

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

**Avaliação Esportiva Utilizando Técnicas Multivariadas:
Construção de Indicadores e Sistema Online**

Alexandre Cristovão Maiorano

São Carlos - SP
17 de Junho de 2015

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Avaliação Esportiva Utilizando Técnicas Multivariadas:
Construção de Indicadores e Sistema Online

Alexandre Cristovão Maiorano

Orientador: Prof. Dr. Francisco Louzada Neto

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Estatística da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos - SP
17 de Junho de 2015

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

M227ae

Maiorano, Alexandre Cristovão.

Avaliação esportiva utilizando técnicas multivariadas:
construção de indicadores e sistemas online/ Alexandre
Cristovão Maiorano. -- São Carlos : UFSCar, 2015.
61 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2015.

1. Estatística. 2. Avaliação esportiva. 3. Métodos
univariados e multivariados. 4. Metodologia. I. Título.

CDD: 519.5 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Estatística

Via Washington Luís, Km 235 - C.P. 676 - CGC 45358058/0001-40

FONE: (016) 3351-8292 - Email: ppgest@ufscar.br

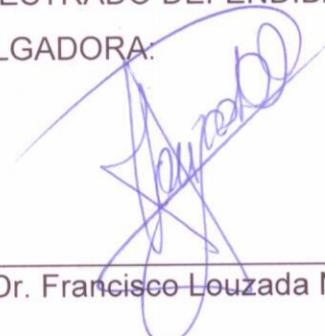
13565-905 - SÃO CARLOS-SP - BRASIL

FOLHA DE APROVAÇÃO

Aluno(a) : Alexandre Cristovão Maiorano

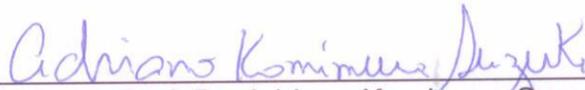
DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 10/10/2014
PELA COMISSÃO JULGADORA.

Presidente



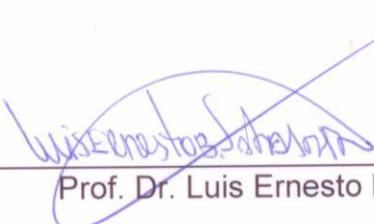
Prof. Dr. Francisco Louzada Neto (ICMC-USP /Orientador)

1º Examinador



Prof. Dr. Adriano Kamimura Suzuki (ICMC-USP)

2º Examinador



Prof. Dr. Luis Ernesto Bueno Salazar (DEs-UFSCar)

Agradecimentos

Agradeço primeiramente à Deus por ter iluminado minha trajetória e feito com que eu tenha escolhido os melhores caminhos durante toda minha vida, além de ter me dado saúde e determinação para conseguir finalizar minha dissertação.

À minha família, principalmente meus pais, Luiz e Wilneide, por todo esforço, amor, compreensão e apoio para dar continuidade aos meus estudos. À minha irmã Amanda pelo companheirismo e motivação desde quando éramos pequenos.

À minha namorada Bárbara, que amo tanto, pelo amor, carinho, companheirismo e grande apoio em todo esse tempo que estamos juntos.

Ao meu orientador Francisco Louzada Neto pelas oportunidades e toda a experiência que tem me passado nesses anos de trabalho.

Agradeço ao Anderson Ara pela ajuda no desenvolvimento do iSports, pela amizade e conhecimento que sempre tem me passado.

À todos os alunos do Grupo CER que participaram da coleta de dados para realização da pesquisa, bem como a equipe da escolinha Meninos da Vila - São Carlos por ter disponibilizado suas dependências para aplicação dos testes de futebol.

Por fim, agradeço aos docentes e funcionários do Departamento de Estatística da UFSCar, pela formação e estrutura disponível, e à CAPES pelo apoio financeiro.

Resumo

O principal objetivo do trabalho é apresentar ferramentas estatísticas que permitam a comparação de indivíduos em uma determinada modalidade esportiva. Particularmente, o estudo exposto é voltado à avaliação de desempenho em futebol, utilizando métodos univariados e multivariados. A abordagem univariada é dada pela metodologia Z-CELAFISCS, desenvolvida com o propósito de identificar talentos no esporte. As abordagens multivariadas são dadas pela construção de indicadores, mais especificamente por meio da análise de componentes principais, análise fatorial e cópulas. A obtenção desses indicadores possibilita a redução da dimensionalidade do estudo, fornecendo melhor interpretação dos resultados e melhor comparabilidade entre o desempenho e ranqueamento dos indivíduos. Para facilitar a utilização da metodologia aqui estudada foi construído um sistema estatístico online chamado de i-Sports.

Palavras-chave: *Avaliação Esportiva, Métodos Univariados e Multivariados, Metodologia Z-CELAFISCS, Análise de Componentes Principais, Análise Fatorial, Teoria de Cópulas.*

Abstract

The main objective of this research is to provide statistical tools that allow the comparison of individuals in a specified sports category. Particularly, the present study is focused on the performance evaluation in football using univariate and multivariate methods. The univariate approach is given by Z-CELAFISCS methodology, which was developed with the purpose of identifying talents in the sport. The multivariate approaches are given by the construction of indicators, specifically by means of principal component analysis, factor analysis and copulas. These indicators allows the reduction of the dimensionality of the data in studying, providing better interpretation of the results and improving comparability between the performance and assortment of individuals. To facilitate the use of the methodology studied here was built an online statistical system called i-Sports.

Keywords: *Sport Evaluation, Univariate and Multivariate Methods, Z-CELAFISCS Methodology, Principal Components Analysis, Factor Analysis, Copula Theory.*

Sumário

1	Introdução	1
1.1	Objetivos	3
1.2	Organização dos Capítulos	3
2	Materiais e Métodos	4
2.1	Testes Práticos para Avaliação de Desempenho	4
2.1.1	Teste de Passe de Mor e Christian	5
2.1.2	Teste de Drible de Cinco Cones	5
2.1.3	Teste de Chute após Passe	6
2.1.4	Teste de 1000 Metros em Pista	7
2.1.5	Teste de Velocidade Cíclica de 20 Metros	8
2.1.6	Teste de Potência Anaeróbica (R.A.S.T.)	9
2.2	Sumarização de Dados	10
2.2.1	Método Z	11
2.2.2	Análise de Componentes Principais	12
2.2.3	Análise Fatorial	14
2.2.4	Teoria de Cópulas	17
3	Resultados	23
3.1	Análise Descritiva	24
3.2	Construção dos Indicadores	25
3.2.1	Via A.C.P.	25
3.2.2	Via Análise Fatorial	29
3.2.3	Via Cópula	35
3.3	Indivíduos com Melhores Desempenho	41
3.3.1	Método Z	41

3.4	Jogadores com Melhores Avaliações	42
4	Sistema i-Sports	44
4.1	Abordagem Web	44
4.2	Implementação e Avaliação	45
5	Considerações Finais	56
	Referências Bibliográficas	57

Lista de Tabelas

3.1	Medidas descritivas para as variáveis coletadas.	25
3.2	Matriz de correlação observada para os testes físicos.	27
3.3	Autovalores e % de variabilidade explicada obtida por meio da A.C.P. para testes físicos.	27
3.4	Autovetores obtidos por meio da A.C.P. para testes físicos.	27
3.5	Matriz de correlação observada para os testes técnicos.	28
3.6	Autovalores e % de variabilidade explicada obtida por meio da A.C.P. para testes técnicos.	28
3.7	Autovetores obtidos por meio da A.C.P. para testes técnicos.	29
3.8	Matriz de correlação dos testes.	30
3.9	Resultado do teste Bartlett de esfericidade.	30
3.10	Medida de adequação da amostra - MSA.	30
3.11	Autovalor, proporção de variabilidade explicada e acumulada.	30
3.12	Cargas fatoriais rotacionadas.	31
3.13	Variância explicada por cada fator rotacionado.	31
3.14	Comunalidade final estimada.	32
3.15	Coefficientes dos escores fatoriais.	33
3.16	Ajuste da cópula Normal.	38
3.17	Teste de qualidade de ajuste.	38
3.18	Indivíduos mais consistentes.	39
3.19	Teste de independência multivariada.	40
3.20	Família de cópulas.	40
3.21	Ajuste da cópula Normal.	40
3.22	Indivíduos que mais se destacaram no CHUTE segundo a metodologia Z. .	41
3.23	Indivíduos que mais se destacaram no DRIBLE segundo a metodologia Z. .	41

3.24	Indivíduos que mais se destacaram no PASSE segundo a metodologia Z. . .	41
3.25	Indivíduos que mais se destacaram em PR segundo a metodologia Z. . . .	41
3.26	Indivíduos que mais se destacaram em VO2MAX segundo a metodologia Z.	41
3.27	Indivíduos que mais se destacaram em CL20 segundo a metodologia Z. . .	42
3.28	Melhores jogadores fisicamente.	42
3.29	Jogadores com maior habilidade técnica.	42
3.30	Jogadores com melhores avaliações - Escore geral.	42
3.31	Indivíduos mais consistentes.	43

Lista de Figuras

2.1	Disposição e ordem de aplicação dos testes aplicados.	5
2.2	Ilustração do teste de passe de Mor e Christian.	6
2.3	Ilustração do teste de drible de 5 cones.	6
2.4	Ilustração do teste de chute após passe.	7
2.5	Ilustração do teste de 1000 metros em pista.	9
2.6	Ilustração do teste de velocidade cíclica de 20 metros.	9
2.7	Ilustração do teste de potência anaeróbica.	10
2.8	Posição de um indivíduo com $Z = 1$ em relação aos demais.	11
2.9	Obtenção dos indicadores por meio da A.C.P.	13
3.1	Histograma - descrição geral dos indivíduos.	25
3.2	Histograma - desempenho nos testes.	25
3.3	Gráfico das duas primeiras componentes - Escore técnico.	29
3.4	Gráfico de cotovelo e proporção de variância explicada.	31
3.5	Cruzamento entre primeiro e segundo fator.	32
3.6	Cruzamento entre primeiro e terceiro fator.	33
3.7	Cruzamento entre segundo e terceiro fator.	33
3.8	Cruzamento entre todas as variáveis da análise.	36
3.9	Normal bivariada ($r = \frac{1}{2}$).	37
3.10	Cópula Normal ($r = \frac{1}{2}$).	37
3.11	Dados originais (esquerda) e transformados pelo rank (direita).	38
3.12	Comparação entre possíveis cópulas simuladas.	39
3.13	QQplot para comparação entre cópula empírica e cópula Normal ($\theta = 0,7813$).	39
4.1	Estrutura geral do sistema, bem como as linguagens utilizadas.	46

4.2	Acesso disponível para cada tipo de usuário.	46
4.3	Área de login da escola cadastrada.	48
4.4	Formulário de cadastro de aluno/atleta no sistema i-Sports.	48
4.5	Menu de relatórios para a escola Meninos da Vila no i-Sports.	49
4.6	Relatório geral dos alunos.	50
4.7	Relatório geral dos testes.	51
4.8	Melhores jogadores fisicamente.	52
4.9	Jogadores com maior habilidade técnica.	52
4.10	Jogadores com melhores avaliações.	53
4.11	Jogadores mais consistentes.	53
4.12	Aluno com maior escore geral e maior consistência.	54
4.13	Aluno com menor escore geral.	55

Capítulo 1

Introdução

Em tempos de eventos esportivos mundiais realizados no país do Futebol, o Brasil, cada vez mais o assunto esporte se torna o centro das atenções. Nesse sentido, surgem diversas abordagens estatísticas sobre o tema e, dentre elas, a utilização de indicadores para comparar o desempenho de indivíduos/atletas.

No escopo geral da composição de um atleta, temos que sua condição física é imprescindível para que seu rendimento em determinado esporte seja de excelência [1, 2, 3, 4, 5, 6]. Um atleta com condição física apurada tende a obter resultados mais satisfatórios quando comparados com aqueles que são desprovidos dessa característica, pois a alta capacidade física está relacionada com o alto desempenho em indicadores de performance, menor fadiga de jogo e até mesmo com o desempenho do próprio time [7]. Além disso, os melhores jogadores devem conseguir se adaptar a demanda física do jogo, que é multifatorial. Eles não precisam ter uma característica extraordinária em uma única área, mas sim possuir um nível elevado em todas elas [1].

Porém, este não é o único fator que influencia em seu desempenho. A capacidade que um atleta/equipe possui está relacionada com muitas outras variáveis como a capacidade psicológica para suportar altas pressões, principalmente quando envolvem consequências negativas [3], e características de personalidade que facilitam o aprendizado, treinamento e competição [2]. Podemos observar também a existência de fatores cognitivos e de inteligência de jogo como antecipação, tomada de decisão e até inteligência e criatividade [2, 8, 9]. Ainda, devemos levar em conta a maturação do atleta [5, 10], fatores genéticos [2, 11] e sociológicos [2] como, por exemplo, a influência familiar, o papel do treinador e a prevenção e detecção de lesões.

Com a finalidade de avaliar o desempenho de atletas praticantes do futebol de campo,

testes de campo específicos podem ser usados e empiricamente analisados, a fim de detectar indivíduos que se destaquem dentre os demais. Esses testes são considerados uma alternativa satisfatória em relação aos testes laboratoriais [12], pois geralmente são simples, populares entre jogadores e técnicos e não necessitam de equipamentos para sua aplicação [7].

Outro fator que contribui na utilização desses testes é que os testes laboratoriais envolvem exercícios que não são específicos do futebol e os protocolos usados impactam na motivação dos jogadores [12]. Além disso, tais testes geralmente são caros [7].

Dessa forma, para compararmos o desempenho de atletas/indivíduos em determinado esporte, devemos utilizar indicadores que reflitam especialmente a realidade da aptidão física e técnica. Como uma das principais características de um bom indicador, temos que este deve ser calculado de forma simples e também ser de fácil interpretação.

Neste contexto, análises estatísticas são utilizadas para mensurar as habilidades físicas e técnicas de indivíduos, permitindo a comparabilidade dos mesmos. Para este fim, propomos um problema de caráter social, a comparação do desempenho de atletas em diferentes testes aplicados, utilizamos uma metodologia proposta por [13], a construção de indicadores por meio da Análise de Componentes Principais, da Análise Fatorial e também pela metodologia de Cópulas e desenvolvemos um sistema online e inovador feito a partir de *softwares* livres e gratuitos.

No nosso conhecimento, não existe na literatura trabalhos que tratam de forma sistemática procedimentos de testes específicos iniciando-se com a coleta dos dados, desenvolvimento teórico, aplicação e desenvolvimento de um *software* que possibilita uma estrutura integrada. Isso nos reporta a uma estrutura de inovação em estatística aplicada ao esporte que possivelmente deverá ser comum no futuro próximo, inclusive para outras áreas do conhecimento.

Trabalho semelhante foi apresentado por [14], em que foi desenvolvido um sistema online utilizando métodos Fuzzy para identificação de aptidão do indivíduo em uma lista específica de esportes. Nosso trabalho se baseia na estrutura empregada pelos autores, mas engloba métodos estatísticos uni e multivariados, diferentes linguagens e *softwares* e a exclusiva utilização do *software* R [15] como meio de cálculo dos resultados em um sistema dinâmico de conexão com o servidor.

Para exemplificar as metodologias utilizadas, fizemos a aplicação de testes para avaliação de desempenho em futebol em uma escola de futebol de campo da cidade de São

Carlos - SP e apresentamos os diferentes relatórios gerados pelo sistema criado, o i-Sports (disponível em <http://www.mwstat.com/isports>).

1.1 Objetivos

O objetivo principal deste trabalho é apresentar ferramentas estatísticas que permitam a comparação de indivíduos em qualquer modalidade esportiva. Inicialmente, o estudo apresentado é voltado à avaliação de desempenho no futebol, sendo que são apresentadas abordagens univariadas e multivariadas.

Como resultado, apresentamos as etapas de criação dos indicadores utilizados, exibimos os indivíduos que se destacaram dos demais nos testes e mostramos a utilização dos indicadores criados em um sistema estatístico online.

1.2 Organização dos Capítulos

O presente trabalho está organizado como segue: no Capítulo 2 apresentamos a metodologia de aplicação e coleta de dados, bem como as técnicas estatísticas utilizadas. No Capítulo 3 exibimos os resultados da criação dos indicadores direcionados à modalidade de futebol de campo e no Capítulo 4 a construção e funcionamento do sistema i-Sports. Por fim, no Capítulo 5 apresentamos as considerações finais e trabalhos futuros.

Capítulo 2

Materiais e Métodos

Esta seção apresenta a metodologia de coleta de dados baseada em testes de desempenho voltado ao futebol, bem como as ferramentas estatísticas utilizadas para a análise das informações relativas aos testes práticos.

2.1 Testes Práticos para Avaliação de Desempenho

Como citado anteriormente testes de campo específicos tem sido bastante utilizados para avaliar o desempenho em futebol [1, 6, 7, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Tais testes são alvos de pesquisa tanto para desenvolvimento de novos testes [25, 26] quanto para validação dos mesmos [25, 27, 28].

Para selecionarmos os testes de avaliação de desempenho a serem utilizados no sistema, uma coleção de testes aplicados à avaliação de atletas de futebol de campo foi exibida a um especialista deste esporte, a fim de que selecionasse os testes físicos e técnicos a serem utilizados.

Os testes de habilidade técnica selecionados foram: teste de passe de Mor e Christian, teste de drible de cinco cones e teste de chute após passe.

Já os testes físicos foram: teste de 1000 metros em pista, teste de velocidade cíclica de 20 metros e o teste de potência anaeróbica (R.A.S.T.).

Estrategicamente, para facilitar a coleta dos dados, os testes foram aplicados e dispostos no campo de futebol da maneira indicada na Figura 2.1.

