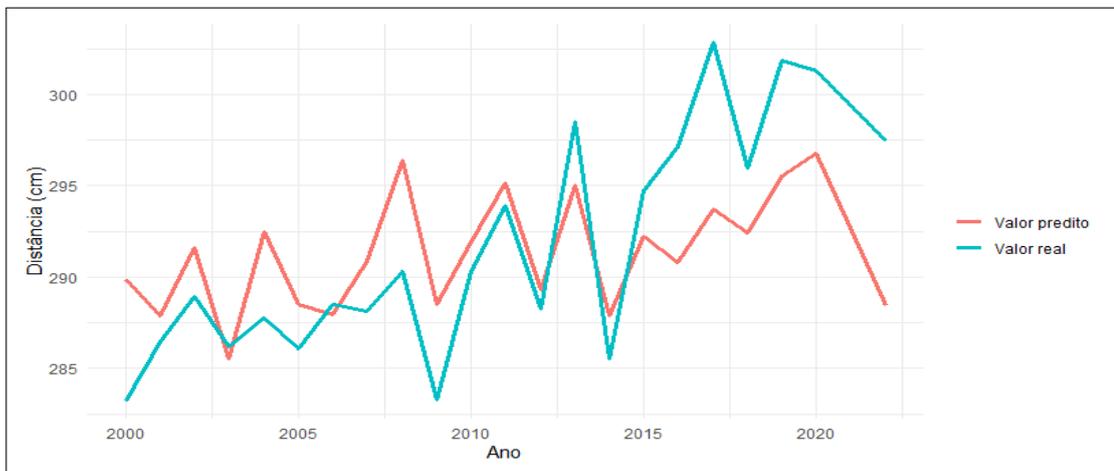
Figura 5.2: Comparação entre a Média Anual e o Média do Valor Predito no Teste do supino dos Jogadores no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

De acordo com aos valores médios no teste de distância e os valores médios previstos na figura 5.3, é perceptível que o modelo consegue capturar as variações que ocorrem ao longo dos anos entre 2000 e 2015. No entanto, após esse período, o modelo não conseguiu acompanhar a evolução dos resultados no teste, resultando em um desempenho inferior ao comparar os valores reais com os previstos.

Figura 5.3: Comparação entre a Média Anual e o Média do Valor Predito no Teste de distância dos Jogadores no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

E por fim, ao comparar aos valores médios no teste dos 3 cones e os valores médios previstos na figura 5.4, é perceptível que o modelo consegue capturar as variações que ocorrem ao longo dos anos entre 2015 a 2022. No entanto, antes desse período, o modelo não conseguiu acompanhar a evolução dos resultados no teste, resultando em um desempenho inferior ao comparar os valores reais com os previstos.

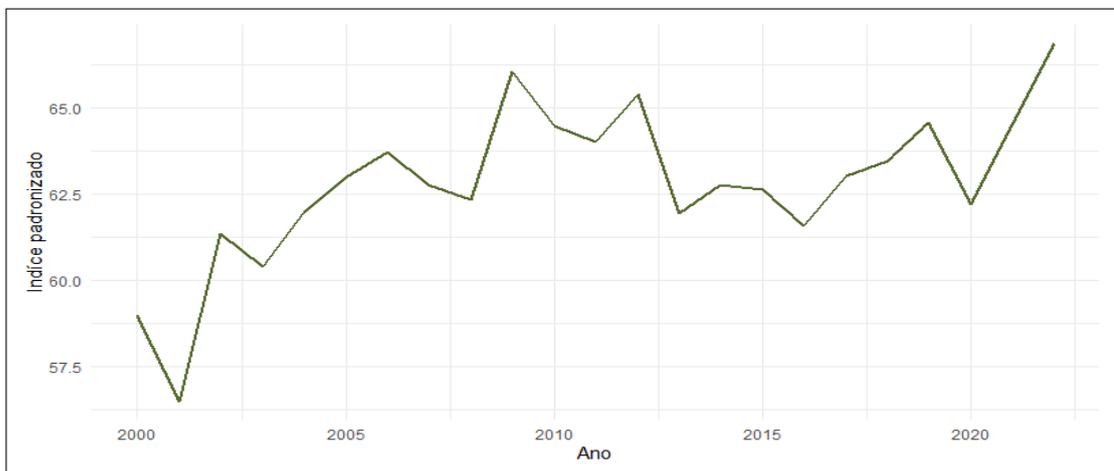
Figura 5.4: Comparação entre a Média Anual e o Média do Valor Predito no Teste dos 3 cones dos Jogadores no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

