



Substituições no jogo de Futebol – Análise de dados em épocas de competição sem e com COVID-19

Ricardo Jorge dos Santos Balteiro

ISCAC | 2021



**Instituto Superior
de Contabilidade
e Administração**

Politécnico de Coimbra

Ricardo Jorge dos Santos Balteiro

Substituições no jogo de Futebol – Análise de dados em épocas de competição
sem e com COVID-19

Coimbra, setembro de 2021



**Instituto Superior
de Contabilidade
e Administração**

Politécnico de Coimbra

Ricardo Jorge dos Santos Balteiro

Substituições no jogo de Futebol – Análise de dados em épocas de competição sem e com COVID-19

Dissertação submetida ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Análise de Dados e Sistemas de Apoio à Decisão realizado sob a orientação do Professor Doutor Alexandre Silva e coorientação do Professor Doutor Rui Mendes.

Coimbra, setembro de 2021

TERMO DE RESPONSABILIDADE

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau acadêmico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

“We are told that talent creates its own opportunities. But it sometimes seems that intense desire creates not only its own opportunities, but its own talents.”

- Eric Hoffer

DEDICATÓRIA

Dedico esta tese ao verdadeiro “mestre” da minha vida, o meu avô José, que tantas vezes me fez rir e que há 3 anos me fez chorar. Obrigado, avô, para sempre no meu coração.

AGRADECIMENTOS

Ao longo desta jornada o contributo, os conselhos, a inspiração e a motivação que recebi foram absolutamente fundamentais para a poder concluir da melhor maneira. Desta forma, queria expressar o meu sincero agradecimento:

Ao Professor Doutor Alexandre Silva, orientador desta dissertação, Professor do Instituto Superior de Contabilidade de Coimbra e coordenador do Mestrado de Análise de Dados e Sistemas de Apoio à Decisão, por toda a paciência, orientação, disponibilidade e críticas construtivas. Desde a primeira aula que mostrou ser um exemplo de um excelente Professor.

Ao Professor Doutor Rui Mendes, coorientador desta dissertação, Professor da Escola Superior de Educação de Coimbra, por toda a disponibilidade e apoio que sempre demonstrou e por todo o *apport* científico e estrutural que deu.

Ao Professor Doutor Fernando Martins, Professor da Escola Superior de Educação de Coimbra, pelo interesse no meu trabalho e pelos conselhos científicos e o acrescento de rigor.

Aos meus colegas e amigos da AGEAS, em especial à Manuela e ao Ricardo, por terem sido os meus primeiros e brilhantes professores no mundo do trabalho e por toda a amizade e conselhos que levarei para a vida.

À Plural, por me dar a oportunidade de começar a minha carreira na análise de dados. Um agradecimento especial ao José Ferreira por toda a tutoria, apoio, conselhos e me proporcionar todas as condições para crescer.

À Dra. Paula Castilho por me ensinar que nunca serei perfeito mas que posso sempre melhorar.

Aos futuros mestres, Pedro Oliveira e Rui Esteves, pelo excelente trio de amizade e entreaajuda que formámos ao longo deste mestrado.

À família do Sporting Clube Povoense, em especial ao Pedro Ratola, que me ensinou a ser competitivo e a trabalhar por uma causa. Vou recordar sempre todos os ensinamentos preciosos e momentos felizes vividos.

Aos meus grandes amigos Benjamin, Bruno, David, Gonçalo, Hugo e Rodrigo por todo o apoio, risos, conversas e amizade. Sem dúvida que não poderia ter amigos melhores.

Ao meu tio Bruno, por todo interesse e ajuda na minha vida acadêmica.

Aos meus tios José e Lúcia, e à minha prima Carolina, pelo apoio constante e carinho que sempre tiveram para comigo.

Aos meus avós, Carlos e Fátima, por todo o carinho e interesse constante, sempre preocupados com o meu bem-estar.

À minha avó Bertina, por ser um pilar fundamental na minha vida, por todos os sacrifícios, dedicação e preocupação constante.

Ao meu irmão Gonçalo, por ser o meu melhor amigo, por todas as vezes que nos rimos e conversamos sobre tudo e sobre nada, pela alegria e preocupação constantes, por ser sempre o meu conselheiro e a pessoa que admiro.

E, por último, os meus pais, porque se hoje aqui cheguei lhes devo a eles. Por todo o amor, apoio e sacrifícios que fizeram por mim. Por estarem sempre lá quando eu mais precisei, por me darem a mão e me mostrarem o caminho certo. Não existem palavras para descrever toda admiração que sinto por vocês. Obrigado por tudo.