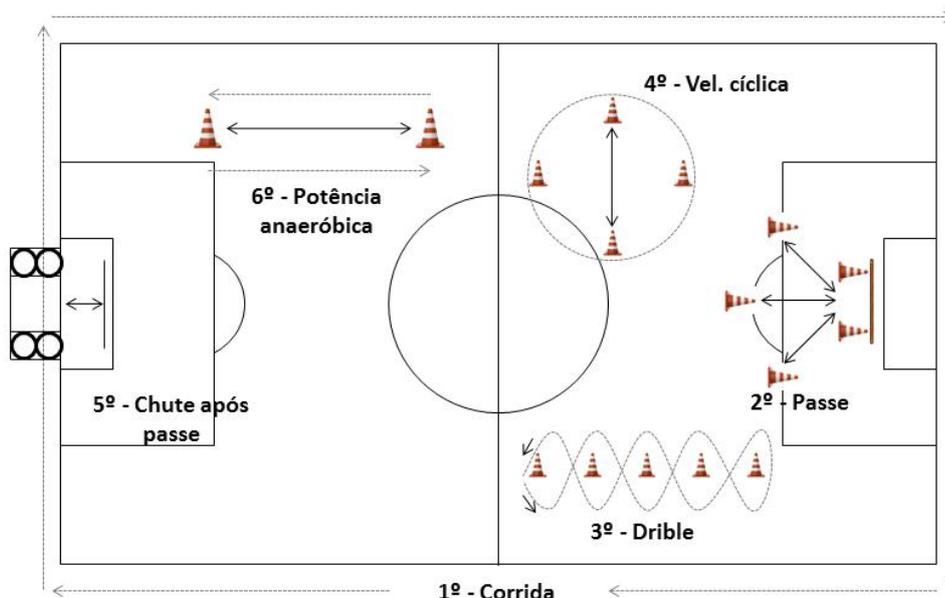


Figura 2.1: Disposição e ordem de aplicação dos testes aplicados.

2.1.1 Teste de Passe de Mor e Christian

O teste de passe de Mor e Christian [29] consiste na demarcação de uma pequena meta, de 91 centímetros de largura e 46 centímetros de altura, feita por dois cones e uma haste de maneira que estará limitando a altura do “gol”. Os indivíduos participantes do teste ficam a uma distância de 14 metros, em três ângulos diferentes: 45°, 90° e 135°, de maneira que cada um terá a disposição quatro tentativas em cada ângulo, resultando em doze no total. Computa-se 1 ponto para cada acerto, que é considerado válido quando a bola passa por dentro da meta, mesmo que tenha tocado os cones e/ou na haste. Dessa forma, a menor pontuação possível é 0 e a maior é 12. A disposição dos cones é exposta na Figura 2.2.

As observações deste teste foram indexadas à variável PASSE.

2.1.2 Teste de Drible de Cinco Cones

Esse teste consiste em uma adaptação do teste de agilidade de [19]. Cinco cones são dispostos a 2,7 metros um do outro em linha reta e o indivíduo deve, dado um lado de partida, ir e voltar driblando os cones em “zigzague”. Este processo é realizado duas vezes e a pontuação obtida é a soma dos tempos realizados nas duas tentativas, de modo que quanto menor for essa soma, maior será o nível de drible e controle de bola do aluno (visto serem ambas capacidades técnicas essenciais ao futebol profissional). A disposição

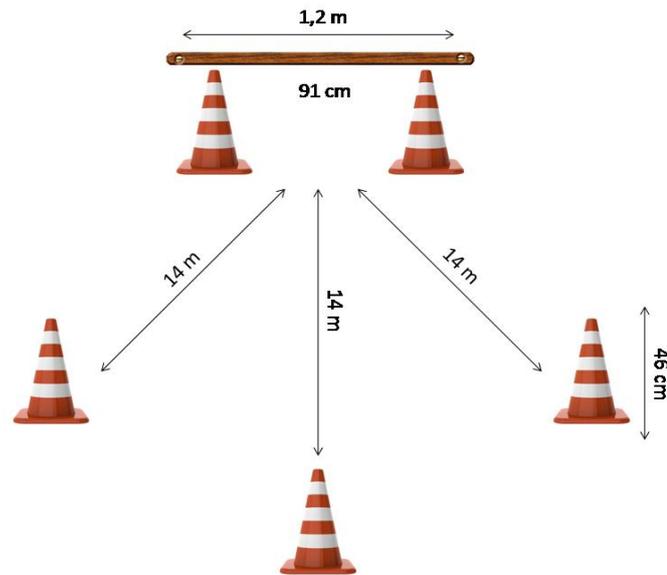


Figura 2.2: Ilustração do teste de passe de Mor e Christian.

dos cones é exposta na Figura 2.3.

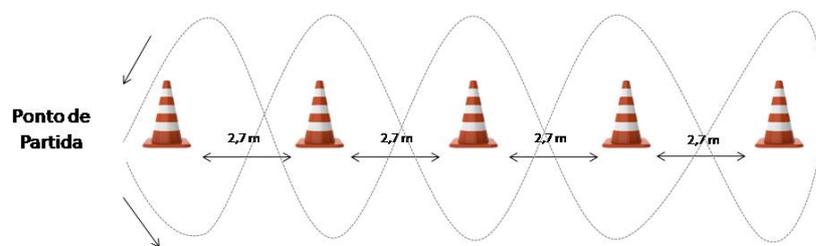


Figura 2.3: Ilustração do teste de drible de 5 cones.

As observações deste teste foram indexadas à variável DRIBLE.

2.1.3 Teste de Chute após Passe

O teste de chute após passe [29] consiste em um gol regulamentar de futebol (7,32 metros x 2,4 metros) dividido em áreas de pontuação por duas cordas fixadas no travessão

a 1,2 metros de cada trave. Cada lado contém dois bambolês de 1,2 metros de diâmetro, dispostos de forma que um fique em cima e o outro embaixo. Assim, temos quatro zonas de pontuação, formadas por quatro circunferências de 1,2 metros de diâmetro.

O objetivo é chutar a bola em movimento, que é passada por um instrutor a cerca de 11 metros de distância do alvo, acertando as zonas de pontuação de modo que o jogador obtenha 6 pontos para cada acerto nas zonas superiores, 1 ponto para cada acerto nas zonas inferiores e nenhum ponto para qualquer outro caso. O atleta tem direito a 5 tentativas, sendo que se for preciso, uma tentativa pode ser feita novamente. Além disso, o jogador pode ter uma chance para treino antes do início da avaliação. Apenas acertar o bambolê já é visto como acerto, ou seja, a bola não precisa necessariamente passar por dentro das circunferências. A disposição dos bambolês é exposta na Figura 2.4.

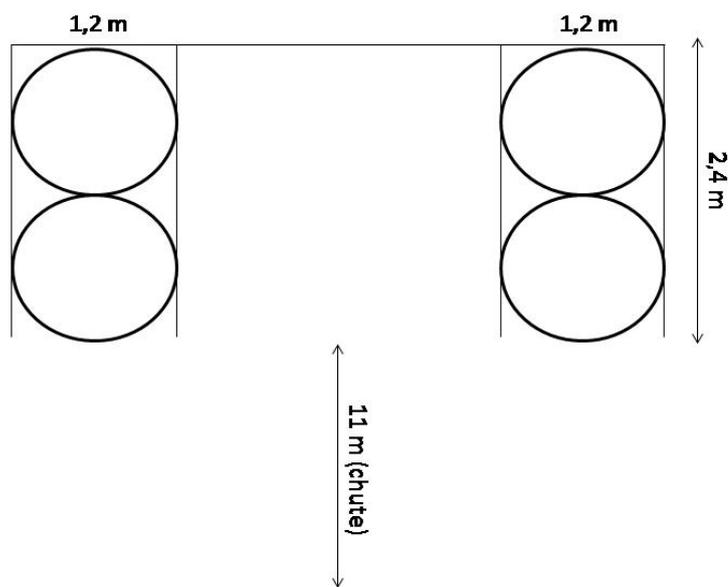


Figura 2.4: Ilustração do teste de chute após passe.

As observações deste teste foram indexadas à variável CHUTE.

2.1.4 Teste de 1000 Metros em Pista

Esse teste é utilizado para medir a potência aeróbica do atleta, uma das capacidades físicas que mais influenciam no desempenho dos jogadores, pois em uma partida de futebol, os jogadores, com exceção do goleiro, percorrem cerca de 10 quilômetros com velocidades

baixas e moderadas em mais de 30% do tempo de jogo [30].

O teste de 1000 metros em pista é indireto (geralmente de campo), ou seja, ele fornece apenas uma estimativa do volume máximo de oxigênio, diferentemente de um teste laboratorial (*gold standard* - padrão ouro), que evidentemente possui mais recursos que o de campo. Entretanto, a validação dos 1000 metros em pista foi comprovada por [31].

A aplicação do teste de 1000 metros é simples: os atletas devem correr uma distância de 1000 metros no menor tempo possível, preferivelmente em uma pista de atletismo ou num campo de futebol, permitindo o registro do tempo (em segundos) gasto pelo aluno para a execução da tarefa. Assim, para o cálculo do valor do consumo máximo de oxigênio (VO2MAX), é utilizada a seguinte fórmula:

$$X = \frac{652,17 - Y}{6,762} \quad (2.1)$$

em que X representa o consumo máximo de oxigênio (ml/kg/min), Y é o tempo de corrida (em segundos) nos 1000 metros e 652,17 e 6,762 são constantes.

Para a medição do tempo que é utilizado na expressão (2.1) são necessários cronômetros ou câmeras que consigam calcular de forma precisa o tempo de cada indivíduo. Além disso, tal atividade deve ser aplicada em horários adequados sem expor os participantes a períodos de alta temperatura, possibilitando a prevenção de problemas físicos devido à intensidade da atividade.

Na Figura 2.5 temos como exemplo um campo de futebol de 100 metros x 65 metros, na qual seriam necessárias aproximadamente três voltas completas para a execução dos 1000 metros.

As observações deste teste foram indexadas à variável VO2MAX.

2.1.5 Teste de Velocidade Cíclica de 20 Metros

Nesse teste o aluno deve partir da linha de saída, após ouvir o apito, e percorrer os 20 metros no menor tempo possível, registrando-se o tempo gasto para completar o percurso. Em outras palavras, o teste de velocidade cíclica de 20 metros consiste numa corrida cíclica, na qual o comprimento da circunferência deve ter aproximadamente 6,37 metros de diâmetro (possibilitando que a circunferência tenha seus 20 metros).

O objetivo é registrar o tempo que o indivíduo gasta para completar o teste. A disposição do teste é exibida na Figura 2.6 e as observações deste teste foram indexadas

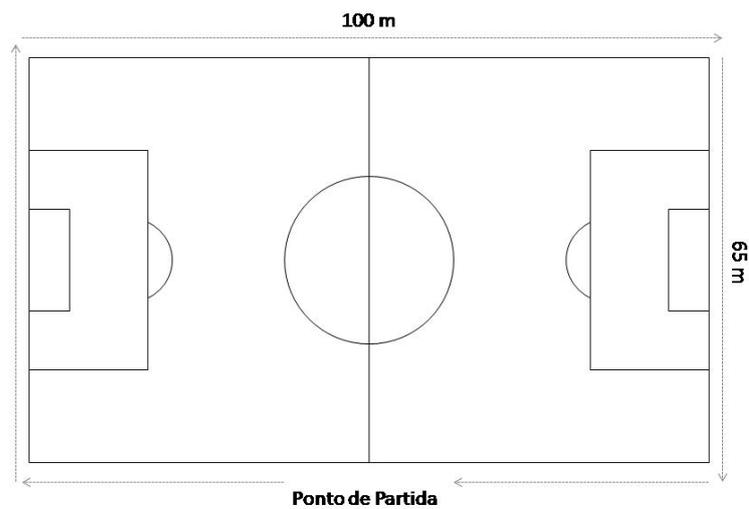


Figura 2.5: Ilustração do teste de 1000 metros em pista.

à variável CL20.

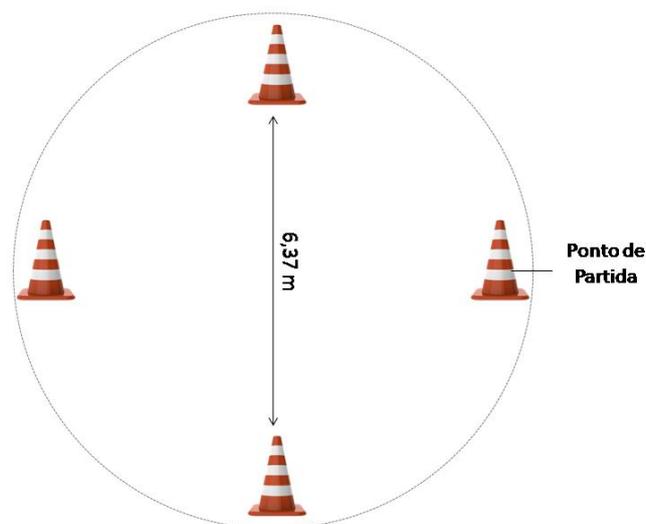


Figura 2.6: Ilustração do teste de velocidade cíclica de 20 metros.

2.1.6 Teste de Potência Anaeróbica (R.A.S.T.)

Esse teste [32] é utilizado para medir a potência anaeróbica do atleta, outra grandeza de extrema importância, uma vez que ela está presente em diversas situações em uma partida de futebol como, por exemplo, quando um zagueiro alto, mas lento, deve acompanhar

incansavelmente um atacante bastante rápido e driblador num contra-ataque durante a parte final do jogo. É possível que tanto o zagueiro quanto o atacante já tenham atingido o limite de consumo de oxigênio na produção energética, proporcionando um estado no qual a resistência é indispensável. É apenas uma circunstância que não só é bastante real, mas também representa categoricamente a necessidade de incluir tal variável entre aquelas que são provavelmente as mais influentes no meio futebolístico. A disposição do teste é exibida na Figura 2.7.

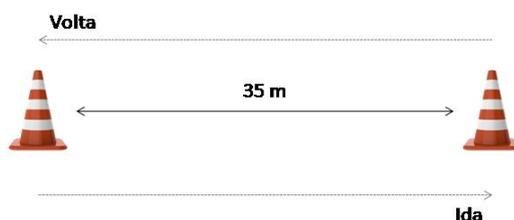


Figura 2.7: Ilustração do teste de potência anaeróbica.

Neste teste, o aluno deve correr seis vezes a distância de 35 metros, na maior velocidade possível, com um intervalo de 10 segundos entre cada deslocamento. Assim, pode-se calcular a potência no tiro i : $P_i = \text{Peso} \times (35^2 / (t_i)^3)$, em que t_i é o tempo no tiro i . A Potência Média (PM) e a Potência Relativa (PR) são dadas por: $PM = \frac{\sum_{j=1}^6 P_i}{6}$ e $PR = \frac{\text{Potência Média}}{\text{Peso}}$, respectivamente.

2.2 Sumarização de Dados

Nesta seção são exibidas as ferramentas estatísticas utilizadas para a análise das informações relativas aos testes práticos de desempenho.

Um conjunto de técnicas estatísticas uni e multivariadas foi aplicado com o intuito de criarmos indicadores capazes de avaliar e comparar as habilidades dos indivíduos.

2.2.1 Método Z

Uma proposta para identificar atletas que se destacam em determinada prática esportiva, denominada “estratégia Z-CELAFISCS”, foi apresentada por [13]. Esse instrumento permite diagnosticar o perfil de aptidão física ou técnica e determinar, em termos percentuais, o quanto um determinado indivíduo se aproxima ou se afasta da média populacional. O “Z” que aparece no título vem da padronização da distribuição Normal:

$$Z = \frac{x - \mu}{\sigma}, \quad (2.2)$$

em que x = valor observado para um indivíduo específico, μ = média populacional e σ = desvio padrão populacional.

Por exemplo, se Z for igual a 1 indica que tal indivíduo estará um desvio padrão acima da média populacional que, em termos percentuais, significaria ter um desempenho melhor que 84,13% da população de sua característica em determinada variável. A Figura 2.8 exibe a posição desse indivíduo no gráfico da densidade da distribuição Normal.

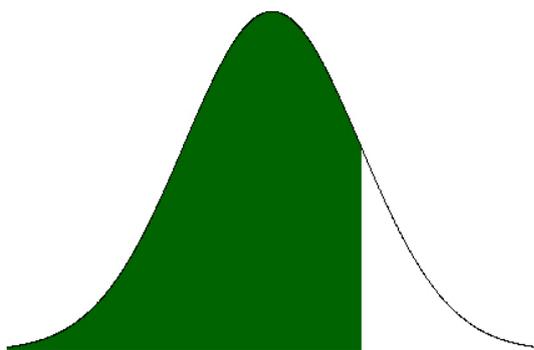


Figura 2.8: Posição de um indivíduo com $Z = 1$ em relação aos demais.

Esta estratégia é um método interessante, pois pode ser considerada como um instrumento de fácil aplicação e baixo custo, podendo ser utilizada em larga escala. Por outro lado, sua eficácia pode ser questionada, pois se trata de uma análise univariada (que considera apenas a informação de uma única variável). Dessa forma, não considera as relações existentes entre as variáveis e nem leva em consideração as diferenças das importâncias relativas dessas variáveis em cada categoria. Esses dois elementos são fundamentais na análise de desempenho do atleta.

Assim, a utilização de escores padronizados nos processos de promoção de talentos esportivos parece ser um passo importante frente à subjetividade ainda presente nesses

processos. Contudo, a busca por análises mais sofisticadas se faz desejável para que se possa caminhar em direção a uma técnica que permita abordar as relações entre as variáveis e a importância das mesmas nos diferentes períodos que envolvem o treinamento em longo prazo.

2.2.2 Análise de Componentes Principais

A Análise de Componentes Principais (A.C.P.) é uma técnica de análise multivariada que consiste em estudar a estrutura de interdependência de um conjunto de variáveis observadas em um determinado conjunto de dados. O estudo desta interdependência considera a análise da estrutura de variâncias e covariâncias das variáveis, buscando a obtenção de combinações lineares que retenham o máximo possível de informação contida nas variáveis originais. Essas combinações são os chamados componentes principais e por meio deles podemos reduzir a dimensionalidade do problema em estudo, facilitando a análise e interpretação das relações entre as variáveis estudadas.