Ao analisar o índice padrão médio na Figura 5.5, é evidente que ao longo dos anos os valores do índice têm melhorado consistentemente. Isso indica que os jogadores têm superado as expectativas nos últimos anos, o que evidencia um aumento no desempenho e nos resultados do teste físico de 40 jardas.

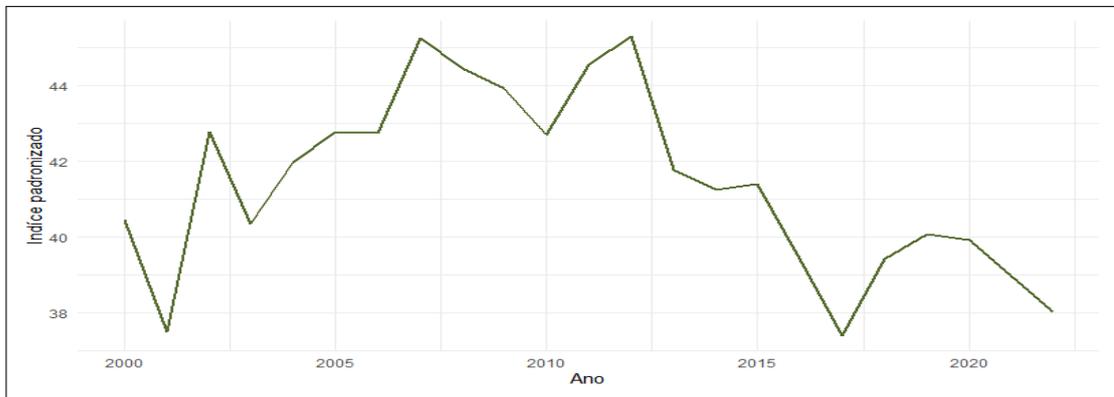
Figura 5.5: Desenvolvimento do Índice Padrão Médio do Modelo de 40 Jardas no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

Analisando os resultados do índice padrão médio na Figura 5.6, é perceptível uma tendência inicial de aumento nos anos iniciais, seguida por uma queda nos valores médios ao longo do tempo. Essa tendência sugere que, após um período inicial de melhoria, o desempenho dos jogadores no teste diminuiu progressivamente ao longo do período analisado, ficando abaixo das expectativas.

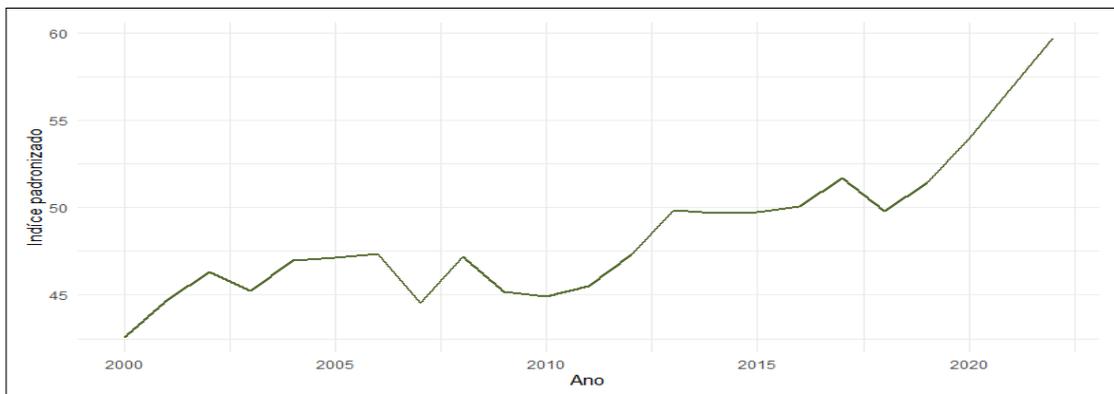
Figura 5.6: Desenvolvimento do Índice Padrão Médio do Modelo de supino no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