RESUMO

O Futebol de 11 é o desporto mais visto e praticado em todo o mundo, onde os jogadores das duas equipas tentam realizar várias ações técnicas e táticas com o intuito de marcar golos e ganhar a partida.

A análise de dados é um procedimento que permite tomar decisões sustentadas em informações retiradas dos dados. No Futebol assume uma crescente importância, estando cada vez mais presente na preparação tática dos jogos, na monitorização do desempenho dos jogadores e *scouting*.

O presente estudo teve como objetivo analisar o impacto da possibilidade de realizar cinco substituições, ao invés das tradicionais três substituições, em virtude das alterações provocadas pela pandemia de COVID-19.

A amostra deste estudo foi composta por 192 jogos oficiais da Liga dos Campeões, respeitante à fase de grupos da edição 2019/20 e 2020/21, onde foram analisadas 40 variáveis.

Conclui-se que apesar desta alteração as variáveis de jogo nas duas épocas são muito homogéneas. Parece ser mais relevante para os treinadores a utilização de uma única substituição adicional, isto é, a quarta substituição. Também é evidenciado que o aumento da possibilidade de realizar mais substituições contribui para renovar a capacidade da equipa, com vantagens táticas e de prevenção da ocorrência de lesões. A primeira e a última substituição de cada edição da competição ocorrem praticamente no mesmo tempo de jogo, mostrando no que diz respeito às substituições de jogadores, que a opção dos treinadores centra-se na estratégia tática.

Palavras-chave: Futebol, análise de dados, COVID-19, substituições, análise de jogo.

ABSTRACT

11 football is the most watched and practiced sport in the world, where players from both teams try to carry out various technical and tactical actions in order to score goals and win the match.

Data analysis is a procedure that allows decision-making based on information taken from the data. In football, it assumes a growing importance, increasingly present in the tactical preparation of games, in monitoring the performance of players and scouts.

The present study aimed to analyze the impact of the possibility of performing five replacements, when applying the traditional three replacements, due to the changes caused by the COVID-19 pandemic.

The study sample consisted of 192 official Champions League games, concerning the group stage of the 2019/20 and 2020/21 editions, where 40 variables were analyzed.

It is concluded that despite this change, the game variables in the two seasons are very homogeneous. It seems to be more relevant for coaches to use a single additional substitution, that is, the fourth substitution. It is also evident that the increase in the possibility of making more substitutions contributes to renew the team's capacity, with tactical advantages and preventing the occurrence of injuries. The first and the last substitution of each edition occur at practically the same time of the game, showing, with regard to player substitutions, that the coaches' option is centered on tactical strategy.

Keywords: Football, data analysis, COVID-19, substitutions, game analysis.

ÍNDICE GERAL

TERMO DE RESPONSABILIDADE.....	i
DEDICATÓRIA	iii
AGRADECIMENTOS	iv
RESUMO.....	vi
ABSTRACT.....	vii
ÍNDICE GERAL.....	viii
ÍNDICE DE FIGURAS	x
ÍNDICE DE TABELAS.....	xi
ABREVIATURAS / ACRÓNIMOS / SIGLAS	xii
INTRODUÇÃO	1
PARTE I – REVISÃO DA LITERATURA	3
1 – Introdução à análise de dados: <i>Data Analytics, Data Mining e Big Data</i>	4
2 – Covid-19	5
3 – Análise de dados em Futebol	6
4 – Substituições e as suas conclusões.....	15
PARTE II – METODOLOGIA.....	21
5 – Caracterização da Amostra	22
6 – Extração e registo dos dados.....	23
7 – Variáveis	23
8 – Procedimentos estatísticos	25
PARTE III – ANÁLISE E DISCUSSÃO DOS RESULTADOS	29
9 – Introdução	30
10 – Estatística descritiva	31
10.1 – ANÁLISE DESCRITIVA DAS VARIÁVEIS LOCAL DE JOGO E RESULTADO	31