Seja Σ a matriz de covariâncias associada ao vetor aleatório $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ e $(\lambda_i, \mathbf{e}_i)$ os pares de autovalor-autovetor de Σ para $i = 1, 2, \dots, p$, em que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, o i -ésimo componente principal é dado por:

$$Y_i = \mathbf{e}_i' \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \quad i = 1, 2, \dots, p. \quad (2.3)$$

Suas principais características são: $Var(Y_i) = \mathbf{e}_i' \Sigma \mathbf{e}_i$ e $Cov(Y_i, Y_k) = \mathbf{e}_i' \Sigma \mathbf{e}_k = 0$, para $i \neq k$ e $\mathbf{e}_i = (e_{i1}, \dots, e_{ip})$, e $\sum_{i=1}^p Var(Y_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p Var(X_i) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}$.

Além disso, podemos obter os coeficientes de correlação entre os componentes Y_i e as variáveis X_k da seguinte maneira:

$$\rho_{X_k, Y_i} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad \text{para } i, k = 1, 2, \dots, p.$$

Dessa forma, a A.C.P. identifica um novo conjunto de eixos ortogonais, cujos novos eixos são os componentes principais e os seus valores são os escores. Cada nova variável é uma combinação linear das variáveis originais, sendo que a primeira nova variável sempre explicará a maior parte da variância dos dados e as p novas variáveis serão não correlacionadas.

Esta metodologia é frequentemente utilizada, uma vez que é um passo intermediário em muitas investigações como, por exemplo, em regressão múltipla, em análise de agrupamentos, etc.



Figura 2.9: Obtenção dos indicadores por meio da A.C.P.

A Figura 2.9 representa o esquema da obtenção do indicador por meio da A.C.P. Os indicadores criados foram formulados a partir da soma dos escores fatoriais padronizados, ponderados pelas respectivas porcentagens de variabilidade explicada de cada componente. A fórmula matemática resultante desse processo é dada por:

$$I_m = \sum_{j=1}^k \frac{\lambda_j}{tr(R)} CP_{jm}, \quad (2.4)$$

em que I_m = valor do indicador do m -ésimo jogador, λ_j = j -ésima raiz característica da matriz de correlação R_{pxp} das variáveis utilizadas, k = número de componentes principais escolhidas, CP_{jm} = escore fatorial do m -ésimo jogador, na componente j e tr = traço da matriz de correlação R_{pxp} .

Buscando facilitar a comparação dos indicadores criados, transformamos a base dos mesmos, de modo que os novos valores sejam apresentados dentro do intervalo entre 0 e 100:

$$I_m^* = \frac{(I_m - I_{min})}{(I_{max} - I_{min})} \times 100, \quad (2.5)$$

em que I_m^* = valor do indicador transformado do m -ésimo jogador, I_m = valor do indicador observado no m -ésimo jogador, I_{max} = indicador máximo observado na amostra e I_{min} = indicador mínimo observado na amostra.

Utilizando a A.C.P., os seguintes índices de desempenho foram criados:

Escore Físico

O escore físico é um indicador que mede o desempenho do indivíduo nos testes físicos. Utilizando a Análise de Componentes Principais, pode-se reduzir a dimensão de variáveis em estudo levando em conta a estrutura de interdependência das variáveis que medem a capacidade física do indivíduo. Assim, por meio da combinação dessas variáveis, obtém-se

um conjunto menor de variáveis que expliquem um mesmo fenômeno. Os testes aplicados para verificar a capacidade física foram o teste de potência anaeróbica (R.A.S.T.), sumarizado pela potência relativa (PR), o teste de 1000 metros em pista, sumarizado pelo consumo máximo de oxigênio (VO2MAX) e o teste de velocidade cíclica de 20 metros (CL20).

Escore Técnico

O escore técnico é um indicador que avalia o desempenho dos indivíduos nos testes de habilidades técnicas. Este indicador também foi desenvolvido a partir da A.C.P. e, para medirmos a habilidade técnica do indivíduo, utilizamos o testes de chute e passe de Mor e Christian (PASSE), teste de drible de 5 cones (DRIBLE) e o teste de chute após o passe (CHUTE).

2.2.3 Análise Fatorial

Segundo [33], a Análise Fatorial tem como principal objetivo descrever a variabilidade de um conjunto de dados utilizando um número menor de variáveis não observáveis, denominados fatores comuns, que estão relacionadas ao conjunto de dados por meio de um modelo linear. Neste modelo, parte da variabilidade dos dados é atribuída aos fatores comuns e o restante é atribuída às variáveis que não foram incluídas no modelo, ou seja, o erro aleatório.

Dado que o vetor aleatório \mathbf{X} , associado aos dados, com p componentes (variáveis) tem vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, o modelo fatorial postula que \mathbf{X} é linearmente dependente de algumas variáveis aleatórias não observáveis F_1, F_2, \dots, F_m chamadas fatores comuns e p adicionais fontes de variação $\epsilon_1, \epsilon_2, \dots, \epsilon_p$, chamados de erros ou, algumas vezes de fatores específicos. Em particular o modelo da Análise Fatorial é, na notação matricial, dado por:

$$(\mathbf{X} - \boldsymbol{\mu})_{p \times 1} = \mathbf{L}_{p \times m} \times \mathbf{F}_{m \times 1} + \boldsymbol{\epsilon}_{p \times 1}, \quad (2.6)$$

em que μ_i = média da variável i , ϵ_i = i -ésimo fator específico, F_j = j -ésimo fator comum e l_{ij} = carga fatorial da i -ésima variável no j -ésimo fator.

A matriz \mathbf{L} é formada pelos coeficientes l_{ij} , em que l_{ij} é a carga fatorial da i -ésima variável no j -ésimo fator, de modo que a matriz \mathbf{L} é a matriz de cargas fatoriais. Note

que o i -ésimo fator específico ϵ_i está associado somente à i -ésima resposta de X_i .

Além disso, os vetores aleatórios \mathbf{F} e $\boldsymbol{\epsilon}$ satisfazem as seguintes condições: \mathbf{F} e $\boldsymbol{\epsilon}$ são independentes, $E(\mathbf{F}) = \mathbf{0}$, $Cov(\mathbf{F}) = \mathbf{I}$, $E(\boldsymbol{\epsilon}) = \mathbf{0}$ e $Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = diag(\Psi_1, \dots, \Psi_p)$.

[33] afirmam que este modelo fatorial ortogonal implica uma estrutura de covariâncias para \mathbf{X} . A partir da expressão (2.6) temos que:

$$\begin{aligned} (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' &= (\mathbf{LF} + \boldsymbol{\epsilon})(\mathbf{LF} + \boldsymbol{\epsilon})' \\ &= (\mathbf{LF} + \boldsymbol{\epsilon})((\mathbf{LF})' + \boldsymbol{\epsilon}') \\ &= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\epsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\epsilon}' + \boldsymbol{\epsilon}\boldsymbol{\epsilon}'. \end{aligned} \quad (2.7)$$

Então,

$$\begin{aligned} \boldsymbol{\Sigma} &= Cov(\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\ &= \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\epsilon}') + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \\ &= \mathbf{LL}' + \boldsymbol{\Psi}. \end{aligned} \quad (2.8)$$

Por meio desta equação temos que $\sigma_{ii} = (l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2) + \Psi_i$, ou seja, a variância de X_i é composta pela porção $(l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2)$ explicada pelos m fatores comuns, chamada de comunalidade, e a porção Ψ_i , chamada de variância específica.

Neste contexto, a i -ésima comunalidade é a soma de quadrados das cargas fatoriais da i -ésima variável nos m fatores comuns, e a variância específica é o que não é explicada por estes fatores.

De acordo com [33], o modelo fatorial assume que $p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2}$ variâncias e covariâncias do vetor aleatório \mathbf{X} podem ser reproduzidas por pm cargas fatoriais l_{ij} e p variâncias específicas Ψ_i . Quando $m = p$, qualquer matriz de covariâncias $\boldsymbol{\Sigma}$ pode ser reproduzida exatamente como \mathbf{LL}' , com $\boldsymbol{\Psi} = \mathbf{0}$. A vantagem da aplicação desta técnica ocorre quando $m < p$.

Nos casos em que as variáveis apresentam diferentes escalas de unidades, é usual trabalhar com variáveis padronizadas, cuja matriz de covariâncias amostral é igual a matriz de correlações amostral \mathbf{R} das observações x_1, x_2, \dots, x_n . A padronização evita o problemas de ter uma variável com variância muito grande influenciando indevidamente a determinação das cargas fatoriais.

Critério para Determinação do Número de Fatores

Embora saibamos que quando $m = p$ a Análise Fatorial é exata, estamos interessados em um m menor que p . Para isso basicamente temos dois principais critérios: critério da raiz latente e critério da percentagem da variância.

O critério da raiz latente aconselha a seleção de fatores que possuam autovalor maior que 1. O raciocínio para este critério é que qualquer fator individual deve explicar a variância de pelo menos uma variável se o mesmo for mantido para interpretação. Cada variável contribui com o valor 1 do autovalor total. Sendo assim, apenas os fatores que tem raiz latente maior que 1 são considerados significantes, enquanto que todos os fatores com raízes latentes menores que 1 são descartados.

Em relação ao critério da percentagem da variância explicada temos que, fixado um valor mínimo ou de base para a variabilidade total explicada, selecionam-se o número de fatores que atingem ou superam este valor.

Rotação de Fatores

Uma ferramenta muito importante na interpretação de fatores é a rotação fatorial, que significa rotacionar os fatores em torno da origem até que alguma outra posição mais interessante seja alcançada. O caso mais simples de rotação é a denominada rotação ortogonal, na qual os eixos são mantidos a 90° . Também é possível rotacionar os eixos sem o ângulo de 90° , esta rotação é denominada rotação oblíqua.

Considerando a rotação ortogonal, o método mais utilizado é o VARIMAX, o qual se concentra na simplificação das colunas da matriz fatorial, maximizando a soma de variâncias de cargas exigidas da matriz fatorial. Segundo [33], o método VARIMAX busca a melhor rotação dos eixos de modo que a nova matriz de cargas fatoriais tenha o maior número de coeficientes nulos.

Escore Geral

O escore geral é um indicador criado a partir dos 6 testes aplicados e tem como característica avaliar o desempenho geral do indivíduo. A metodologia empregada nesse indicador é similar à utilizada nos anteriores e, nesse caso, utilizamos a combinação de fatores comuns para a composição do escore geral.

2.2.4 Teoria de Cópulas

A metodologia de cópulas [34] se insere ao contexto de análise multivariada, sendo que as cópulas são utilizadas para descrever a dependência entre um grupo de variáveis aleatórias e possuem a capacidade de “conectar” a distribuição conjunta às distribuições marginais univariadas.

Seja $\mathbf{X} = (X_1, \dots, X_d)$ um vetor aleatório com funções de distribuição marginais acumuladas contínuas F_1, \dots, F_d . A função de distribuição acumulada de \mathbf{X} pode ser representada como:

$$H(\mathbf{x}) = C \{F_1(x_1), \dots, F_d(x_d)\}, \quad \mathbf{x} \in \mathbb{R}^d, \quad (2.9)$$

em termos de uma única função $C : [0, 1]^d \rightarrow [0, 1]$, chamada de cópula. Nesse caso, C é uma função de distribuição acumulada com dimensão d e distribuições marginais Uniforme (0,1).

O processo de modelagem de uma cópula é baseado principalmente em dois estágios: estimar as marginais $F_1(x_1), \dots, F_d(x_d)$ e estimar os parâmetros da cópula C .

Para o ajuste, assumimos que os dados consistem em n réplicas $\mathbf{X}_1, \dots, \mathbf{X}_n$ do vetor \mathbf{X} cuja f.d.a. H admite a representação dada em (2.9).

A fim de construir um modelo para C devemos testar se $\mathbf{X}_1, \dots, \mathbf{X}_n$ é mutuamente independente, isto é, se podem ser considerados como uma amostra aleatória de H . Esse fato é de extrema importância quando temos, por exemplo, um conjunto de testes aplicados em um mesmo indivíduo em mais de um momento. Caso a hipótese de independência seja rejeitada, devemos trabalhar com outros tipos de modelos que levam em consideração a dependência entre os dados (ex: modelos de séries temporais).

Se a hipótese é aceita, devemos testar a presença de dependência entre os componentes de \mathbf{X} . Nesse caso, testar:

$$H_0 : C = \Pi \text{ contra } H_1 : C \neq \Pi,$$

em que $\Pi(\mathbf{u}) = \prod_{i=1}^d u_i$, $\mathbf{u} \in [0, 1]^d$, é a cópula independente.

Se a independência é rejeitada, devemos ajustar uma família paramétrica de cópula adequada aos dados. Para isso, utilizamos testes da forma:

$$H_0 : C \in \mathcal{C} \text{ contra } H_1 : C \notin \mathcal{C},$$

para várias famílias paramétricas $\mathcal{C} = \mathcal{C}_\theta$.

Ao final, devemos escolher as candidatas em que tal hipótese não é rejeitada e, se possível, avaliar o erro padrão das estimativas dos parâmetros obtidos.

Dentre as possibilidades de estimação da cópula C , temos o método baseado no vetor de ranks, ou postos, que são os números de ordem das observações quando estas estão ordenadas. De acordo com [35], o uso dos ranks faz a estimação de C ser marginalmente livre e, dessa forma, a má especificação de uma das marginais $F_1(x_1), \dots, F_d(x_d)$ não terá consequências sobre a estimativa da cópula.

Cópula Empírica

Uma importante característica dos testes citados é que eles são baseados na cópula empírica, que é um estimador consistente de C [35]. Esta cópula é baseada na própria distribuição empírica dos dados, sendo que nesse caso, não fazemos nenhuma suposição sobre a distribuição dos mesmos.

Assim, supondo que $\widehat{\mathbf{U}}_1, \dots, \widehat{\mathbf{U}}_n$ são as pseudo-observações de C , definidas por $\widehat{\mathbf{U}}_i = \mathbf{R}_i/(n+1)$, em que $\mathbf{R}_i = (R_{i1}, \dots, R_{in})$ e R_{ij} é o rank de X_{ij} entre X_{ij}, \dots, X_{nj} , definimos a cópula empírica como a f.d.a. empírica das pseudo-observações, isto é,

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n 1(\widehat{\mathbf{U}}_i \leq \mathbf{u}), \quad u \in [0, 1]^d. \quad (2.10)$$

Equivalentemente, os componentes das pseudo-observações de C podem ser escritos como $\widehat{U}_{ij} = n\widehat{F}_j(X_{ij})/(n+1)$, em que \widehat{F}_j é a f.d.a. empírica calculada em X_{1j}, \dots, X_{nj} e $n/(n+1)$ é um fator de escala introduzido para evitar problemas na fronteira de $[0, 1]^d$.

Teste de Independência Multivariada

Inspirados em trabalhos de outros autores, [36] sugeriram um teste de independência mútua entre os componentes de \mathbf{X} utilizando a seguinte estatística:

$$I_n = \int_{[0,1]^d} n \left\{ C_n(\mathbf{u}) - \prod_{i=1}^d u_i \right\}^2 d\mathbf{u}. \quad (2.11)$$

Uma das características desse teste é que, sob a hipótese de independência entre X_1, \dots, X_d de \mathbf{X} , pode-se decompor a estatística I_n em $2^d - d - 1$ estatísticas e realizar o teste de independência mútua em determinado subconjunto de variáveis a partir do próprio conjunto de dados utilizado.

Mais detalhes sobre a decomposição da estatística I_n , as suposições levadas em conta, bem como a distribuição da estatística em questão e também sobre como é possível calcular um p-valor para o teste pode ser verificado no trabalho dos próprios autores.

Goodness of Fit

Os testes de qualidade do ajuste considerados neste trabalho são baseados no processo empírico:

$$\mathbb{C}_n(\mathbf{u}) = \sqrt{n} \{C_n(\mathbf{u}) - C_{\hat{\theta}_n}(\mathbf{u})\}, \quad \mathbf{u} \in [0, 1]^d, \quad (2.12)$$

em que C_n é a cópula empírica definida em (2.10) e $C_{\hat{\theta}_n}$ é uma estimativa da cópula C sob a hipótese $H_0 : C \in \{C_\theta\}$. O estimador $\hat{\theta}_n$ de θ que aparece em (2.12) é exclusivamente baseado em ranks e pode ser estimado tanto pelo método dos momentos, quanto por máxima pseudo-verossimilhança.

No entanto, baseado em simulações de Monte Carlo, [37] e [38] verificaram que a estatística

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(\mathbf{u})^2 dC_n(\mathbf{u}) = \sum_{i=1}^n \left\{ C_n(\hat{\mathbf{U}}_i) - C_{\hat{\theta}_n}(\hat{\mathbf{U}}_i) \right\}^2, \quad (2.13)$$

apresenta melhores resultados. Como podemos observar, esta estatística mede a proximidade entre valores da cópula empírica e valores da cópula que está sendo testada.

Um p-valor aproximado para a estatística S_n pode ser obtido por meio de procedimentos baseados em bootstrap paramétrico, cuja validade foi confirmada por [39]. Entretanto, o esforço computacional exigido é alto à medida que o tamanho da amostra aumenta devido aos cálculos necessários como, por exemplo, gerar números aleatórios e estimar parâmetros da cópula a cada interação utilizada. Alternativamente, [40] propuseram uma forma baseada em multiplicadores do teorema central do limite, que apresenta maior eficiência computacional.

As seguir apresentamos os procedimentos necessários para o cálculo do p-valor levando em conta os dois métodos citados.

Abordagem Baseada em Bootstrap Paramétrico Um p-valor aproximado para o teste de qualidade de ajuste definido em (2.13) pode ser obtido da seguinte maneira (para mais detalhes, ver [39]):

1. Calcular a cópula empírica C_n por meio das pseudo-observações $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_n$ e estimar θ , também a partir de $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_n$, usando o estimador $\hat{\theta}_n$ (baseado em ranks).