De acordo com os resultados obtidos na Figura 5.7, é possível afirmar que, a partir dos índices padrão médio analisados, houve um avanço significativo ao longo dos anos, evidenciando, assim, que o desempenho dos jogadores no período analisado tem melhorado de forma significativa.

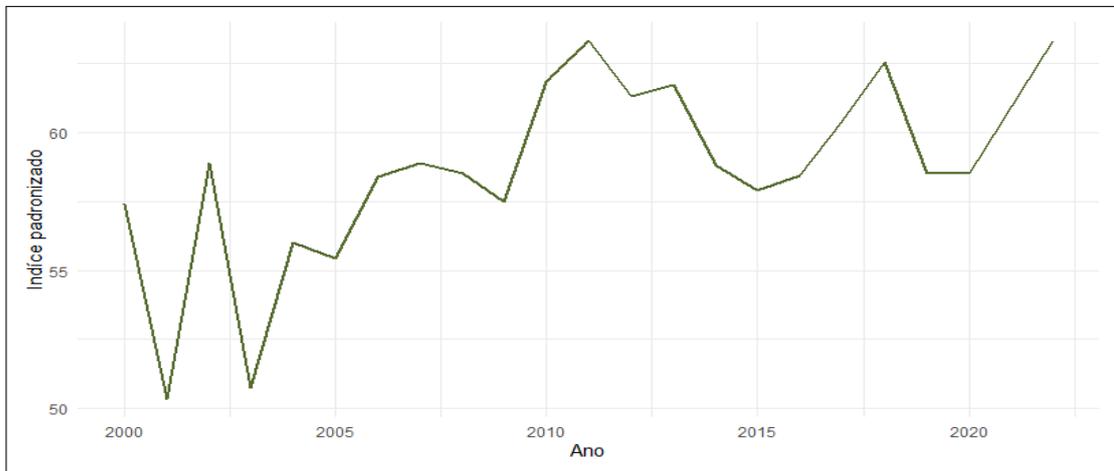
Figura 5.7: Desenvolvimento do Índice Padrão Médio do Modelo de salto em distância no *Combine* (2000-2022)



Fonte: Elaborado pelo autor

Com base nos resultados observados na Figura 5.8, pode-se afirmar que houve uma oscilação no desempenho dos jogadores nos anos iniciais. No entanto, ao longo do período analisado, após essa fase de oscilação, nota-se uma melhoria significativa nos resultados obtidos, refletindo um avanço no rendimento médio dos jogadores.