10.2 – ANÁLISE DESCRITIVA DAS VARIÁVEIS DE JOGO NAS DUAS EDIÇÕES DA LC (2019/20 E 2020/21).....	32
10.3 – ANÁLISE DESCRITIVA DO NÚMERO E TEMPO DE REALIZAÇÃO DAS SUBSTITUIÇÕES NAS DUAS EDIÇÕES DA LC (2019/20 E 2020/21).....	33
10.4 – ANÁLISE DESCRITIVA DO TIPO DE SUBSTITUIÇÕES.....	35
10.5 – ANÁLISE DESCRITIVA DOS GOLOS APÓS AS SUBSTITUIÇÕES.....	36
10.6 – ANÁLISE DESCRITIVA DO NÚMERO DE LESÕES.....	36
11 – Estatística inferencial.....	37
11.1 – REGRESSÃO LOGÍSTICA.....	37
11.2 – TESTE U DE MANN-WHITNEY.....	38
11.3 – EFFECT SIZE (TESTE d DE COHEN).....	38
11.4 – ANÁLISE DE SOBREVIVÊNCIA.....	39
CONCLUSÃO.....	43
REFERÊNCIAS BIBLIOGRÁFICAS.....	45

ÍNDICE DE FIGURAS

Figura 1. Logotipo da Liga dos Campeões (UEFA, 2021).....	22
Figura 2. Distribuição geográfica das equipas por país edição LC 2019/20 (elaboração própria).....	30
Figura 3. Distribuição geográfica das equipas por país edição LC 2020/21 (elaboração própria).....	30
Figura 4. Análise de sobrevivência: curvas do tempo da 1ª substituição para ambas as edições da LC.....	39
Figura 5. Análise de sobrevivência: curvas do tempo da 3ª substituição e da 5ª substituição para ambas as edições da LC	40

ÍNDICE DE TABELAS

Tabela 1. Síntese de estudos relativos à análise de dados em Futebol (elaboração própria)	6
Tabela 2. Frequência de variáveis investigadas nos estudos de análise de dados no Futebol (elaboração própria).....	11
Tabela 3. Análise do valor percentual das variáveis local de jogo e resultado nas duas edições da LC (LC 2019/20 e 2020/21) (elaboração própria)	31
Tabela 4. Análise do valor percentual da relação entre as variáveis local de jogo e resultado nas duas edições da LC (LC 2019/20 e 2020/21) (elaboração própria)	32
Tabela 5. Comparação dos valores da estatística descritiva das variáveis de jogo nas duas edições da LC (2019/20 e 2020/21) (elaboração própria)	32
Tabela 6. Análise do valor absoluto, médio e coeficiente de variação de cada substituição nas duas competições (LC 2019/20 e 2020/21) (elaboração própria)	33
Tabela 7. Valores absolutos e percentuais de substituições neutras, ofensivas e defensivas nas duas competições (LC 2019/20 e 2020/21) (elaboração própria)	35
Tabela 8. Análise do golo após a substituição (elaboração própria).....	36
Tabela 9. Análise do número de lesões (elaboração própria)	37
Tabela 10. Output regressão logística SPSS (elaboração própria)	37
Tabela 11. Output teste Mann-Whitney SPSS (elaboração própria)	38
Tabela 12. Output teste d de Cohen SPSS (elaboração própria).....	39
Tabela 13. Regressão de Cox.....	40

ABREVIATURAS / ACRÓNIMOS / SIGLAS

COVID-19 – *Coronavirus Disease 2019*

CV – Coeficiente de Variação

DGS – Direção Geral da Saúde

FIFA – *Fédération Internationale de Football Association*

INE – Instituto Nacional de Estatística

LC – Liga dos Campeões

SPPS – *Statistical Package for the Social Sciences*

SARS-CoV-2 – *Severe Acute Respiratory Syndrome Coronavirus 2*

UEFA – União das Associações Europeias de Futebol

INTRODUÇÃO

A análise de dados é uma das etapas mais importantes para a tomada de decisões nas organizações. Todos os dias, milhões de empresas geram milhões de dados que fornecem informações fundamentais para análises preditivas e não preditivas. Contudo, a maioria das empresas deparam-se com o problema de terem muitos dados, mas poucas informações. É então fundamental transformar esses dados em informações (Mcafee & Brynjolfsson, 2012).

O desporto, em particular, enfrenta uma revolução trazida pela análise de dados em que se procura encontrar as melhores estratégias para ganhar (Aguiar, 2018). A quantidade de dados existentes e gerados em todos os desportos é verdadeiramente substancial. Estes dados podem ter origem na *performance* individual de cada atleta, na forma como a equipa joga, nas decisões táticas do treinador ou nos próprios eventos desportivos (Schumaker *et al.*, 2010). Começa a ser perentório para qualquer equipa de desporto, sobretudo de alta competição, utilizar os dados em seu benefício e estabelecer estratégias competitivas para alcançar o sucesso a curto e longo prazo (Morgulev *et al.*, 2018).