2. Calcular a estatística teste S_n definida em (2.13).

3. Para um número inteiro N , repetir o seguinte passo para todo $k \in \{1, \dots, N\}$:

- Gerar uma amostra aleatória $\mathbf{X}_1^{(k)}, \dots, \mathbf{X}_n^{(k)}$ da cópula $C_{\hat{\theta}_n}$ e calcular as pseudo-observações $\mathbf{U}_1^{(k)}, \dots, \mathbf{U}_n^{(k)}$.

- Fazer:

$$C_n^{(k)}(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{\mathbf{U}}_i^{(k)} \leq \mathbf{u}), \quad u \in [0, 1]^d,$$

e calcular a estimativa $\theta_n^{(k)}$ de θ a partir de $\mathbf{U}_1^{(k)}, \dots, \mathbf{U}_n^{(k)}$ usando o mesmo estimador do passo 1.

- Calcular um valor aproximado de S_n , sob H_0 , por meio de:

$$S_n^{(k)} = \sum_{i=1}^n \left\{ C_n^{(k)}(\hat{\mathbf{U}}_i^{(k)}) - C_{\theta_n^{(k)}}(\hat{\mathbf{U}}_i^{(k)}) \right\}^2.$$

4. Um p-valor aproximado para o teste é dado por:

$$N^{-1} \sum_{k=1}^N \mathbf{1}(S_n^{(k)} \geq S_n).$$

Como podemos observar, esse procedimento exige um esforço computacional intenso a cada interação do passo 3, que envolve geração de números aleatórios e estimativa dos parâmetros da cópula. De acordo com [35], esse fato é particularmente verdadeiro se $\hat{\theta}_n$ for o estimador de máxima pseudo-verossimilhança.

No pacote **copula** [41] do *software* R o p-valor apresentado é calculado utilizando uma expressão um pouco modificada, dada por:

$$\frac{1}{N+1} \left\{ \sum_{k=1}^N \mathbf{1}(S_n^{(k)} \geq S_n) + \frac{1}{2} \right\}.$$

Segundo [35], essa modificação garante que o valor obtido esteja no intervalo aberto $(0,1)$ de modo que as transformações pelas funções inversas das f.d.a.'s de distribuições contínuas sejam sempre bem definidas.

Abordagem Baseada em Multiplicadores do Teorema Central do Limite A eficiência computacional desse método segue o fato que, sob determinadas condições de

regularidades, o processo \mathbb{C}_n pode ser escrito como:

$$\mathbb{C}_n(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{J}_i(\mathbf{u}) + R_n(\mathbf{u}), \quad (2.14)$$

em que $\mathbb{J}_1, \dots, \mathbb{J}_n$ é um processo estocástico i.i.d. (um processo será do i.i.d. se as variáveis que o constituem são i.i.d.), cuja forma depende do estimador θ_n e da família de cópula $\{C_\theta\}$, e $\sup |R_n(\mathbf{u})|$ tende a 0 em probabilidade.

Seja $\widehat{\mathbb{J}}_{i,n}$ a versão de $\mathbb{J}_{i,n}$ em que todas as quantidades desconhecidas são substituídas por suas estimativas, a abordagem multiplicadora modifica o passo 3 da versão baseada em bootstrap paramétrico da seguinte maneira:

3. Para um número inteiro N , repetir os seguintes passos para todo $k \in \{1, \dots, N\}$:

- Gerar n variáveis aleatórias $Z_1^{(k)}, \dots, Z_n^{(k)}$ i.i.d. com média 0 e variância 1.
- Definir a seguinte aproximação \mathbb{C}_n sob H_0 :

$$\mathbb{C}_n^{(k)}(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i^{(k)} \widehat{\mathbb{J}}_{i,n}(\mathbf{u}).$$

- Calcular uma aproximação de S_n , sob H_0 , utilizando:

$$S_n^{(k)} = \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{C}_n^{(k)}(\widehat{U}_i) \right\}^2.$$

Este procedimento é rápido, pois os termos $\widehat{\mathbb{J}}_{i,n}$ são calculados apenas uma vez. Entretanto, o cálculo desses termos não é trivial e envolve derivadas parciais da função de distribuição acumulada e função de distribuição de probabilidade da cópula com respeito a u_i e também com respeito aos parâmetros (para mais detalhes, ver [40] e [42])

A partir de testes realizados observamos a diferença entre os tempos de cada método para o teste considerando uma cópula Normal com 6 variáveis. Utilizando um notebook com Windows 7, processador Intel[®] Core[™] i7 com dois núcleos de 2,66 GHz e 4 GB de memória RAM, o método baseado em bootstrap paramétrico levou cerca de 19,89 minutos para ser calculado, enquanto que pelo outro método levou 3,47 segundos.

Consistência

A consistência é um indicador que exhibe o desempenho geral do atleta considerando conjuntamente os valores apresentados nos 6 testes avaliados. Essa metodologia tem como essência a metodologia Z , mas nesse caso calculamos a probabilidade acumulada no ponto, obtida a partir de uma cópula multivariada ajustada aos dados.

Um indivíduo com consistência igual a um, é o que obteve a máxima avaliação nos testes em relação à população total. Similarmente, um indivíduo que apresenta consistência igual a zero, é o com o pior desempenho em pelo menos um teste quando comparado com a população.

Capítulo 3

Resultados

A aplicação dos testes foi realizada na escola Meninos da Vila, unidade São Carlos-SP, possibilitando assim, a coleta dos dados e o desenvolvimento prático de toda a metodologia proposta.

Essa rede de escolinhas de futebol foi inaugurada no dia 3 de março de 2004 e possui um espaço de 15 mil metros quadrados que abrange, além dos locais para jogos e treinamentos, arquibancada de concreto, dois vestiários grandes, sala de secretaria e um grande alojamento.

Os agentes que participaram da aplicação dos testes foram alunos da equipe do Centro de Estudos do Risco (CER), que é um laboratório de pesquisa do Departamento de Estatística (DEs-CCET) criado na UFSCar desde 2004 e atualmente em parceria com o Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), alunos do curso de Educação Física, também da UFSCar, e funcionários da escola citada, tais como o professor e seus auxiliares.

Antes da realização dos testes, os alunos eram cadastrados e um número era atribuído a cada um, de tal forma que o aluno fosse facilmente identificado para cada teste.

Ao todo, 102 indivíduos participaram dos testes realizados dentro de um campo de futebol de 100 metros x 65 metros. A disposição dos testes e a utilização de uma câmera de vídeo digital permitiram que a execução dos testes ocorresse de forma adequada e em um pequeno espaço de tempo.

Dentre os participantes, 6 deles não completaram todas as provas devido a irregularidades não programadas, como a necessidade do aluno ir embora mais cedo ou não cumprir a sequência dos 6 testes propostos. Desta forma, a análise final que foi realizada pelo i-Sports contou com a participação integral de 96 alunos.

Nos tópicos a seguir, apresentamos uma análise descritiva do conjunto de dados, as etapas de criação dos indicadores e a aplicação da metodologia Z. Com os resultados obtidos, mostramos os indivíduos com melhores desempenhos nos testes e os que apresentaram os maiores valores nos indicadores criados.

3.1 Análise Descritiva

Por meio da Tabela 3.1 podemos observar a descrição geral dos indivíduos e também o desempenho nos testes.

Analisando os valores apresentados, verificamos que a média de idade dos indivíduos foi igual a 12,89 anos, sendo que o aluno mais velho possuía 22 anos e o mais novo apenas 7 anos. Para altura, observamos que a média foi 1,61 metros, com máximo igual a 1,98 metros e mínimo 1,22 metros. Tratando-se do peso, temos que a média foi de 52,2 kg, com maior e menor valor iguais a 92 kg e 22 kg, respectivamente.

Quanto aos testes aplicados, observamos que a pontuação média na variável CHUTE foi 3,32, com máximo igual a 24 e mínimo igual a 0. O tempo médio na variável DRIBLE foi 25,97 segundos, com maior valor igual a 40,16 segundos e menor valor igual a 14,78 segundos. Para a variável PASSE, verificamos que a pontuação média foi igual a 3,03, com máximo e mínimo iguais a 7 e 0.

Em relação à variável PR, notamos que sua média foi de 4,02 w/kg, com máximo igual a 7,14 w/kg e mínimo igual a 1,07 w/kg. A variável VO2MAX apresentou média de 56,17 ml/kg/min, com maior valor igual a 67,47 ml/kg/min e menor igual a 39,84 ml/kg/min. Por fim, verificamos que a variável CL20 apresentou média igual a 6,12 segundos, com máximo e mínimo iguais a 9,59 segundos e 4,81 segundos, respectivamente.

As Figuras 3.1 e 3.2 apresentam histogramas das variáveis relacionadas à descrição geral dos indivíduos e também do desempenho dos mesmos nos testes realizados. Nota-se destaque para a grande maioria de jogadores com pontuação entre 0 e 5 na variável CHUTE, que apresentou comportamento distinto das demais.

Tabela 3.1: Medidas descritivas para as variáveis coletadas.

Variável	Média	Máximo	Mínimo	DP
IDADE (anos)	12,89	22	7	2,74
ALTURA (metros)	1,61	1,98	1,22	0,16
PESO (kg)	52,20	92	22	15,92
CHUTE (pontuação)	3,32	24	0	4,57
DRIBLE (segundos)	25,97	40,16	14,78	5
PASSE (pontuação)	3,04	7	0	1,51
PR (w/kg)	4,02	7,14	1,07	1,46
VO2MAX (ml/kg/min)	56,17	67,43	39,84	7,12
CL20 (segundos)	6,12	9,59	4,81	0,70

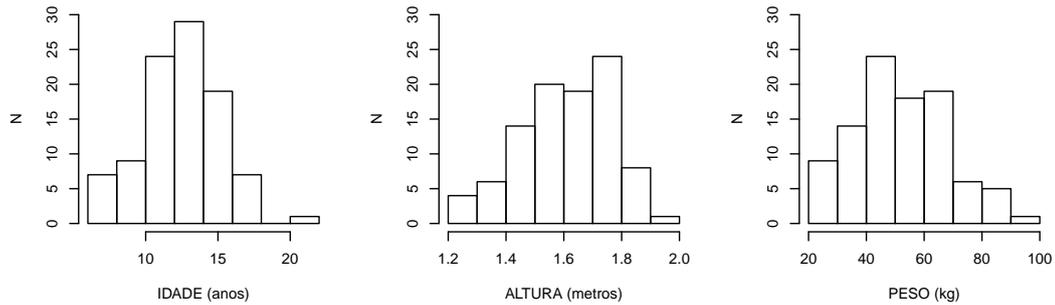


Figura 3.1: Histograma - descrição geral dos indivíduos.

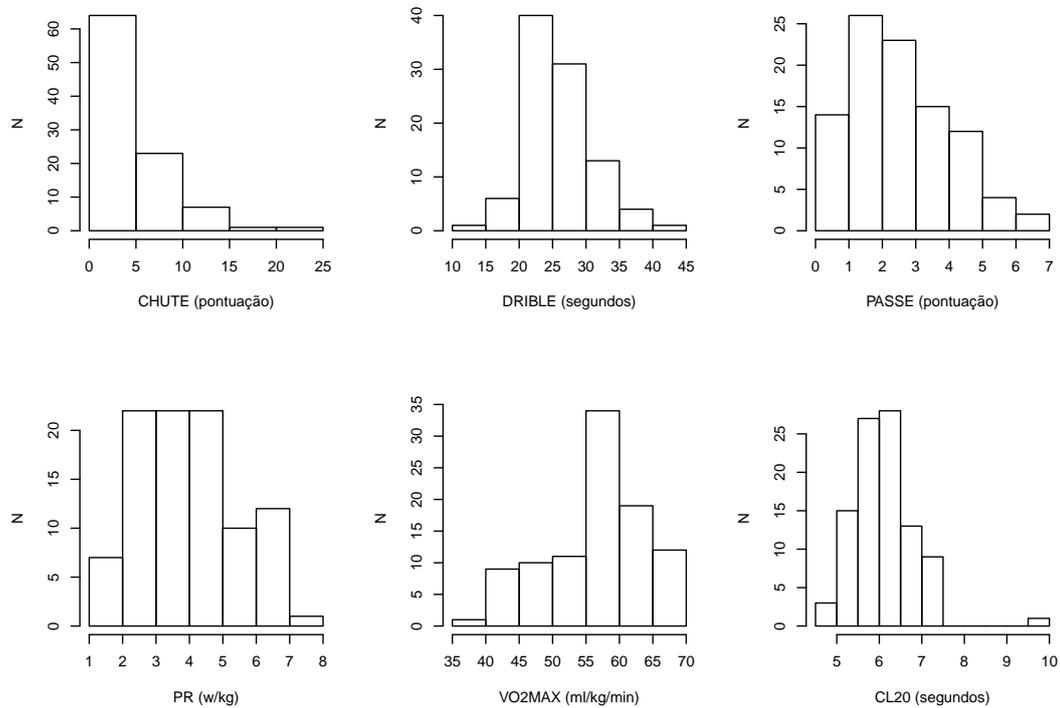


Figura 3.2: Histograma - desempenho nos testes.

3.2 Construção dos Indicadores

3.2.1 Via A.C.P.

Escore Físico

Observamos que existe associação entre as 3 variáveis, sendo que a maior associação está entre as variáveis PR e VO2MAX. Além disso, por meio das correlações apresentadas nota-se que essas duas medidas apresentaram associação negativa com a variável CL20.

Os resultados da A.C.P. aplicada às três variáveis estão dispostos nas Tabelas 3.3 e 3.4. Notamos que com apenas a primeira componente temos 77% da variabilidade total dos dados explicada e que a componente selecionada explica conjuntamente todas as variáveis, com um contraste entre PR e VO2MAX com CL20. Como essa componente apresentou alto valor para variabilidade explicada, optamos por utilizá-la na criação do escore físico.

Metodologicamente, ponderando a componente pela sua respectiva variabilidade explicada, chegamos à seguinte expressão para o indicador que explica 77% da variabilidade total das variáveis que quantificam as características físicas do indivíduo:

$$\begin{aligned} \text{Escore Físico} &= 0,77 \times e_1 \\ &= 0,4543 \times ((PR - 4,02)/1,46) + 0,4566 \times ((VO2MAX - 56,17)/7,12) \\ &\quad - 0,4312 \times ((CL20 - 6,12)/0,70). \end{aligned} \quad (3.1)$$

As quantidades que são subtraídas e divididas em cada variável na fórmula final são as respectivas médias e desvios padrão observadas na amostra. Tal fato foi necessário, pois a A.C.P. foi aplicada na matriz de correlação, que considera as variáveis padronizadas.

Por meio dos coeficientes apresentados na equação final do indicador, vemos que a variável CL20 apresenta peso negativo, ou seja, quanto maior o tempo do indivíduo nesse teste, menor o valor do indicador e vice-versa. Por outro lado, quanto maior o valor das variáveis PR e VO2MAX, maior o valor do indicador.

Assim, para calcularmos o valor máximo desse índice utilizamos os seguintes valores nas respectivas variáveis:

- $PR = 7,58$;
- $VO2MAX = 67,43$;
- $CL20 = 4,81$;

e para o valor mínimo:

- $PR = 1,04$;
- $VO2MAX = 39,84$;
- $CL20 = 9,59$.

Aplicando essas medidas diretamente na equação (3.1) teremos: $I_{max} = 2,50$ e $I_{min} =$

-4,09. Vale ressaltar que os valores considerados como máximo e mínimo foram observados na análise descritiva.

Tabela 3.2: Matriz de correlação observada para os testes físicos.

	PR	VO2MAX	CL20
PR	1	0,7196	-0,6361
VO2MAX	0,7196	1	-0,6122
CL20	-0,6361	-0,6122	1

Tabela 3.3: Autovalores e % de variabilidade explicada obtida por meio da A.C.P. para testes físicos.

Autovalores	2,31	0,41	0,28
% de variância explicada	0,77	0,14	0,09

Tabela 3.4: Autovetores obtidos por meio da A.C.P. para testes físicos.

e_1	e_2	e_3
0,59	-0,32	-0,74
0,58	-0,47	0,66
-0,56	-0,82	-0,09

Escore Técnico

Analisando a correlação entre as variáveis relacionadas às características técnicas dos indivíduos, dispostas na Tabela 3.5, notamos que as 3 variáveis apresentaram correlação baixa.

Por meio da A.C.P. aplicada aos dados, verificamos que as duas primeiras componentes explicam conjuntamente 75% da variabilidade dos dados (Tabela 3.6) e pela Tabela 3.7 e Figura 3.3, observamos que a primeira componente explica principalmente as variáveis PASSE e DRIBLE, com contraste desta última variável com CHUTE e PASSE, enquanto que a segunda componente explica essencialmente a variável CHUTE, sendo contrastada com a variável PASSE.

Caso utilizássemos apenas a primeira componente na composição do escore técnico, teríamos baixa explicação das três variáveis consideradas (42%). Sendo assim, optamos por utilizar as duas primeiras componentes para criação do índice, que explicam conjuntamente 75% da variabilidade total dos dados.

Ponderando a primeira e a segunda componente pela respectiva porcentagem de vari-

abilidade explicada, chegamos à seguinte fórmula para o escore técnico:

$$\begin{aligned} \text{Escore Técnico} &= 0,42 \times e_1 + 0,33 \times e_2 \\ &= 0,4365 \times ((CHUTE - 3,32)/4,57) - 0,2874 \times ((DRIBLE - 25,97)/5) \\ &\quad + 0,11161 \times ((PASSE - 3,04)/1,51). \end{aligned} \quad (3.2)$$

Analisando os coeficientes exibidos na equação final do escore técnico, verificamos que a variável DRIBLE apresenta peso negativo, indicando que quanto maior o tempo do indivíduo nesse teste, menor o valor do indicador e vice-versa. Para CHUTE e DRIBLE ocorre o contrário, ou seja, quanto maior a pontuação apresentada nesses testes, maior o valor do indicador ou o contrário.