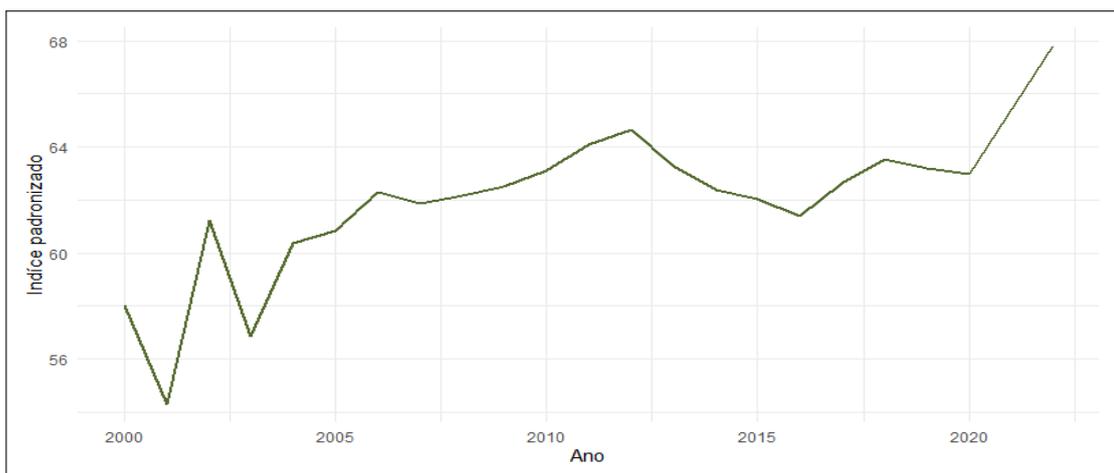
Figura 5.8: Desenvolvimento do Índice Padrão Médio do Modelo de 3 cones no *Combine* (2000-2022)



**Fonte:** Elaborado pelo autor

Ao analisar os resultados, é possível afirmar que o índice global de desempenho desenvolvido aumenta em média ao longo do tempo, indicando uma evolução dos jogadores. O índice global é a componente 1 do PCA (Análise de Componentes Principais) dos 4 índices listados anteriormente, combinando assim os resultados dos 4 testes analisados. Isso favorece a visualização do avanço dos resultados médios dos jogadores, pois considera todos os testes realizados. A partir deste índice, é possível analisar em conjunto e ter um contexto geral dos testes físicos realizados.

Figura 5.9: Desenvolvimento do Índice Global Padrão Médio dos Modelos MLG no *Combine* (2000-2022)



**Fonte:** Elaborado pelo autor

Diante dos resultados obtidos, foi crucial validar as informações previamente adquiridas. Nesse sentido, o cálculo do *Annual Percentage Change* (APC) foi realizado com o

intuito de verificar não apenas a solidez das etapas de avaliação, mas também a robustez e abrangência da estrutura fornecida para a seleção dos métodos de MLG mais apropriados. Isso assegura a obtenção de resultados confiáveis e de alta qualidade, essenciais para impulsionar significativamente o avanço da pesquisa nessa área específica.

Sendo assim, ao implementar o cálculo a partir dos dados obtidos, a porcentagem de crescimento anual do rendimento dos jogadores nos testes físicos realizados foi de 85,39% (IC95%: 58,82% a 116,39%). Esses resultados indicam uma tendência de crescimento significativo ao longo do período analisado, reforçando a eficácia das estratégias adotadas e a consistência dos avanços alcançados no desempenho dos jogadores.

Além disso, a análise minuciosa dos resultados revelou informações valiosas acerca dos padrões de desempenho dos jogadores ao longo do tempo. Ficou claro que alguns períodos tiveram melhorias consistentes, enquanto outros tiveram variações mais significativas. A compreensão mais aprofundada dos dados não apenas confirma a eficiência das medidas utilizadas, mas também fornece uma base sólida para futuras pesquisas e intervenções visando aprimorar ainda mais o desempenho desportivo. Com essas informações em mente, é possível direcionar os esforços e recursos para áreas específicas que possam se beneficiar mais com intervenções direcionadas, contribuindo para o crescimento constante e o aperfeiçoamento do desempenho esportivo.

A supervisão longitudinal de atletas é essencial para compreender os padrões de desempenho ao longo do tempo. Como apontado por Bailey (2019) no seu artigo 'Longitudinal Monitoring of Athletes: Statistical Issues and Best Practices', existem diversas considerações estatísticas que devem ser consideradas ao analisar dados de desempenho esportivo. Bailey ressalta a importância de uma rápida incorporação dos dados no processo de tomada de decisão, juntamente com a utilização de visualizações de dados claras e simples. Ao levar em conta as melhores práticas recomendadas por Bailey, é possível aperfeiçoar a eficiência das intervenções e programas de desenvolvimento de atletas, proporcionando um crescimento constante no desempenho esportivo.