O presente trabalho tem como objetivo compreender o impacto da possibilidade das equipas de Futebol realizarem cinco substituições, em virtude das alterações provenientes da pandemia COVID-19, ao invés das tradicionais três substituições.

Em concreto, pretende-se obter informação para responder às seguintes questões:

1. São esgotadas nos dois ambientes competitivos todas as substituições possíveis;
2. A possibilidade de mais substituições na época com COVID-19 é potenciada pelos treinadores;
3. O *timing* da primeira substituição é distinto nos dois contextos de competição (sem e com COVID-19);
4. O número de lesões ocorridas em jogo não é igual nos dois contextos competitivos.

A nível estrutural esta dissertação está dividida em quatro partes: revisão de literatura, metodologia, análise e discussão dos resultados e conclusões.

A primeira parte centra-se na revisão da literatura onde se contextualiza o tema e informações através de estudos, artigos e trabalhos já publicados sobre esta temática de forma a efetuar uma comparação com os resultados desta tese.

No capítulo seguinte são apresentadas as características metodológicas de investigação utilizada onde é feita uma caracterização da amostra e é referenciada a origem, a recolha e o tratamento dos dados, bem como a caracterização das variáveis extraídas e os procedimentos estatísticos para as analisar.

A terceira parte tem por base a análise e a discussão dos resultados.

Por fim, o trabalho é terminado com as conclusões. São também listadas as limitações desta pesquisa e as perspetivas de desenvolvimento de trabalhos futuros a partir desta dissertação.

PARTE I – REVISÃO DA LITERATURA

1 – Introdução à análise de dados: *Data Analytics*, *Data Mining* e *Big Data*

O mundo está cada vez mais formatado para a utilização dos dados como informação vital para o desempenho das organizações. Existem quantidades infinitas de dados disponíveis que permitem um maior conhecimento da realidade dos setores e ajudam a resolver problemas (Hinshelwood, 2018).

A análise de dados, em inglês vulgarmente conhecida por *Data Analysis*, consiste no processo de tratamento, transformação e avaliação dos dados utilizando ferramentas analíticas e estatísticas para descobrir informações úteis. O objetivo passa por facilitar a tomada de decisões (Grant, 2020).

Associado à análise de dados existem três conceitos que importa definir:

1. ***Data Analytics*** – é a ciência de analisar dados em bruto para retirar informações e conclusões sobre os mesmos. Muitas das técnicas e processos de análise de dados foram automatizados em processos mecânicos e algoritmos. As técnicas de análise de dados podem revelar tendências e métricas que, de outra forma, seriam perdidas na massa de dados. Estas informações podem então ser utilizadas para otimizar processos de modo a aumentar a eficiência geral de um negócio ou sistema (Frankenfield, 2020).
2. ***Big Data*** – pode ser definido como a recolha, o processamento, a análise e a visualização associada a conjuntos de dados muito grandes (Emmanuel & Stanier, 2016). É um termo utilizado em dados massivos com uma estrutura grande, variada e complexa em que existe dificuldade de os armazenar, analisar e visualizar através dos processos manuais e tradicionais. O processo de pesquisa em grandes quantidades de dados para revelar padrões e correlações ocultas denomina-se por *Big Data Analytics* (Sagiroglu & Sinanc, 2013).
3. ***Data Mining*** – em português significa Mineração de Dados, consiste num processo de análise projetado para explorar um grande volume de dados, tipicamente relacionados com negócios, pesquisas científicas e mercado, com o intuito de descobrir padrões e/ou relações sistemáticas entre as variáveis e, posteriormente, validá-los aplicando os padrões detetados aos novos subconjuntos de dados (Han *et al.*, 2012). O processo consiste em três etapas: exploração,

construção de modelo ou definição do padrão e validação/verificação (Carvalho *et al.*, 2003).

2 – Covid-19

Coronavirus Disease 2019 (Covid-19) é uma doença descoberta em 2019 e provocada pelo *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2), um agente que nunca tinha sido detetado anteriormente em seres humanos e cuja fonte de infeção é ainda desconhecida (Direção Geral da Saúde, 2020).

Como fator de combate à transmissão do vírus os governos de cada país têm tomado um conjunto de medidas que restringem a circulação e que obrigam à contração económica. No caso português, estas medidas são influenciadas por diversos indicadores tais como o número de novos casos, o R (número de reprodução que mede o número médio de infeções geradas por cada pessoa infetada), o número de doentes internados e em cuidados intensivos e com sequentes taxas de ocupação hospitalares e o número de infetados por um determinado valor científico e politicamente deliberado (e.g., 100 mil habitantes em cada concelho), permitindo que as medidas possam ser ajustadas de concelho para concelho mediante tal indicador.