Para calcularmos o valor máximo desse índice utilizamos os seguintes valores para as variáveis:

- $CHUTE = 24$;
- $DRIBLE = 14,78$;
- $PASSE = 7$;

e para o valor mínimo:

- $CHUTE = 0$;
- $DRIBLE = 40,16$;
- $PASSE = 0$.

Aplicando essas medidas na fórmula exibida na equação (3.2) temos: $I_{max} = 2,91$ e $I_{min} = -1,36$. Assim como feito no escore físico, os valores considerados como máximo e mínimo foram observados na análise descritiva.

Tabela 3.5: Matriz de correlação observada para os testes técnicos.

	CHUTE	DRIBLE	PASSE
CHUTE	1	-0,1154	0,0102
DRIBLE	-0,1154	1	-0,2270
PASSE	0,0102	-0,2270	1

Tabela 3.6: Autovalores e % de variabilidade explicada obtida por meio da A.C.P. para testes técnicos.

Autovalores	1,26	0,99	0,75
% de variância explicada	0,42	0,33	0,25

Tabela 3.7: Autovetores obtidos por meio da A.C.P. para testes técnicos.

e_1	e_2	e_3
0,34	0,89	-0,30
-0,70	0,02	-0,71
0,63	-0,45	-0,63

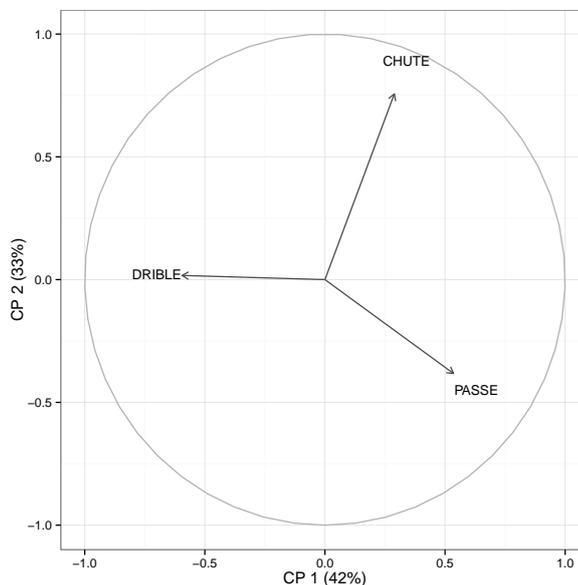


Figura 3.3: Gráfico das duas primeiras componentes - Escore técnico.

3.2.2 Via Análise Fatorial

Escore Geral

Por meio da Análise Fatorial reduzimos as 6 variáveis em apenas um indicador, dado pela combinação dos fatores comuns escolhidos, que representa o desempenho geral nos testes.

Assim como na Análise de Componentes Principais, a Análise Fatorial leva em conta apenas observações completas, invalidando indivíduos com pelo menos uma observação faltante em determinada variável. Entretanto, os indivíduos que não completaram determinados testes já foram retirados da análise antecipadamente.

Com a finalidade de determinar a adequação da Análise Fatorial, examinamos a matriz de correlação (Tabela 3.8) utilizando o teste Barlett de esfericidade. Tal teste verifica se a matriz de correlação apresenta correlações significativas entre pelo menos alguma das variáveis. Analisando o p-valor apresentado na Tabela 3.9, podemos observar presença significativa de correlações entre pelo menos duas variáveis.

Outra medida para quantificar o grau de intercorrelações e adequação da Análise Fatorial aos dados é a medida de adequação da amostra (MSA). Esse índice varia de

0 a 1, alcançando 1 quando cada variável é prevista sem erro pelas outras variáveis. Avaliando os valores apresentados na Tabela 3.10 e baseando-se nas orientações sobre a interpretação da MSA ([43]), notamos que tanto para o MSA Total quanto para as variáveis individualmente, temos valores aceitáveis.

Dessa forma, seguimos com a Análise Fatorial sabendo que a aplicação desta análise ao conjunto de dados é adequada.

Tabela 3.8: Matriz de correlação dos testes.

	PR	VO2MAX	CL20	CHUTE	DRIBLE	PASSE
PR	1	0,7196	-0,6361	0,1009	-0,6608	0,2662
VO2MAX	0,7196	1	-0,6122	0,1265	-0,6047	0,1830
CL20	-0,6361	-0,6122	1	-0,0967	0,6826	-0,2226
CHUTE	0,1009	0,1265	-0,0967	1	-0,1154	0,0102
DRIBLE	-0,6608	-0,6047	0,6826	-0,1154	1	-0,2270
PASSE	0,2662	0,1830	-0,2226	0,0102	-0,2270	1

Tabela 3.9: Resultado do teste Bartlett de esfericidade.

G.L	Qui-Quadrado	P-Valor
15	208,1529	0,000

Tabela 3.10: Medida de adequação da amostra - MSA.

Total	PR	VO2MAX	CL20	CHUTE	DRIBLE	PASSE
0,8261	0,8025	0,8179	0,8421	0,8890	0,8345	0,8863

A Tabela 3.11 nos fornece os autovalores, proporção de variabilidade explicada e acumulada de cada um dos fatores presentes na análise. Considerando a proporção da variabilidade explicada pelos fatores como critério de seleção do número de fatores, temos que, ao utilizar os três primeiros fatores, conseguimos explicar aproximadamente 83% da variabilidade total dos dados. Tal critério de escolha pode ser melhor visualizado por meio da Figura 3.4, sendo que os três primeiros fatores explicam, conjuntamente, acima de 80% da variabilidade total dos dados.

Tabela 3.11: Autovalor, proporção de variabilidade explicada e acumulada.

Fator	Autovalor	Proporção Explicada	Proporção Acumulada
1	3,08	0,5133	0,5133
2	0,99	0,1658	0,6791
3	0,89	0,1482	0,8273
4	0,45	0,0748	0,9021
5	0,32	0,0538	0,9558
6	0,27	0,0442	1,0000

Dado que o número de fatores a serem utilizados foi escolhido, tornou-se possível rotacionar os eixos e calcular as cargas fatoriais referentes a cada um dos fatores. Tais valores, dispostos na Tabela 3.12, foram obtidos a partir da rotação VARIMAX.

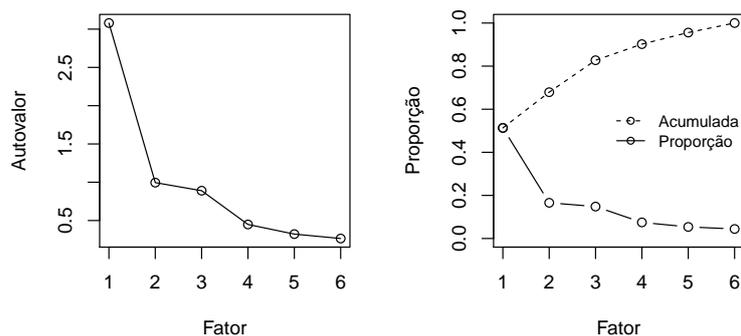


Figura 3.4: Gráfico de cotovelo e proporção de variância explicada.

Observando as variáveis e os fatores conjuntamente, temos que o primeiro fator explica principalmente a potência relativa (PR), o volume de oxigênio máximo (VO2MAX), a velocidade cíclica 20 metros (CL20) e drible (DRIBLE), sendo que há um contraste dos dois primeiros testes com os dois últimos citados. O segundo fator explica exclusivamente o passe (PASSE) e o terceiro exclusivamente o chute (CHUTE).

A Tabela 3.13 apresenta a variabilidade explicada por cada fator rotacionado. Dessa forma, o primeiro fator explica 48,98%, o segundo 17,01% e o terceiro 16,74%.

Temos na Tabela 3.14 as comunalidades (variabilidade explicada) de cada variável, medida utilizada para verificar se as variáveis estão sendo bem explicadas pelos fatores escolhidos. Por meio dessas medidas, verificamos que os três fatores explicam, conjuntamente, aproximadamente 77,33% de PR, 73,56% de VO2MAX, 72,46% de CL20, 100% do CHUTE, 73,28% do DRIBLE e 100% do PASSE.

Tabela 3.12: Cargas fatoriais rotacionadas.

	Fator 1	Fator 2	Fator 3
PR	0,8656	0,1508	0,0353
VO2MAX	0,8539	0,0353	0,0718
CL20	-0,8451	-0,0983	-0,0274
CHUTE	0,0737	0,0001	0,9971
DRIBLE	-0,8476	-0,1070	-0,0540
PASSE	0,1502	0,9876	-0,0003

Tabela 3.13: Variância explicada por cada fator rotacionado.

Fator 1	Fator 2	Fator 3
0,4898	0,1701	0,1674

A Figura 3.5 exhibe o cruzamento entre os fatores 1 e 2 rotacionados. Ao observá-la, é possível verificar que as variáveis VO2MAX e PR são muito correlacionadas entre si e inversamente correlacionadas com as variáveis DRIBLE e CL20. Nota-se ainda que

Tabela 3.14: Comunalidade final estimada.

PR	VO2MAX	CL20	CHUTE	DRIBLE	PASSE
0,7733	0,7356	0,7246	0,9996	0,7328	0,9980

essas variáveis são quase que independentes do PASSE, visto que formam um ângulo de aproximadamente 90° .

Comportamento semelhante pode ser observado na Figura 3.6, sendo que, nesse caso, as variáveis PR, VO2MAX, DRIBLE e CL20 são quase que independentes do CHUTE, apresentando também um ângulo de aproximadamente 90° .

Para as variáveis CHUTE e PASSE, observamos na Figura 3.7 que o ângulo entre as variáveis é exatamente igual a 90° , indicando que são variáveis independentes.

Por meio desses resultados, podemos dizer que com aumento de VO2MAX e PR temos diminuição em CL20 e DRIBLE, e vice-versa. Resultado esperado para esses testes dado que com maior potência e desempenho anaeróbico, o atleta consegue tempos mais baixos nos testes que envolvem habilidade e físico. Além disso, o desempenho no CHUTE e no PASSE são independentes entre si e entre as demais variáveis.

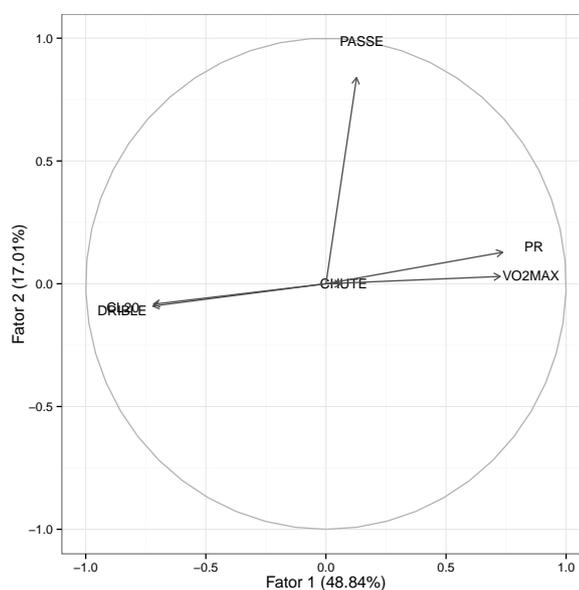


Figura 3.5: Cruzamento entre primeiro e segundo fator.

Para calcularmos os escores fatoriais a serem utilizados na análise, utilizamos os coeficientes presentes na Tabela 3.15.

Vale ressaltar que as variáveis devem estar padronizadas, ou seja, apresentarem média

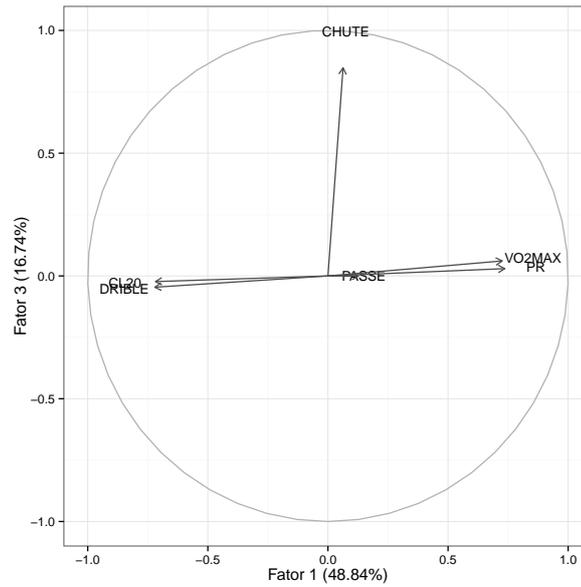


Figura 3.6: Cruzamento entre primeiro e terceiro fator.

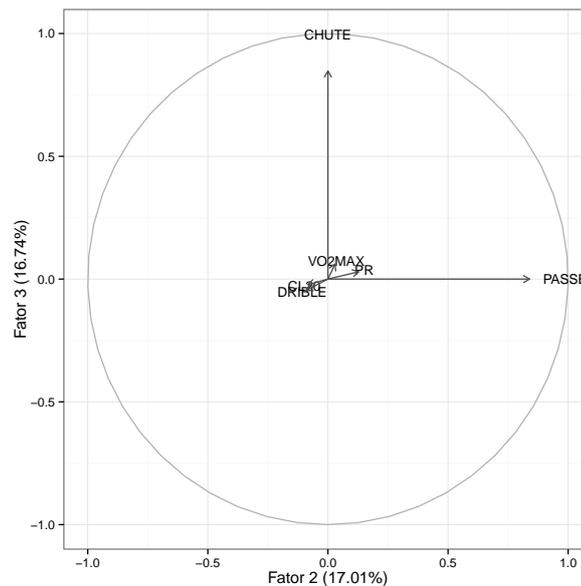


Figura 3.7: Cruzamento entre segundo e terceiro fator.

Tabela 3.15: Coeficientes dos escores fatoriais.

	F1	F2	F3
PR	0,2959	0,0084	-0,0340
VO2MAX	0,3088	-0,1115	0,0013
CL20	-0,2981	0,0440	0,0415
CHUTE	-0,0570	0,0111	1,0060
DRIBLE	-0,2952	0,0346	0,0145
PASSE	-0,1179	1,0234	0,0108

0 e variância 1. Tal transformação pode ser feita da seguinte maneira:

$$X^* = \frac{X - \text{Média}(X)}{\text{Desvio Padrão}(X)}. \quad (3.3)$$

Sendo assim, os valores finais são obtidos aplicando-se as seguintes fórmulas para cada um dos fatores:

$$\begin{aligned} F1 = & 0,2959 \times PR^* + 0,3088 \times VO2MAX^* - 0,2981 \times CL20^* - 0,0570 \times CHUTE^* \\ & - 0,2952 \times DRIBLE^* - 0,1179 \times PASSE^*, \end{aligned} \quad (3.4)$$

$$\begin{aligned} F2 = & 0,0084 \times PR^* - 0,1115 \times VO2MAX^* + 0,0440 \times CL20^* + 0,0111 \times CHUTE^* \\ & 0,0346 \times DRIBLE^* + 1,0234 \times PASSE^*, \end{aligned} \quad (3.5)$$

$$\begin{aligned} F3 = & -0,0340 \times PR^* + 0,0013 \times VO2MAX^* + 0,0415 \times CL20^* + 1,0060 \times CHUTE^* \\ & 0,0145 \times DRIBLE^* + 0,0108 \times PASSE^*. \end{aligned} \quad (3.6)$$

Dada a proporção de variabilidade explicada por cada fator e as fórmulas de cada escore fatorial, calculamos o valor do indicador por meio da ponderação dos escores fatoriais de cada um dos fatores pela proporção de variabilidade explicada pelo respectivo fator, após a rotação:

$$\begin{aligned} \text{Escore Geral} = & 0,4898 \times F1 + 0,1701 \times F2 + 0,1674 \times F3 \\ = & 0,1406 \times PR^* + 0,1325 \times VO2MAX^* - 0,1316 \times CL20^* + 0,1424 \times CHUTE^* \\ & - 0,1363 \times DRIBLE^* + 0,1181 \times PASSE^*. \end{aligned} \quad (3.7)$$

Assim como observado no escore físico e técnico, novamente as variáveis CL20 e DRIBLE apresentaram coeficiente negativo. Dessa forma, quanto maior o tempo para percorrer os 20 metros e completar o teste de drible, menor o valor do escore geral e vice-versa. Esse comportamento era esperado, pois um jogador com capacidade física e técnica apurada tende a apresentar melhor será o desempenho em drible e velocidade.

Para o escore geral utilizamos os seguintes valores para calcular o máximo:

- $PR = 7,14$;
- $VO2MAX = 67,43$;
- $CL20 = 4,81$;
- $CHUTE = 24$;
- $DRIBLE = 14,78$;

- $PASSE = 7$;

e os seguintes para o mínimo:

- $PR = 1,07$;
- $VO2MAX = 39,84$;
- $CL20 = 9,59$;
- $CHUTE = 0$;
- $DRIBLE = 40,16$;
- $PASSE = 0$.

Aplicando os valores considerados para o máximo e mínimo na equação (3.7) temos: $I_{max} = 2,02$ e $I_{min} = -1,97$. Assim como feito nos dois indicadores da A.C.P., os valores considerados como máximo e mínimo foram observados na análise descritiva.

3.2.3 Via Cópula

Para exemplificar o conceito de consistência, primeiramente apresentamos o ajuste de uma cópula Normal bivariada. Na Figura 3.8 podemos verificar que as variáveis PR e VO2MAX foram as mais associadas e, por esse motivo, foram escolhidas para representarem a aplicação da metodologia bivariada.