E por fim, ao analisar a revisão sistemática realizada por José Pino-Ortega et al. (2021) sobre 'Desenho de Treinamento, Análise de Desempenho e Identificação de Talentos – Uma Revisão Sistemática sobre as Variáveis Mais Relevantes por meio da Análise de Componentes Principais em Futebol, Basquete e Rugby', destaca-se a importância da Análise de Componentes Principais (PCA) na identificação de variáveis determinantes no desempenho esportivo. Os resultados mostram que diferentes esportes coletivos exigem conjuntos específicos de habilidades físicas e técnicas para o sucesso, o que reforça a necessidade de abordagens personalizadas no treinamento e na análise de desempenho.

## 6.1 Conclusão

O evento do *COMBINE* é de fundamental importância para o futuro profissional dos jogadores envolvidos, sendo assim um bom desempenho atlético dos mesmos no evento é primordial, por conta disso é fundamental que seja realizada uma análise para avaliar os padrões atléticos ao longo dos anos do evento.

O estudo realizado neste trabalho apresentou métricas com o objetivo de realizar essa avaliação de forma sistemática, com critérios bem definidos, a fim de analisar os padrões atléticos de maneira abrangente.

Com base na avaliação dos jogadores que estavam incluídos na base de dados do *Pro Football Reference* (PFR), referentes ao período de 2000 a 2022, os resultados nos testes físicos avaliados (40 jardas, supino, salto em distância e 3 cones), foram melhorando em média ao longo dos anos, sendo notável a variabilidade desses resultados. Essas flutuações sugerem uma interação complexa de fatores que podem influenciar o desempenho dos atletas, incluindo mudanças na composição dos jogadores, evolução dos métodos de treinamento, ajustes nas políticas de recrutamento das universidades e até mesmo fatores externos, como lesões ou condições meteorológicas durante os testes.

Dessa forma, os modelos estatísticos mostraram padrões específicos relacionados ao desempenho dos jogadores nos testes físicos ao longo do período investigado. Foi constatada uma forte conexão entre as variáveis explicativas e os resultados dos testes físicos,

indicando a influência significativa da composição física e de seu subgrupo posicional no desempenho atlético.

Com base nos modelos estatísticos gerados, a comparação entre os valores médios dos testes físicos realizados e os valores médios previstos ao longo dos anos para cada modelo destacou-se tanto pontos positivos quanto áreas a serem aprimoradas.

No que se refere aos testes individuais, o modelo do teste de 40 jardas apresentou um desempenho consistente e preciso, como evidenciado pela menor raiz do erro quadrático médio e pelo alto pseudo- $R^2$ . Isso indica uma boa capacidade de previsão e uma boa explicação para a variação nos resultados desse teste ao longo dos anos.

Entretanto, os resultados do teste do Supino e teste de Salto em distância evidenciaram desafios na precisão das previsões dos modelos, especialmente em captar as variações ao longo do tempo. Enquanto o modelo do teste do Supino apresentou um pseudo- $R^2$  relativamente baixo, indicando uma explicação inferior da variabilidade dos dados, o modelo do teste de Salto em distância mostrou uma maior variabilidade nas previsões em relação aos dados reais.

A análise dos resultados destacou a relevância de levar em conta o contexto histórico e as tendências ao longo do tempo. A análise dos índices médios ao longo dos anos revelou variações de desempenho, incluindo períodos de melhoria consistente e variações temporais. Essas informações fornecem percepções valiosas para entender não apenas as mudanças no desempenho dos jogadores, mas também os fatores subjacentes que podem influenciar essas tendências.