Desta forma, a COVID-19 tem tido um enorme impacto no mundo, o que tem levado a alterações não só a nível económico como também no comportamento das pessoas e das organizações. No caso concreto de Portugal, a pandemia provocou uma queda de 7,6% na economia portuguesa em 2020 face a 2019, divulgou o Instituto Nacional de Estatística (INE) (Laja, 2021).

Esta pandemia permitiu a nível mundial consciencializar e aumentar a importância da análise dos dados como fator de informações cruciais para uma boa tomada de decisão.

Nunca como até agora a análise de dados foi tão decisiva, tornando-se um fator crucial para a gestão da pandemia. É a informação recolhida que depois de tratada e analisada, permite identificar padrões de mobilidade das pessoas, zonas de risco, previsão de contágios ou até a taxa de utilização de transportes públicos (Aguiar, 2020^a).

No que diz respeito ao Futebol, este vírus obrigou a que fossem tomadas medidas por parte da *Fédération Internationale de Football Association* (FIFA) que alteraram várias variáveis neste jogo desportivo coletivo. As alterações mais importantes prendem-se com

a proibição de público nos estádios, o que obriga a repensar se o fator “casa” ainda existe, tanto para os jogadores como para os próprios árbitros que podem ser menos parciais, e a possibilidade de realizar cinco substituições por jogo, em três momentos, tendo em vista a proteção dos jogadores e das equipas, ao invés das tradicionais três (Aguiar, 2020^b).

3 – Análise de dados em Futebol

Nos últimos anos tem-se vindo a assistir a uma crescente pesquisa e publicação científica de trabalhos que relacionam a matemática e a estatística com o Futebol.

Na Tabela 1 sistematizou-se a literatura dos principais estudos publicados e validados na comunidade científica neste domínio.

Tabela 1. Síntese de estudos relativos à análise de dados em Futebol (elaboração própria)

Autor	Objetivo	Amostra	Variáveis	Resultados/conclusões
Carling <i>et al.</i> (2010)	Investigar a taxa/ritmo de esforço (<i>work-rate</i>) de jogadores suplentes no Futebol profissional. O desempenho dos suplentes foi comparado com o dos jogadores que substituíram, com os companheiros na mesma posição que permaneceram em campo após a substituição e em relação ao seu desempenho habitual no início dos jogos	26 jogadores (11 médios e 14 avançados) num clube da primeira divisão francesa (Lille FC)	1 – Distância percorrida 2 – Posição dos jogadores 3 – Tempo de recuperação entre esforços de alta intensidade 4 – Taxa de esforço (<i>work-rate</i>)	Os substitutos do meio-campo percorreram uma distância total e distância em altas intensidades maiores e tiveram um tempo de recuperação menor entre os esforços de alta intensidade em comparação com outros jogadores de meio-campo que permaneceram em campo. Os atacantes percorreram menos distância nos primeiros 10 minutos como substitutos em comparação com seu perfil de <i>work-rate</i> habitual nos primeiros 10 minutos ao iniciar as partidas, enquanto este facto não foi observado nos meio-campistas. Em comparação com os substitutos do meio-campo, os avançados substitutos não utilizam todo o seu potencial físico
Myers (2011)	Propor regra de decisão para o tempo das substituições no Futebol	485 observações de jogos de Futebol (155 da Premier League Inglesa, 172 da Serie A Italiana e 158 da La	11 – Número de substituições 12 – Tempo de substituição 15 – Local de jogo 17 – Resultado Corrente 35 – Diferença de golos antes e	É usada a técnica Data Mining de árvores de decisão para desenvolver uma regra de decisão para orientar os treinadores sobre quando fazer cada uma das três substituições. A regra de decisão proposta demonstra entre 38% e 47% de eficácia quando utilizada e quando não utilizada entre 17% e 24% de eficácia