A distribuição da densidade da distribuição Normal (padrão) bivariada é dada por:

$$h_r(x, y) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left\{-\frac{x^2 + y^2 - 2rxy}{2(1-r^2)}\right\}, \quad (3.8)$$

com função de distribuição não explícita:

$$H_r(x, y) = \int_{-\infty}^x \int_{-\infty}^y h_r(t, s) dt ds, \quad (3.9)$$

e marginais também não explícitas, tradicionalmente denotadas por:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp^{-\frac{t^2}{2}} dt. \quad (3.10)$$

Ainda que H_r não seja explícita, a cópula associada a H_r é dada por:

$$C_r(u, v) = H_r\{\Phi^{-1}(u), \Phi^{-1}(v)\}, r \in [-1, 1], \quad (3.11)$$

isto é,

$$C_r(u, v) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{x^2 + y^2 - 2rxy}{2(1-r^2)}\right\} dy dx. \quad (3.12)$$

A sua função densidade de probabilidade é dada por:

$$c_r(u, v) = \frac{h_r\{\Phi^{-1}(u), \Phi^{-1}(v)\}}{\varphi\{\Phi^{-1}(u)\}\varphi\{\Phi^{-1}(v)\}}. \quad (3.13)$$

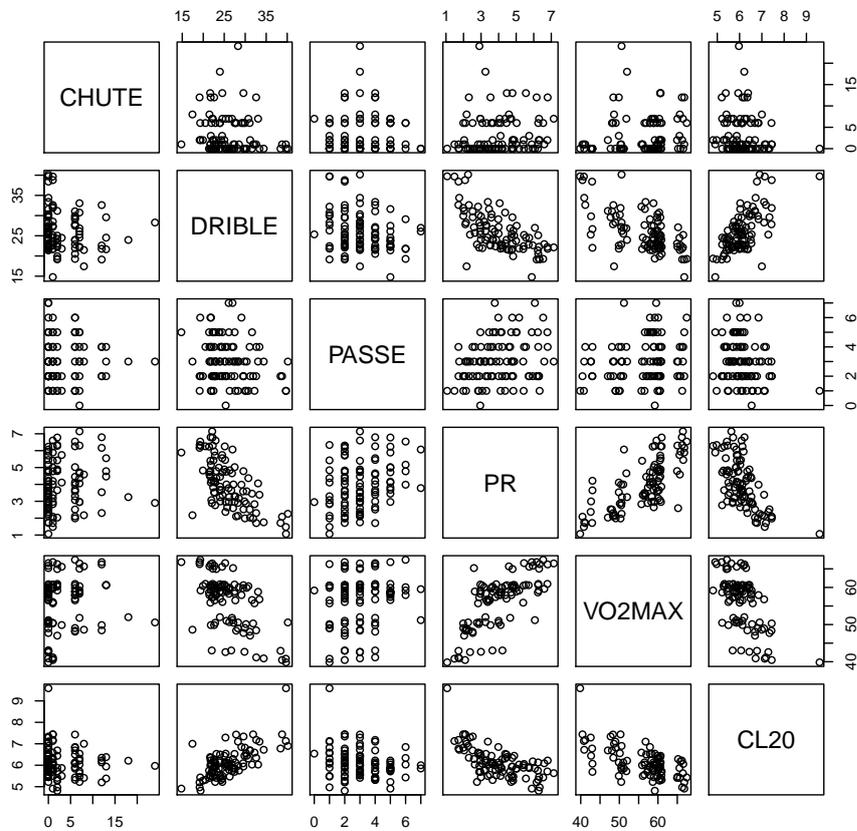


Figura 3.8: Cruzamento entre todas as variáveis da análise.

As Figuras 3.9 e 3.10 apresentam uma visualização gráfica do comportamento da densidade, curvas de níveis da densidade, probabilidade acumulada e curvas de níveis da probabilidade acumulada para a distribuição Normal bivariada e cópula Normal, respectivamente. Em todos os casos consideramos coeficiente de correlação igual $1/2$, ou seja, $r = 1/2$.

Por meio da Figura 3.11 observamos os pares de observações para as variáveis em questão e também os pares normalizados pelo rank.

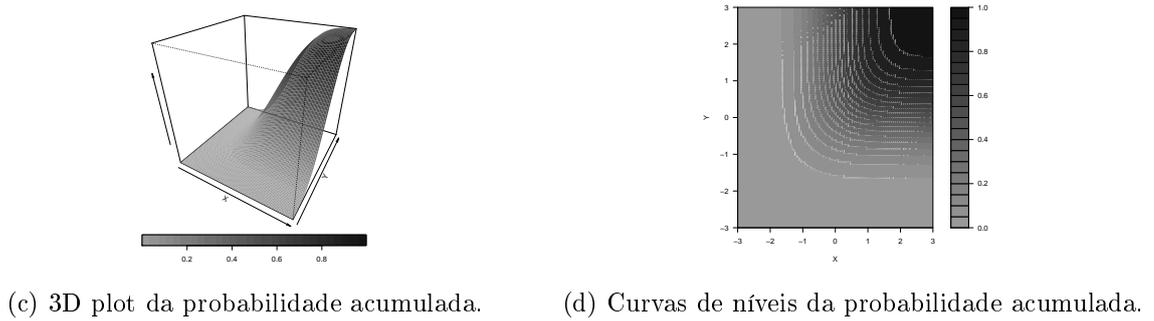
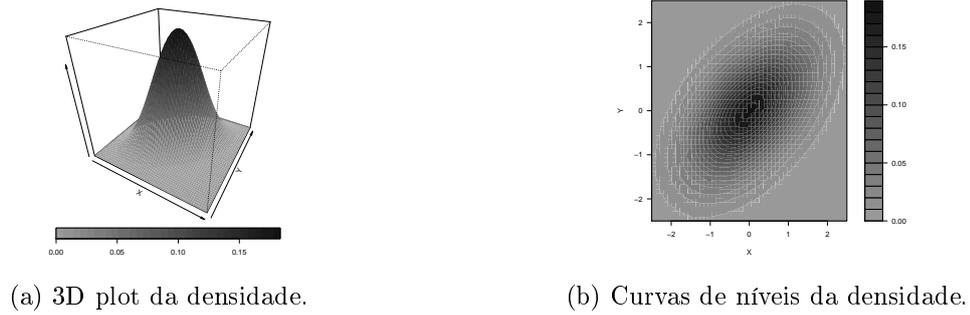


Figura 3.9: Normal bivariada ($r = \frac{1}{2}$).

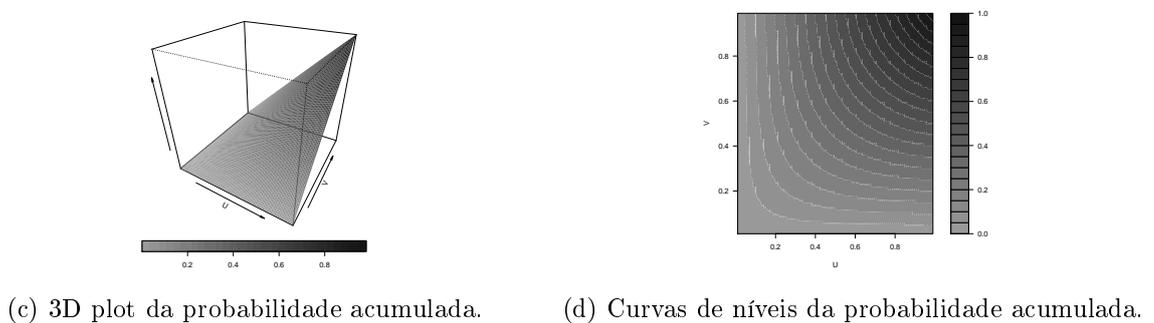
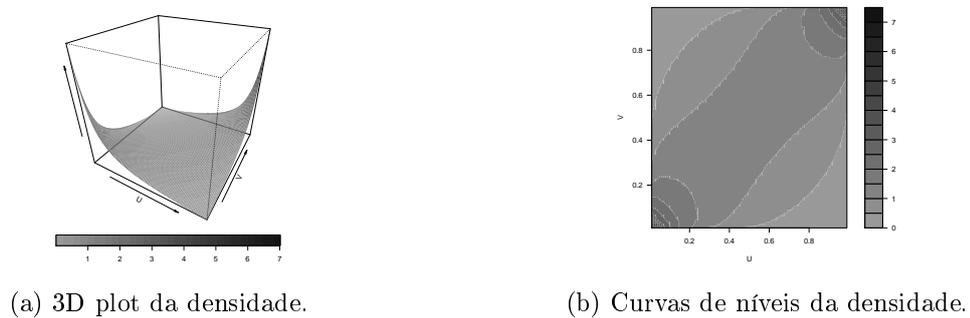


Figura 3.10: Cópula Normal ($r = \frac{1}{2}$).

A Figura 3.12 apresenta a comparação entre os dados utilizados e possíveis cópulas a serem ajustadas. Tais valores foram gerados para as cópulas de Cauchy, t, Normal, Gumbel, Clayton, Frank e Plackett utilizando o *tau de Kendall* obtido a partir de PR e

VO2MAX ($\tau = 0,5609$). Por meio das cópulas geradas podemos observar que a cópula Normal, Cauchy e Plackett são possíveis candidatas a se ajustarem corretamente aos dados, pois apresentam comportamento semelhantes aos dados analisados. Optamos por utilizar a cópula Normal pelo fato dessa família de cópula já ter sido apresentada no trabalho e também por ser uma das mais adequadas às 6 variáveis conjuntamente (Tabela 3.20).

As Tabelas 3.16 e 3.17 são referentes ao ajuste da cópula Normal. A primeira fornece a estimativa do parâmetro θ da cópula, seu erro padrão, a estatística teste e o p-valor para a significância do parâmetro, enquanto que a segunda apresenta o teste de qualidade de ajuste para a cópula Normal. Por meio desses valores podemos observar que o valor de θ estimado é igual a 0,7813, sendo significativo ao nível de 1%, e também que a cópula escolhida é adequada aos dados, uma vez que o teste de qualidade de ajuste indicou ajuste satisfatório.

Confirmamos a adequação da cópula por meio da comparação entre a cópula empírica e a cópula Normal pelo gráfico QQplot (Figura 3.13). Nota-se claramente a aproximação entre ambas.

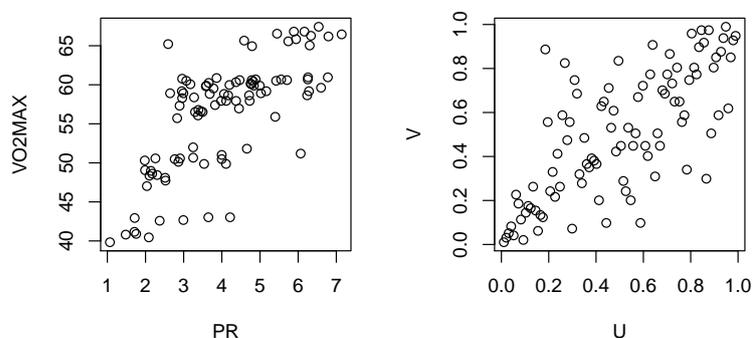


Figura 3.11: Dados originais (esquerda) e transformados pelo rank (direita).

Tabela 3.16: Ajuste da cópula Normal.

$\hat{\theta}$	Erro padrão	Estatística teste	P-valor
0,7813	0,0317	24,63	<2e-16

Tabela 3.17: Teste de qualidade de ajuste.

P-valor
0,3741

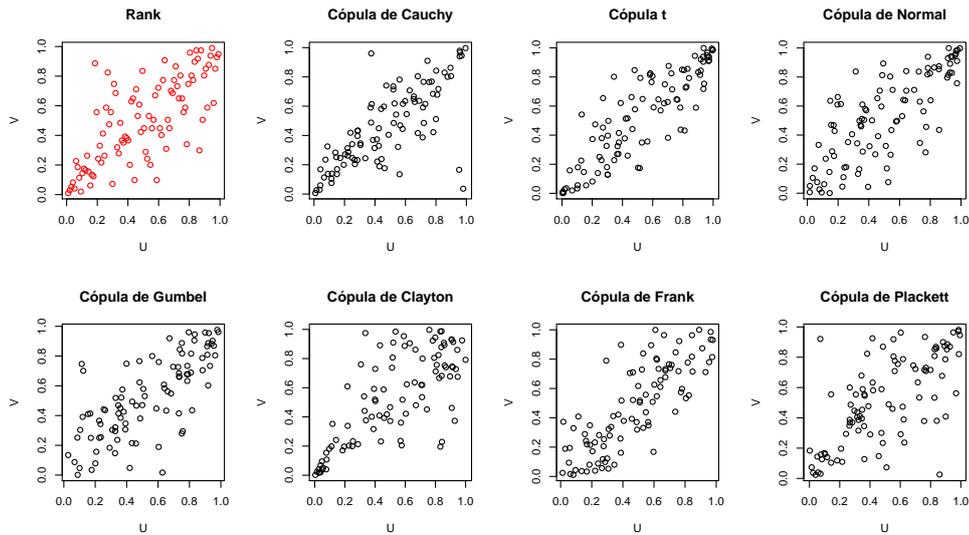


Figura 3.12: Comparação entre possíveis cópulas simuladas.

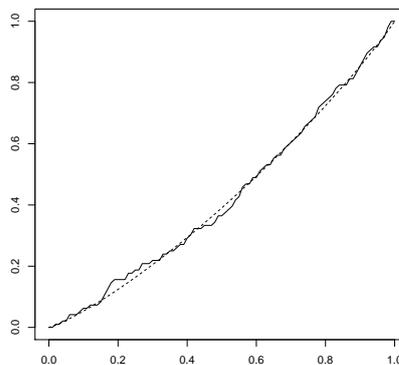


Figura 3.13: QQplot para comparação entre cópula empírica e cópula Normal ($\theta = 0,7813$).

A Tabela 3.18 fornece os resultados dos indivíduos com melhores desempenhos em PR e VO2MAX conjuntamente. Tais valores são resultados da cópula Normal bivariada ajustada ao dados.

Tabela 3.18: Indivíduos mais consistentes.

NOME	IDADE	ALTURA	PESO	PR	VO2MAX	CONSISTÊNCIA
Joaquim	16	1,72	59	6,5381	67,4312	94,58
João Lucas	15	1,75	70	7,1410	66,4552	94,58
Calebe	15	1,65	50	6,7935	66,1890	92,23
Francisco	17	1,70	68	6,3352	66,2777	90,68

Para o ajuste da cópula multivariada consideramos as seis variáveis que representam os testes aplicados nos atletas: CHUTE, DRIBLE, PASSE, PR, VO2MAX e CL20.

Por utilizarmos as pseudo-observações calculadas por meio dos ranks, não foi necessário fazer transformações nos dados. Nesse caso, utilizamos o rank médio para os empates.

Caso usássemos o máximo, estaríamos favorecendo os indivíduos em relação aos com rank menores que estes, e o mínimo desfavoreceríamos em relação à seus superiores. Quanto à atribuição aleatória para empates, perderíamos a comparação entre os indivíduos.

Como as variáveis DRIBLE e CL20 são medidas de tempo, o melhor indivíduo é o indivíduo que apresenta menor valor. Dessa forma, para invertermos a ordem de classificação dos indivíduos de maneira que o melhor apresentasse o maior valor, utilizamos o escore reverso para duas variáveis:

- DRIBLE = - DRIBLE,
- CL20 = - CL20.

Sabendo que as observações foram feitas em indivíduos diferentes, descartamos o teste de aleatoriedade.

A Tabela 3.19 apresenta o teste de independência multivariada. Por meio desse teste observamos que a hipótese de independência foi rejeitada, indicando que existe associação entre as variáveis.

Analisando a Tabela 3.20, que representa os testes de qualidade de ajuste para as famílias de cópulas Normal, cópula t, cópula de Cauchy, cópula de Clayton, cópula de Gumbel-Hougaard e cópula de Frank, notamos que por apresentar maior p-valor, a cópula Normal é a mais adequada aos dados.

As estimativas para a cópula Normal adequada estão dispostas na Tabela 3.21.

Tabela 3.19: Teste de independência multivariada.

Estadística teste	P-valor
0,3038	0,0005

Tabela 3.20: Família de cópulas.

Família	Estadística teste	P-valor
Gumbel	0,4206	0,0654
Clayton	0,2546	0,2033
Frank	0,3561	0,0934
Normal	0,1490	0,3701
t	0,2803	0,1394
Plackett	-	-

Tabela 3.21: Ajuste da cópula Normal.

$\hat{\theta}$	Erro padrão	Estadística teste	P-valor
0,4330	0,0404	10,72	<2e-16

3.3 Indivíduos com Melhores Desempenho

3.3.1 Método Z

As Tabelas de 3.22 a 3.27 fornecem os resultados obtidos por meio da metodologia Z. Em todos os casos são apresentados os 3 jogadores que mais se destacaram individualmente em cada teste.

Destaca-se que na variável CHUTE dois indivíduos apresentaram desempenho bem acima do esperado, sendo que um deles apresentou valor de Z maior do que 4.

Como visto anteriormente, essa metodologia leva em conta apenas uma variável. Dessa forma, é essencial a avaliação multivariada feita por meio dos quatro indicadores criados.

Tabela 3.22: Indivíduos que mais se destacaram no CHUTE segundo a metodologia Z.

Nome	IDADE	ALTURA	PESO	CHUTE	Z
Théo	12	1,52	40	24	4,53
Pietro	14	1,66	46	18	3,21
Vinicius	14	1,83	72	13	2,12

Tabela 3.23: Indivíduos que mais se destacaram no DRIBLE segundo a metodologia Z.

Nome	IDADE	ALTURA	PESO	DRIBLE	Z
Davi Lucca	15	1,77	67	14,78	2,24
Juan	11	1,50	58	17,43	1,71
Igor	17	1,74	70	19,13	1,37

Tabela 3.24: Indivíduos que mais se destacaram no PASSE segundo a metodologia Z.

Nome	IDADE	ALTURA	PESO	PASSE	Z
Ryan	15	1,69	81	7	2,61
Luan	13	1,68	56	7	2,61
Joaquim	16	1,72	59	6	1,95

Tabela 3.25: Indivíduos que mais se destacaram em PR segundo a metodologia Z.

Nome	IDADE	ALTURA	PESO	PR	Z
João Lucas	15	1,75	70	7,14	2,14
Calebe	15	1,65	50	6,79	1,90
Bruno	16	1,65	58	6,78	1,89

Tabela 3.26: Indivíduos que mais se destacaram em VO2MAX segundo a metodologia Z.