Além disso, a aplicação do cálculo do *Annual Percentage Change* (APC) reforçou que existe uma tendência de crescimento significativo no desempenho atlético dos jogadores nos testes físicos ao longo do período analisado. Esse resultado confirma a eficácia das estratégias adotadas e a consistência dos avanços alcançados, fornecendo uma base sólida para futuras pesquisas e intervenções no campo do desempenho esportivo.

Em suma, a análise abrangente dos resultados dos testes físicos e das previsões dos modelos estatísticos a partir das métricas definidas oferece visões valiosas sobre a evolução do desempenho atlético dos jogadores ao longo do tempo. Essas descobertas não apenas validam a solidez das etapas de avaliação e a robustez das metodologias empregadas, mas também fornecem uma base sólida para orientar futuras pesquisas.

## 6.2 Trabalhos Futuros

Para projetos futuros, com base nos resultados deste estudo, recomenda-se a realização de testes com uma variedade maior de modelos aplicados aos testes físicos, especialmente nos casos em que os modelos de MLG não apresentaram um bom ajuste. Deste modo, para evitar um possível viés no estudo, é sugerível testar formas diferente de agrupar as posições dos atletas.

Além disso, a respeito da análise temporal, sugere-se segmentar o período analisado para examinar a tendência e padrão do desempenho atlético em intervalos menores, já que o APC mostra uma tendência única ao longo de todo o período.

## 6.3 Anexo

Diversos materiais complementares e informações adicionais foram de relevância ao longo da elaboração deste presente trabalho, uma vez que contribuíram para o embasamento teórico do tema abordado. Abaixo, segue o link do repositório no Github com documentos, gráficos e script que fundamentaram a construção do trabalho. A inserção do repositório se faz importante para complementar a análise realizada, destacando de forma mais ampla e detalhada para o público leitor que possua interesse pelo tema apresentado. A leitura complementar deste repositório é indispensável para a compreensão integral do trabalho.

### 6.3.1 Script

Link para acesso : [GitHub TCC - Análise do Atletismo: Uma Investigação sobre a Performance de Jogadores Universitários Norte Americanos no combine ao Longo do Tempo 2000 - 2022](#)

## CAPÍTULO 7

### REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, A. O que é o combine da NFL? Tudo sobre Futebol Americano. Disponível em: <https://tudosobrefutebolamericano.com/o-que-e-o-combine-da-nfl>. Acesso em: 25.03.2024.

ABDI, H.; WILLIAMS, L.J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 2, n. 4, p. 433-459, 2010.

ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. *Epidemiol. Serv. Saúde*, 24(3), 565-576, 2015.

BAILEY, C. Longitudinal Monitoring of Athletes: Statistical Issues and Best Practices. *Journal of Science in Sport and Exercise*, v. 1900, n. 209, p. 1-12, 2019.

DOBSON, A.J.; BARNETT, A. G. *An Introduction to Generalized Linear Models*, 3rd Ed. Chapman & Hall/CRC, Boca Raton, 2008.

DOMPIER, T. P. et al. Incidence of Concussion During Practice and Games in Youth, High School, and Collegiate American Football Players. *JAMA Pediatrics*, 169(7), 659–665, 2015.

GABBETT, T. J.; JENKINS, D. G; ABERNETHY, B. Physical Performance Tests That Best Predict Repetitive High-Intensity Effort Ability in Professional Rugby League Players. *Journal of Strength and Conditioning Research*, v. 30, n. 2, p. 319–325, 2016.

GALLUP. Football Still Americans' Favorite Sport to Watch. 2018. Disponível em: <https://news.gallup.com/poll/224864/football-americans-favorite-sport-watch.aspx>. Acesso em: 15.03.2024.