		Liga Espanhola)	depois de cada substituição 36 – Tipo de competição	
Coelho <i>et al.</i> (2012)	Estudar o efeito das substituições, realizadas na segunda parte, na intensidade do jogo de Futebol, comparando a intensidade de esforço dos jogadores de Futebol nas situações: primeiro tempo; segundo tempo; segundo tempo com substituições	45 jogadores de Futebol do sexo masculino; 29 jogos oficiais	5 – Peso 6 – Altura 7 – Percentagem de massa gorda 8 – Frequência cardíaca 9 – Consumo de oxigénio 10 – Intensidade de esforço	A diminuição da intensidade de esforço na segunda parte foi atenuada com a realização de substituições e foi evidenciado um maior tempo de permanência em zonas de alta intensidade quando comparado com a segunda parte sem substituições
Bradley <i>et al.</i> (2014)	Avaliar o desempenho de jogadores substitutos no Futebol de elite	Jogos da Premier League Inglesa; 810 jogadores que fizeram todo o jogo; 286 substituição; 286 jogadores que entraram no jogo; 94 jogadores que fizeram jogos completos vs jogos em que estes começaram no banco e depois entraram	2 – Posição do jogador 1 – Distância percorrida 11 – Número de substituições 12 – Tempo de substituição 13 – Velocidade máxima de corrida com e sem bola 14 – Passes bem efetuados	A maioria das substituições ocorrem entre os 60 e 85 minutos, sendo que são substituídos com mais frequência jogadores de posição de ataque; os jogadores que entram percorrem maiores distâncias de corrida em alta intensidade dos que os que completam o jogo inteiro ou são substituídos
Russiano (2016)	Análise o contexto e o efeito das substituições no Futebol de alto rendimento	1584 substituições de 279 jogos da fase de grupos da Liga dos Campeões das épocas de 2012/13, 2013/14 e 2014/15	15 – Local do jogo 16 – Qualidade do adversário 17 – Resultado corrente 18 – Período do jogo 19 – Tempo do golo sofrido 20 – Tempo de substituição 21 – Diferença do tempo entre o golo sofrido e a substituição 22 – Diferença do tempo entre a	Associação significativa entre as variáveis: qualidade do adversário*período do jogo; local do jogo*resultado corrente; resultado corrente*qualidade do adversário e o período do jogo*resultado corrente e a realização das substituições no Futebol de alto rendimento. As variáveis local do jogo e o tempo da substituição têm um poder significativo na predição do sucesso das substituições

			substituição e a alteração do resultado 22 – Alteração do resultado depois da substituição	
Gomez <i>et al.</i> (2016)	Determinar a influência das substituições na <i>performance</i> das equipas de elite	292 substituições, de 50 jogos de Futebol de equipas de elite	12 – Minuto da substituição 11 – Número de substituições 23 – Cartões amarelos 2 – Posição do jogador que sai e do jogador que entra 24 – Percentagem de tempo de posse de bola 25 – Remates na baliza 26 – Bolas recuperadas 16 – Qualidade do adversário 15 – Local do jogo 27 – Resultado do jogo	Primeira e segunda substituição ocorreram durante os 61 a 90 minutos, enquanto a terceira ocorre predominantemente no período de 76 a 90 minutos; Quando estão a perder as equipas fazem substituições mais cedo do que quando estão empatadas ou a ganhar; A equipa da casa faz substituições mais cedo do que a equipa visitante; A maioria das substituições envolveu jogadores do meio-campo (57,1%) e atacantes (27,1%); quanto melhor for a qualidade da equipa adversária, mais tarde ocorrem as substituições
Varela-Quintana <i>et al.</i> (2016)	Estudar o impacto de uma substituição adicional no prolongamento de um jogo de futebol (possível no Mundial de 2018) e como os treinadores reagiram a uma substituição adicional em situações anteriores (em 1995 passou a ser possível substituir três jogadores e não apenas dois, durante todo o jogo)	2623 substituições de 612 jogos da Série A Italiana, durante as épocas de 1994-95 e 1995-96	33 – Tipo de substituição 16** – Qualidade do adversário 37 – Jogos rivais 38 – Jornada 12 – Tempo de substituição 17* – Resultado corrente 39 – Valor do 11 inicial	Observou-se uma mudança no comportamento tático dos treinadores que mostraram preferência por substituições neutras (jogadores da mesma posição) ao invés de táticas, provavelmente para manter a intensidade do jogo e prevenir lesões
Caballero <i>et al.</i> (2017)	Analisar a influência de variáveis no rendimento das equipas de Futebol sub-18, em concreto no resultado final	132 jogos de Liga Regional Espanhola Sub-18	28 – Diferença de pontuação 29 – Classificação final na liga 15 – Localização da partida 30 – Primeira equipa a marcar 23 – Cartões	A equipa da casa quando ganha marca o primeiro golo do jogo em 79,4% das vezes, faz mais substituições (67,3%) e recebe menos cartões do que a equipa adversária (66,7%). Se a equipa da casa não for a primeira a marcar a probabilidade de ganhar é de 21,7%. As substituições são uma das variáveis que mais tem