Nome	IDADE	ALTURA	PESO	VO2MAX	Z
Joaquim	16	1,72	59	67,43	1,58
Igor	17	1,74	70	66,81	1,50
Davi Lucca	15	1,77	67	66,81	1,50

Tabela 3.27: Indivíduos que mais se destacaram em CL20 segundo a metodologia Z.

Nome	IDADE	ALTURA	PESO	CL20	Z
Igor	16	1,69	75	4,81	1,88
Davi Lucca	15	1,77	67	4,91	1,74
Francisco	17	1,70	68	4,97	1,65

3.4 Jogadores com Melhores Avaliações

Apresentamos nas Tabelas de 3.28 a 3.31 os indivíduos que mais se destacaram em cada indicador criado.

Analisando os valores para o escore físico (Tabela 3.28) notamos que o atleta que obteve o melhor desempenho apresentou valor igual a 93,58, enquanto que para o escore técnico (Tabela 3.29) o maior valor foi igual a 73,20. Isso indica que, no geral, o desempenho técnico foi abaixo do desempenho físico.

Com relação à avaliação geral, que considera conjuntamente as variáveis CHUTE, DRIBLE, PASSE, PR, VO2MAX e CL20, vemos na Tabela 3.30 que o maior valor foi igual a 76,50.

Por fim, exibimos na Tabela 3.31 os resultados dos indivíduos mais consistentes. Pelos valores apresentados, observamos que a consistência máxima foi igual a 63,06, obtida pelo mesmo atleta que apresentou maior avaliação geral.

Tabela 3.28: Melhores jogadores fisicamente.

Ranking	Escore físico	Nome	IDADE	ALTURA	PESO	PR	VO2MAX	CL20
1	93,58	Francisco	17	1,70	68	6,34	66,28	4,97
2	87,64	Fernando	15	1,74	69	6,30	65,04	5,46
3	81,66	Davi Luiz	13	1,56	46	5,74	65,57	5,87

Tabela 3.29: Jogadores com maior habilidade técnica.

Ranking	Escore técnico	Nome	IDADE	ALTURA	PESO	CHUTE	DRIBLE	PASSE
1	73,20	Pietro	14	1,66	46	18	23,97	3
2	72,39	Théo	12	1,52	40	24	28,28	3
3	63,08	Antonio	17	1,90	80	13	21,65	3

Tabela 3.30: Jogadores com melhores avaliações - Escore geral.

Ranking	Escore geral	Nome	CHUTE	DRIBLE	PASSE	PR	VO2MAX	CL20
1	76,50	Joaquim	6	19,28	6	6,54	67,43	5,43
2	74,31	Davi Lucca	1	14,78	5	5,89	66,81	4,91
3	73,36	Igor	12	19,13	2	6,16	66,81	5,20

Tabela 3.31: Indivíduos mais consistentes.

Ranking	Consistência	Nome	CHUTE	DRIBLE	PASSE	PR	VO2MAX	CL20
1	63,06	Joaquim	6	19,28	6	6,54	67,43	5,43
2	50,03	Erick	6	20,52	4	6,26	60,95	5,32
3	45,32	Bernardo	8	21,48	4	4,59	65,66	5,41

Capítulo 4

Sistema i-Sports

Nesta seção apresentamos o conjunto de *softwares* utilizados na construção do sistema i-Sports e o funcionamento do sistema com a utilização dos indicadores desenvolvidos nesse trabalho.

4.1 Abordagem Web

Nos últimos anos a internet apresentou grande crescimento no número de usuários (676,3% de 2000 a 2014), sendo que no fim de 2013 cerca de 40% da população mundial possuía acesso à mesma [44]. A partir de uma simples conexão o indivíduo pode visitar páginas que estejam armazenadas em servidores em qualquer lugar do mundo e visualizar diversos tipos de conteúdo. Devido à essa facilidade, amplitude e praticidade existente, este trabalho propõe um sistema computacional inovador, denominado i-Sports, como uma ferramenta estatística online direcionada ao auxílio da prática do futebol.

Cada jogador cadastrado no sistema pode ter seu desempenho globalmente comparado a qualquer outro também cadastrado. Neste sentido, o sistema é capaz de realizar um acompanhamento contínuo dos aspectos físicos e técnicos, bem como identificar indivíduos que possuem habilidades acima da média, isto é, atletas que estatisticamente se destaquem da população de praticantes comuns à sua modalidade.

Dado que todo o sistema está estruturado em “nuvem”, não existe necessidade de instalação de *softwares* nos locais de utilização do sistema para que o i-Sports possa ser utilizado, basta uma conexão à internet. Dessa forma, uma das principais características desse tipo de abordagem é que o i-Sports pode ser utilizado em larga escala e acessado por diferentes localidades.

Sua construção é feita a partir de *softwares* essencialmente livres e gratuitos que podem ser facilmente encontrados na web. Basicamente utilizamos um servidor online com sistema operacional Linux Ubuntu Server 9.10 e o *software* R como ambiente computacional para realização de análises estatísticas e construção de gráficos. As web linguagens HTML e PHP foram consideradas para a formulação das páginas e conexões com os banco de dados. Para gerenciamento dos dados empregamos o MySQL 5.0, que usa a linguagem SQL como interface, e o phpMyAdmin versão 2.7.0-pl2, que permite a administração do MySQL pela internet. Por meio deste último pode-se criar e remover bases de dados; criar, remover e alterar tabelas; inserir, remover e editar campos, executar códigos SQL e manipular campos chaves.

A Figura 4.1 exibe a estrutura geral do sistema, desde o acesso feito pelos usuários até a estrutura do banco de dados.

A conexão entre o usuário e o sistema é dada pelos seguintes passos:

Passo 1: O usuário acessa o site e fornece o login e senha nos campos disponíveis;

Passo 2: O sistema recebe as informações e faz uma checagem com as informações presentes no banco de dados;

Passo 3: Se o login e a senha forem autenticados, o sistema permite o acesso. Caso contrário, uma mensagem é retornada ao usuário informando que as informações não são válidas;

Passo 4: Feito o acesso, o sistema checa se o usuário é aluno ou professor. Se for aluno, exibe apenas as informações do próprio indivíduo, e se for professor redireciona o usuário à uma página com as opções disponíveis.

Uma vez logado no sistema o professor pode inserir novos alunos, avaliar os resultados de cada aluno individualmente ou globalmente e comparar turmas específicas. As opções disponíveis para cada tipo de utilizador são exibidas na Figura 4.2.

4.2 Implementação e Avaliação

Feita a coleta de dados, o administrador da escola pôde acessar o sistema usando logins e senhas fornecidos pelo i-Sports, como indicado na Figura 4.3, área a direita.

Para o leitor acessar e visualizar todo o conteúdo deste relatório, basta entrar no site e acessar com o login: ESCOLA e senha: ABC999.

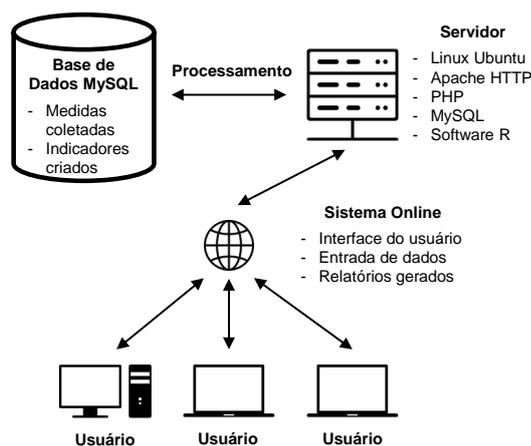


Figura 4.1: Estrutura geral do sistema, bem como as linguagens utilizadas.

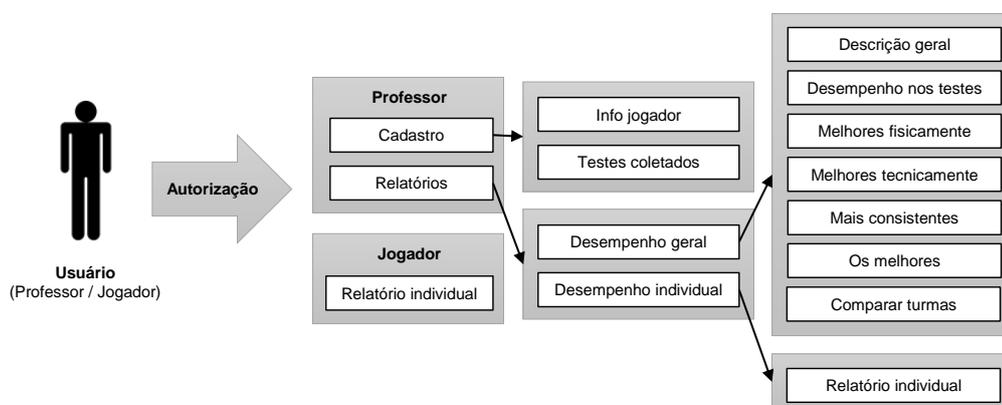


Figura 4.2: Acesso disponível para cada tipo de usuário.

Os dados coletados serviram de aplicação para a metodologia desenvolvida no trabalho, permitindo avaliarmos a viabilidade e efetividade do sistema. Eles foram inseridos no sistema após o login do administrador, que preencheu a ficha de cadastro de cada aluno individualmente, exibida na Figura 4.4.

Após a inserção de cada indivíduo no banco, o usuário pode avaliá-los a partir de diversos relatórios dinâmicos disponíveis em um menu do sistema (Figura 4.5). Vale ressaltar que o i-Sports realiza o cálculo final das variáveis utilizadas como, por exemplo, o cálculo da potência relativa (PR) que é função dos tempos de realização do teste de potência anaeróbia (R.A.S.T).

Os relatórios disponíveis exibem as informações globais e individuais dos atletas, sendo possível avaliar as turmas separadamente ou então fazer uma avaliação geral com todos os indivíduos conjuntamente. As análises são feitas a partir do *software* R e exibidas no sistema.

A seguir apresentamos a descrição de cada teste disponível:

- **Descrição geral dos alunos (Figura 4.6):** expõe uma análise geral das características básicas dos indivíduos, tais como idade, altura e peso;
- **Desempenho geral dos testes (Figura 4.7):** apresenta as informações sobre o rendimento dos indivíduos, de uma forma global, em todos os 6 testes aplicados;
- **Melhores jogadores fisicamente (Figura 4.8):** exibe as informações dos indivíduos que mais se destacaram fisicamente. Inicialmente, o relatório exibe os 5 melhores, mas existe a possibilidade da exibição do ranking de todos os alunos cadastrados no i-Sports;
- **Jogadores com maior habilidade técnica (Figura 4.9):** mostra as informações dos indivíduos que mais se destacaram tecnicamente. Assim como no relatório para os melhores jogadores fisicamente, este relatório permite visualizar tanto os 5 melhores tecnicamente quanto o ranking de todos os indivíduos;
- **Jogadores com melhores avaliações (Figura 4.10):** exibe as informações dos indivíduos que mais se destacaram globalmente, ou seja, os jogadores com maior aptidão ao futebol de campo. É possível exibir tanto os 5 melhores quanto o ranking de todos os indivíduos cadastrados no i-Sports;
- **Jogadores mais consistentes (Figura 4.11):** apresenta as informações dos indivíduos mais consistentes, ou seja, os jogadores mais completos considerando os testes aplicados. Novamente é possível visualizar os 5 melhores e o ranking de todos os alunos cadastrados sistema;
- **Desempenho individual (Figuras 4.12 e 4.13):** expõe um relatório individual de cada indivíduo, o qual irá exibir todas as análises estatísticas de desempenho do mesmo;
- **Comparar turmas:** compara os indicadores das turmas e exibe algumas medidas descritivas dos testes aplicados em cada turma.

Como observado na Figura 4.2, o acesso dos alunos é redirecionado apenas para o relatório de desempenho individual.



Figura 4.3: Área de login da escola cadastrada.

i-sports **Cadastro de novos alunos**

Características Gerais:

Nome: Idade (anos):

Tempo de permanência na escola (anos): Altura (metros): Peso (kg):

Teste de chute após passe [?]:

Pontos na 1ª tentativa: Pontos na 2ª tentativa: Pontos na 3ª tentativa: Pontos na 4ª tentativa: Pontos na 5ª tentativa:

Teste de drible de 5 cone [?]:

1º tempo (em segundos): 2º tempo (em segundos):

Teste de passe de Mor e Christian [?]:

Pontuação total na 1ª meta: Pontuação total na 2ª meta: Pontuação total na 3ª meta:

Teste de potência anaeróbica - Running Anaerobic Sprint Test (R.A.S.T.) [?]:

1º tempo (em segundos): 2º tempo (em segundos): 3º tempo (em segundos): 4º tempo (em segundos): 5º tempo (em segundos): 6º tempo (em segundos):

Teste de 1000 metros em pista [?]: Tempo (em segundos):

Teste de velocidade cíclica de 20 metros [?]: Tempo (em segundos):



© 2014 grupocer.org
 Centro de Estudos do Risco (CER-USP/UFSCar) ([Google map](#))
 Email: avaliacao@cer.ufscar.br



Figura 4.4: Formulário de cadastro de aluno/atleta no sistema i-Sports.



Figura 4.5: Menu de relatórios para a escola Meninos da Vila no i-Sports.

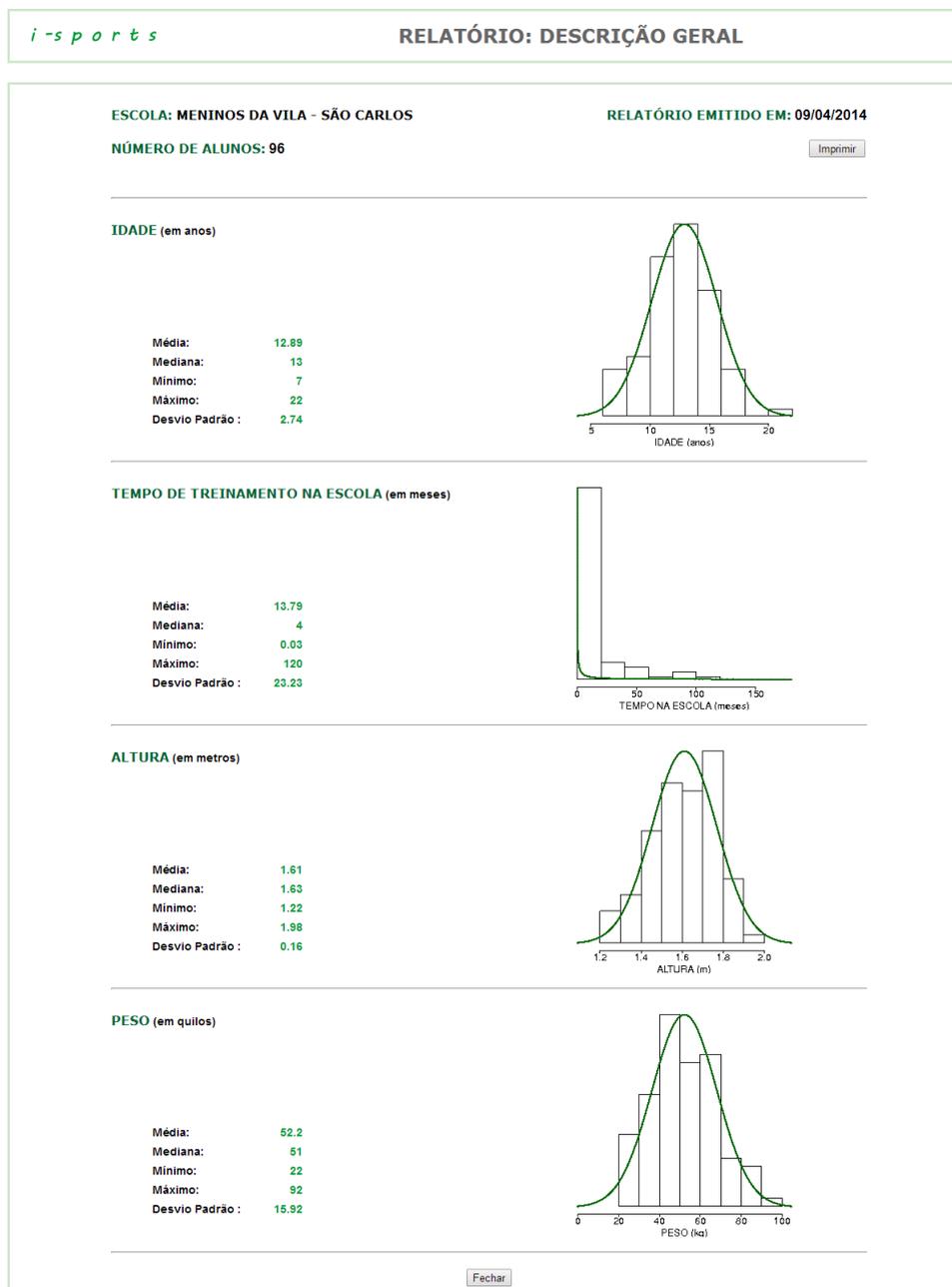


Figura 4.6: Relatório geral dos alunos.