HOOTMAN, J.M.; DICK, R.; AGEL, J. Epidemiology of Collegiate Injuries for 15 Sports: Summary and Recommendations for Injury Prevention Initiatives. *Journal of Athletic Training*, v. 42, n. 2, p. 311–319, 2007.

MACHADO, J.C.; BARREIRA, D.; GARGANTA, J. Eficácia ofensiva e variabilidade de padrões de jogo em futebol. *Revista Brasileira de Educação Física e Esporte*, v. 27, n. 4, p. 667-677, 2013.

MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. CRC Press, 1989.

NELDER, J.A.; WEDDERBURN, R. W; M. *Generalized Linear Models*. *Journal of the Royal Statistical Society. Series A (General)*, v. 135, n. 3, p. 370-384, 1972.

NCAA. Results from the 2015 GOALS study of the student-athlete experience. Indianapolis: NCAA, 2016. Disponível em: <https://www.ncaa.org/sites/default/files/GOALS-convention-slidebank-jan2016-public.pdf>. Acesso em: 20.03.2024.

NCAA. Estimated probability of competing in professional athletics. 2018. Disponível em: <http://www.ncaa.org/about/resources/research/estimated-probability-competing-professionalathletics>. Acesso em: 25.03.2024.

NFL. Official 2020 National Football League Record Fact Book. 2020. Disponível em: <https://nflcommunications.com/Documents/2020%20NFL%20Record%20and%20Fact%20Book.pdf>. Acesso em: 01.04.2024.

PATI, C.; GASPARINI, C. As 50 universidades de maior prestígio no mundo. Exame. 2016. Disponível em: <https://exame.com/carreira/as-50universidades-de-maior-prestigio-no-mundo/>. Acesso em: 05.04.2024.

PINO-ORTEGA, J. et al. Desenho de Treinamento, Análise de Desempenho e Identificação de Talentos – Uma Revisão Sistemática sobre as Variáveis Mais Relevantes por meio da Análise de Componentes Principais em Futebol, Basquete e Rugby. *International Journal of Environmental Research and Public Health*, v. 18, n. 5, p. 2642, 2021.

Pro Football Network. What Is the NFL Combine? Since 1985, the NFL Combine has given college football players the platform to show off their skills. What all happens at the Combine? 2024. Disponível em:

<https://www.profootballnetwork.com/what-is-the-nfl-combine/>. Acesso em: 28.03.2024.

ROBBINS, D. W. Positional Physical Characteristics of Players Drafted Into the National Football League. *Journal of Strength and Conditioning Research*, v. 25, n. 10, p. 2661-2667, out. 2011.

RUGG, ADAM ERNEST. Ruptures in the field: The NFL as global corporate media organization and American cultural industry. 2016.

SCHRÖDINGER, E. Zur theorie der fall-und steigversuche an teikhen mit brownscber bewegung. *Physikalische Zeitschrift*, v. 16, p. 133-134, 1915.

STAMBULOVA, N. et al. Career development and transitions of athletes: The ISSP position stand. *\*International Journal of Sport and Exercise Psychology\**, v. 7, p. 395-412, 2009.

STAMBULOVA, N. Talent development in sport: The perspective of career transitions. In: HUNG, E. Tsung-Min; LIDOR, R.; HACKFORT, D. (Eds.). *Psychology of sport excellence*. Morgantown, WV: Fitness Information Technology, p. 63-74. 2009.

SUCHOMEL, T. J.; Nimphius, S.; Stone, M. H. The Importance of Muscular Strength in Athletic Performance. *Sports Medicine*, v. 49, n. 10, p. 1419–1449, 2019.

TEMPLE, V. J.; BAILEY, B. W. Physical activity patterns of youth with Down syndrome. *Intellectual and Developmental Disabilities*, v. 48, n. 3, p. 188-199, 2010.

ZEILEIS, A.; KLEIBER, C.; JACKMAN, S. Regression models for count data in R. *Journal of Statistical Software*, v. 27, n. 8, p. 1–25, 2008.