			11 – Número de substituições 16* – Qualidade do adversário 31 – Superfície de campo	impacto na vitória da equipa. Se a equipa marca primeiro, joga em casa e faz mais substituições, a diferença de golos aumenta
Hills <i>et al.</i> (2018)	Revisão de literatura sobre a <i>performance</i> e a resposta dos substitutos no Futebol	13 artigos	Variáveis de revisão de literatura	Os substitutos do Futebol são normalmente introduzidos no intervalo ou durante o segundo tempo, geralmente na tentativa de compensar os efeitos da fadiga e/ou alterar a tática, embora possam existir outras motivações. Os suplentes, especialmente os médios, parecem conseguir ultrapassar o work rate dos jogadores que iniciaram o jogo. As observações das práticas utilizadas antes da entrada no campo, as próprias preocupações dos jogadores e a incapacidade dos jogadores introduzidos em campo de excederem o seu próprio ritmo de trabalho habitual do que quando iniciam, questionam a eficácia da preparação física e psicológica dos substitutos
Flôres <i>et al.</i> (2019)	Analisar a relação entre as substituições e os resultados de jogos de Futebol em campeonatos e taças	129 jogos de 16 campeonatos e 129 jogos de 16 taças nacionais (258 jogos de Futebol), os quais decorreram durante a temporada de 2016/17 em 16 países. 737 substituições em campeonatos e 721 em taças nacionais	11 – Número de substituições 32 – Golos marcados 15 – Local de jogo	Os jogadores que entraram marcaram 10,76% dos golos, correspondendo a 26,19% dos golos marcados após a primeira substituição. Nas taças nacionais, os substitutos foram responsáveis por 10% dos golos marcados, representando 24,84% dos golos após a primeira substituição. As equipas que jogam os campeonatos fizeram mais golos após a primeira substituição, após a terceira substituição, e após todas as substituições, comparado com as equipas nas taças nacionais
Lorenzo-Martínez <i>et al.</i> (2020)	Analisar o efeito das substituições de jogadores no comportamento tático no Futebol de alto rendimento	659 substituições em 234 partidas disputadas na Bundesliga alemã na	33 – Tipo de substituição 15 – Local de jogo 27 – Resultado do jogo 34 – Variáveis táticas	A análise do modelo linear misto mostrou diferentes impactos para as substituições neutras, defensivas e ofensivas. As equipas exibiram índice de alongamento significativamente menor após as substituições defensivas

		época 2016/17		A relação entre o comprimento e a largura aumentou com substituições neutras e ofensivas, enquanto a distância entre equipas diminuiu. O controlo do espaço no meio campo e no terço atacante também foram maiores após uma substituição atacante
Mota <i>et al.</i> (2020)	Avaliar o impacto das exigências físicas induzidas pelas regras do Futebol e de outros desportos coletivos com base nas substituições	Não aplicável	Não aplicável	Analisou-se e comparou-se as regras de jogo, como as dimensões do campo, tempo total de jogo, substituições permitidas, as alterações efetuadas devido à pandemia do COVID-19, entre outras, no Futebol, futsal, basquetebol e andebol. O Futebol tem uma exigência física muito superior devido às suas regras (e.g., comparativamente com os outros desportos, a taxa de substituições é substancialmente mais baixa e a dimensão do campo é mais elevada). Mesmo com a mudança de três para cinco substituições, as diferenças continuam a ser extremamente grandes

O Futebol é o jogo mais popular do mundo quer em número de praticantes, quer em número de espetadores. Este jogo consiste numa disputa entre duas equipas de 11 jogadores em campo, que podem utilizar todas as partes do corpo, exceto as mãos e os braços, de forma a tentar manobrar a bola e marcar golo à equipa adversária. Apenas o guarda-redes tem permissão para manusear a bola dentro da sua área. A regra do jogo é simples, a equipa que marcar mais golos é a vencedora (Weil *et al.*, 2020).

Desta forma, muitos estudos têm sido feitos nesta área proporcionando um conhecimento mais profundo sobre o Futebol. A maioria dos estudos foca-se sobretudo no comportamento do sistema coletivo e no perfil e atividade dos jogadores (Wu *et al.*, 2020). Para se ter uma melhor perceção dos elementos da amostra que foram averiguados, na Tabela 2 encontra-se a lista das variáveis utilizadas nesses estudos.