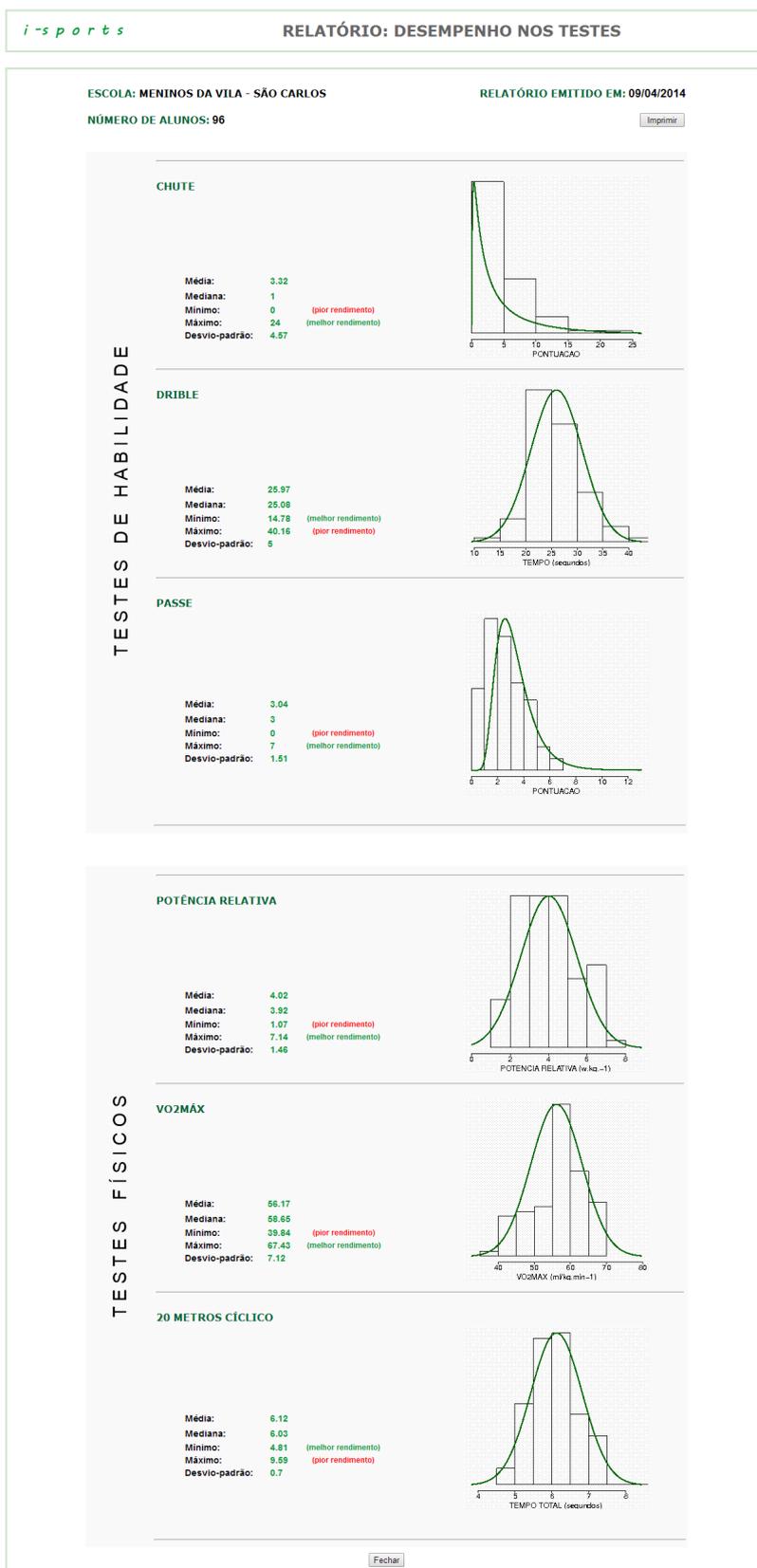


Figura 4.7: Relatório geral dos testes.



Figura 4.8: Melhores jogadores fisicamente.



Figura 4.9: Jogadores com maior habilidade técnica.



Figura 4.10: Jogadores com melhores avaliações.



Figura 4.11: Jogadores mais consistentes.

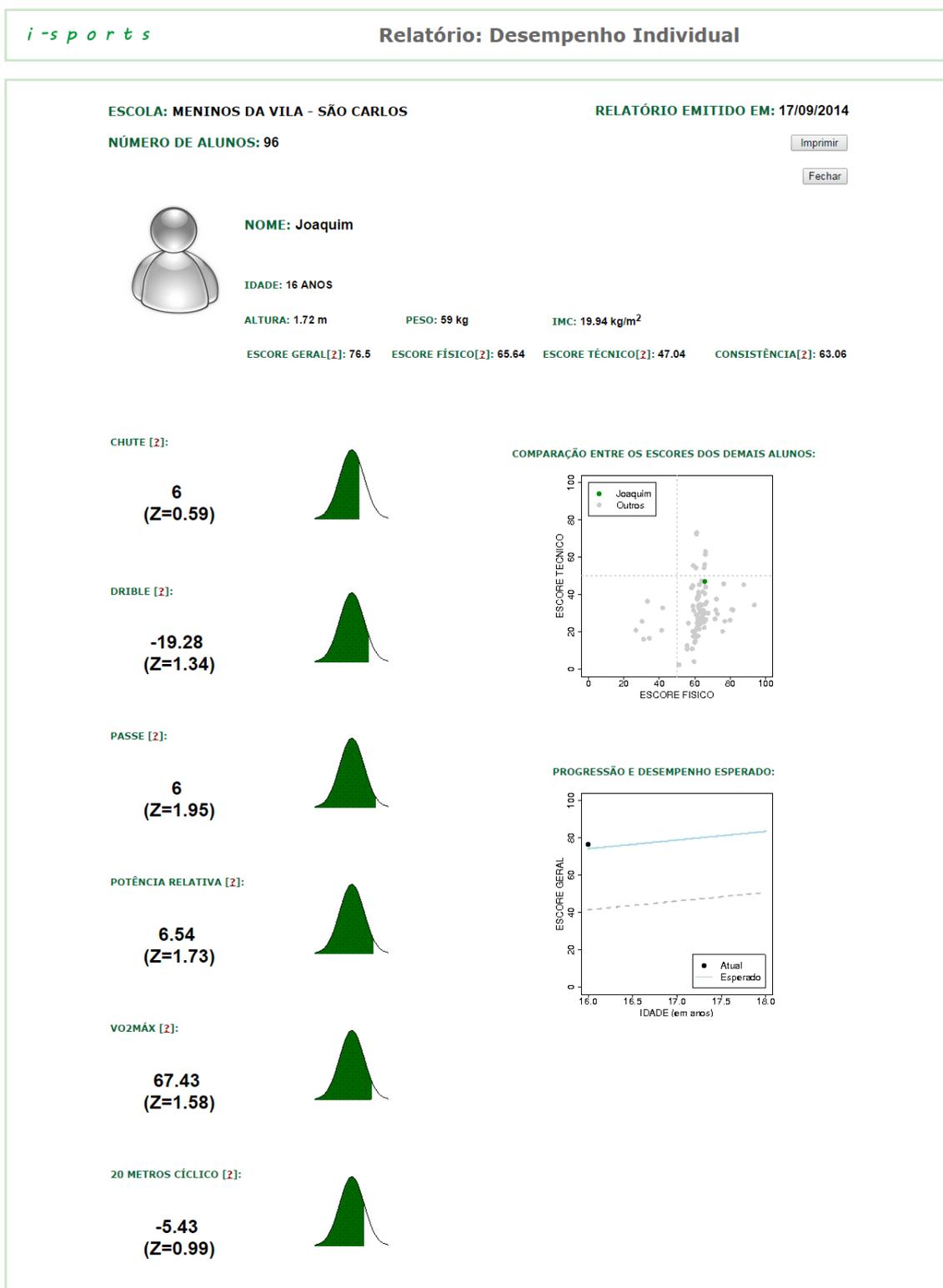


Figura 4.12: Aluno com maior escore geral e maior consistência.



Figura 4.13: Aluno com menor escore geral.

Capítulo 5

Considerações Finais

Neste trabalho, visamos a criação de indicadores para avaliar o desempenho de indivíduos praticantes da modalidade esportiva de futebol de campo utilizando técnicas estatísticas. As ferramentas desenvolvidas aqui possibilitam que as comparações entre os atletas sejam muito eficazes, permitindo ordenar e identificar indivíduos com diferentes habilidades.

Dentre as metodologias aplicadas, observamos que a abordagem multivariada se mostrou mais adequada, pois considera a interação entre as variáveis analisadas. Como exemplo desse fato podemos citar os alunos que se destacaram na metodologia Z, mas não apareceram dentre os melhores com base nos indicadores propostos.

Além disso, a implantação dos indicadores no Sistema Online i-Sports foi extremamente útil, uma vez que permitiu analisar os indivíduos tanto globalmente, quanto individualmente, ordenando-os de maneira que se possa observar quais são os atletas mais aptos à prática do futebol.

Vale ressaltar a importância de aspectos interdisciplinares envolvidos, em uma conjugação das seguintes áreas: Computação, Educação Física e Estatística. Ainda, o sistema criado é completamente inovador, construído a partir de *softwares* livres, sendo viável replicá-lo para qualquer modalidade esportiva.

Trabalhos futuros devem levar em consideração procedimentos vinculados à Teoria de Resposta ao Item, possivelmente utilizando variáveis contínuas [45]. Sabidamente, métodos que também consideram a dificuldade dos diferentes testes avaliados devem apresentar melhores resultados em termos avaliativos.

Referências Bibliográficas

- [1] REILLY, T.; BANGSBO, J.; FRANKS, A. Anthropometric and physiological predispositions for elite soccer. *Journal of Sports Sciences*, v. 18, n. 9, p. 669–683, 2000.
- [2] WILLIAMS, A. M.; REILLY, T. Talent identification and development in soccer. *Journal of Sports Sciences*, v. 18, n. 9, p. 657–667, 2000.
- [3] BALAGUER, I.; GONZÁLEZ, L.; FABRA, P.; CASTILLO, I.; MERCÉ, J.; DUDA, J. L. Coaches' interpersonal style, basic psychological needs and the well-and ill-being of young soccer players: A longitudinal analysis. *Journal of Sports Sciences*, v. 30, n. 15, p. 1619–1629, 2012.
- [4] LE GALL, F.; CARLING, C.; WILLIAMS, M.; REILLY, T. Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, v. 13, n. 1, p. 90–95, 2010.
- [5] F. HELSEN, W.; HODGES, N. J.; WINCKEL, J. V.; STARKES, J. L. The roles of talent, physical precocity and practice in the development of soccer expertise. *Journal of Sports Sciences*, v. 18, n. 9, p. 727–736, 2000.
- [6] VAEYENS, R.; MALINA, R. M.; JANSSENS, M.; VAN RENTERGHEM, B.; BOURGOIS, J.; VRIJENS, J.; PHILIPPAERTS, R. M. A multidisciplinary selection model for youth soccer: the ghent youth soccer project. *British Journal of Sports Medicine*, v. 40, n. 11, p. 928–934, 2006.
- [7] WALKER, S.; TURNER, A. A one-day field test battery for the assessment of aerobic capacity, anaerobic capacity, speed, and agility of soccer players. *Strength & Conditioning Journal*, v. 31, n. 6, p. 52–60, 2009.

- [8] WILLIAMS, A. M. Perceptual skill in soccer: Implications for talent identification and development. *Journal of Sports Sciences*, v. 18, n. 9, p. 737–750, 2000.
- [9] WILLIAMS, A. M.; ERICSSON, K. A. Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science*, v. 24, n. 3, p. 283–307, 2005.
- [10] HIROSE, N. Relationships among birth-month distribution, skeletal age and anthropometric characteristics in adolescent elite soccer players. *Journal of Sports Sciences*, v. 27, n. 11, p. 1159–1166, 2009.
- [11] HOWE, M. J.; DAVIDSON, J. W.; SLOBODA, J. A. Innate talents: Reality or myth? *Behavioral and Brain Sciences*, v. 21, n. 03, p. 399–407, 1998.
- [12] CASTAGNA, C.; MANZI, V.; IMPELLIZZERI, F.; WESTON, M.; ALVAREZ, J. C. B. Relationship between endurance field tests and match performance in young soccer players. *The Journal of Strength & Conditioning Research*, v. 24, n. 12, p. 3227–3233, 2010.
- [13] MATSUDO, V. K.; RIVET, R. E.; PEREIRA, M. H. Standard score assessment on physique and performance of brazilian athletes in a six tiered competitive sports model. *Journal of Sports Sciences*, v. 5, n. 1, p. 49–53, 1987.
- [14] PAPIĆ, V.; ROGULJ, N.; PLEŠTINA, V. Identification of sport talents using a web-oriented expert system with a fuzzy module. *Expert Systems with Applications*, v. 36, n. 5, p. 8830–8838, 2009.
- [15] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [16] BRADLEY, P. S.; CARLING, C.; GOMEZ DIAZ, A.; HOOD, P.; BARNES, C.; ADE, J.; BODDY, M.; KRUSTRUP, P.; MOHR, M. Match performance and physical capacity of players in the top three competitive standards of english professional soccer. *Human Movement Science*, v. 32, n. 4, p. 808–821, 2013.
- [17] REILLY, T.; WILLIAMS, A. M.; NEVILL, A.; FRANKS, A. A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*, v. 18, n. 9, p. 695–702, 2000.

- [18] REBELO, A.; BRITO, J.; SEABRA, A.; OLIVEIRA, J.; KRUSTRUP, P. Physical match performance of youth football players in relation to physical capacity. *European Journal of Sport Science*, v. 14, n. sup1, p. S148–S156, 2014.
- [19] MUJIKA, I.; SANTISTEBAN, J.; IMPELLIZZERI, F. M.; CASTAGNA, C. Fitness determinants of success in men's and women's football. *Journal of Sports Sciences*, v. 27, n. 2, p. 107–114, 2009.
- [20] RÖSCH, D.; HODGSON, R.; PETERSON, L.; GRAF-BAUMANN, T.; JUNGE, A.; CHOMIAK, J.; DVORAK, J. Assessment and evaluation of football performance. *The American Journal of Sports Medicine*, v. 28, n. suppl 5, p. S–29, 2000.
- [21] ALI, A. Measuring soccer skill performance: a review. *Scandinavian Journal of Medicine & Science in Sports*, v. 21, n. 2, p. 170–183, 2011.
- [22] ELFERINK-GEMSER, M. T.; HUIJGEN, B. C.; COELHO-E-SILVA, M.; LEM-MINK, K. A.; VISSCHER, C. The changing characteristics of talented soccer players—a decade of work in groningen. *Journal of Sports Sciences*, v. 30, n. 15, p. 1581–1591, 2012.
- [23] JANKOVIC, S.; MATKOVIC, B.; MATKOVIC, B. Functional abilities and process of selection in soccer. In: . c1997. p. 23–26.
- [24] JANSSENS, M.; VAN RENTHERGEM, B.; BOURGOIS, J.; VRIJENS, J. Physical fitness and specific motor performance of young soccer players aged 11-12 years. *Science and Football IV*, p. 258–262, 2002.
- [25] ROSTGAARD, T.; IAIA, F. M.; SIMONSEN, D. S.; BANGSBO, J. A test to evaluate the physical impact on technical performance in soccer. *The Journal of Strength & Conditioning Research*, v. 22, n. 1, p. 283–292, 2008.
- [26] HUIJGEN, B. C.; ELFERINK-GEMSER, M. T.; POST, W.; VISSCHER, C. Development of dribbling in talented youth soccer players aged 12–19 years: A longitudinal study. *Journal of Sports Sciences*, v. 28, n. 7, p. 689–698, 2010.
- [27] ALI, A.; WILLIAMS, C.; HULSE, M.; STRUDWICK, A.; REDDIN, J.; HOWARTH, L.; ELDRED, J.; HIRST, M.; MCGREGOR, S. Reliability and validity of two tests of soccer skill. *Journal of Sports Sciences*, v. 25, n. 13, p. 1461–1470, 2007.

- [28] DARDOURI, W.; AMIN SELMI, M.; HAJ SASSI, R.; GHARBI, Z.; REBHI, A.; MOALLA, W. Reliability and discriminative power of soccer-specific field tests and skill index in young soccer players. *Science & Sports*, v. 29, n. 2, p. 88–94, 2014.
- [29] MOR, D.; CHRISTIAN, V. The development of a skill test battery to measure general soccer ability. *NCJ Health Phys Ed*, v. 15, p. 30–39, 1979.
- [30] MOHR, M.; KRUSTRUP, P.; BANGSBO, J. Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences*, v. 21, n. 7, p. 519–528, 2003.
- [31] DIAZ C, F. J.; MONTANO C, J. G.; MELCHOR M, M. T.; GUERRERO G, J. H.; TOVAR G, J. A. Validation and confiability of aerobic test of 1,000 meters. *Revista de Investigacion Clinica*, v. 52, n. 1, p. 44–51, 2000.
- [32] ZACHAROGIANNIS, E.; PARADISIS, G.; TZIORTZIS, S. An evaluation of tests of anaerobic power and capacity. *Medicine & Science in Sports & Exercise*, v. 36, n. 5, p. S116, 2004.
- [33] JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 6. ed. Pearson, 2007.
- [34] NELSEN, R. B. *An introduction to copulas*. Springer, 1999.
- [35] KOJADINOVIC, I.; YAN, J. et al. Modeling multivariate distributions with continuous margins using the copula r package. *Journal of Statistical Software*, v. 34, n. 9, p. 1–20, 2010.
- [36] GENEST, C.; RÉMILLARD, B. Test of independence and randomness based on the empirical copula process. *Test*, v. 13, n. 2, p. 335–369, 2004.
- [37] BERG, D. Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, v. 15, n. 7-8, p. 675–701, 2009.
- [38] GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, v. 44, n. 2, p. 199–213, 2009.
- [39] GENEST, C.; RÉMILLARD, B. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In: . c2008. v. 44. p. 1096–1127.

- [40] KOJADINOVIC, I.; YAN, J.; HOLMES, M. Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica*, v. 21, n. 2, p. 841, 2011.
- [41] HOFERT, M.; KOJADINOVIC, I.; MAECHLER, M.; YAN, J. *copula: Multivariate dependence with copulas*, 2014. R package version 0.999-12.
- [42] KOJADINOVIC, I.; YAN, J. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, v. 21, n. 1, p. 17–30, 2011.
- [43] HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. *Multivariate data analysis*. Pearson Prentice Hall Upper Saddle River, NJ, 2006. v. 6.
- [44] GROUP, M. M. *World internet users and population stats*, Dez. 2013. Acessado em: 22 de Novembro de 2014.
- [45] NOEL, Y.; DAUVIER, B. A beta item response model for continuous bounded responses. *Applied Psychological Measurement*, v. 31, n. 1, p. 47–73, 2007.