Tabela 2. Frequência de variáveis investigadas nos estudos de análise de dados no Futebol (elaboração própria)

Variável		Definição/Operacionalização	Frequência
1	Distância percorrida	Trajatória dos jogadores em campo medida em metros	2
2	Posição dos jogadores	Função dos jogadores no campo	2
3	Tempo de recuperação de esforços de alta intensidade	Tempo em segundos necessário para executar o próximo esforço	1
4	Taxa de esforço (<i>work-rate</i>)	A quantidade de esforço que um jogador coloca no jogo, especialmente quando não tem a bola. Esta taxa foi calculada com base na distância percorrida pelo jogador.	1
5	Peso	Número de quilogramas que cada jogador apresenta	1
6	Altura	Tamanho em metros de cada jogador	1
7	Porcentagem de Massa Gorda	Divisão entre a massa total de gordura corporal e a massa total do atleta	1
8	Frequência cardíaca	Número de batimentos cardíacos por minuto	1
9	Consumo máximo de oxigénio	Volume máximo de oxigénio consumido durante a realização de atividade física pelo jogador. Calculado através de uma corrida de 2400m na máxima velocidade	1
10	Intensidade de esforço	Considerada como a média da frequência cardíaca, expressa na percentagem da frequência cardíaca máxima de cada jogador e o tempo despendido em cada zona de intensidade	1
11	Número de substituições	Quantidade de jogadores suplentes que entram em campo	4
12	Tempo de substituição	Tempo de jogo aquando a realização da substituição	4

Substituições no jogo de Futebol – Análise de dados em época de competição sem e com COVID-19

13	Velocidade máxima de corrida com e sem bola	Combinação da corrida em alta intensidade e do sprint medido em metros por minuto	1
14	Passes bem efetuados	Passes efetuados que cheguem ao colega de equipa	1
15	Local de jogo	Indica se a equipa que solicitou a substituição estava a jogar em casa ou fora	4
16	Qualidade do adversário	Qualidade das equipas avaliada pelo ranking final da UEFA. Para classificar os jogos em função da qualidade do adversário, foi utilizado o valor da diferença dos grupos das equipas que se defrontavam entre si: 0 – jogo realizado contra adversários do mesmo nível; -1, -2 – jogo realizado contra adversário mais fraco; 1, 2 – jogo realizado contra adversário mais forte 16* – cálculo da diferença entre as classificações das equipas na liga todas as semanas durante a temporada 16** – Diferença na classificação nas últimas cinco temporadas	4
17	Resultado Corrente	Representa o resultado do jogo registado no momento da substituição solicitada pela equipa observada 17* – Se a equipa está a ganhar no momento da substituição (variável <i>dummy</i>)	3
18	Período do jogo	Momento do jogo, dividido em 4 períodos (0-45 min; 46-60 min; 61-75 min; 76-90 min), em que acontecem as substituições	1
19	Tempo de golo sofrido	Tempo decorrido entre o início do jogo e a marcação do golo pela equipa adversária	1
20	Diferença do tempo entre o golo sofrido e a substituição	Tempo decorrido entre a marcação do golo pela equipa adversária e a realização da substituição	1

21	Diferença do tempo entre a substituição e a alteração do resultado	Tempo decorrido entre a realização da substituição e a marcação do golo (pela própria equipa ou pela equipa adversária) ou o final do jogo se o golo não fosse marcado	1
22	Alteração do resultado depois da substituição	Incluiu duas classes: Golo marcado depois da substituição (situação de sucesso); Golo não marcado ou golo sofrido depois da substituição (situação de insucesso)	1
23	Cartões amarelos	Se o jogador substituído tem cartão amarelo	2
24	Percentagem de tempo de posse de bola	Percentagem da duração que uma equipa tem a bola sem interrupções	1
25	Remates na baliza	Tentativas de golo que requerem a intervenção do guarda redes ou que resultam em golo)	1
26	Bolas recuperadas	Quando uma equipa recupera a bola do adversário	1
27	Resultado do jogo	Resultado final da partida	2
28	Diferença de pontuação	Vitória, empate ou derrota em função do saldo de golos entre as duas equipas	1
29	Classificação final na liga	Classificação final da equipa de 1 a 12	1
30	Primeira equipa a marcar	Utilização da variável <i>dummy</i> , onde 0 não pontua primeiro e 1 pontua primeiro	1
31	Superfície do campo	Se a equipa joga na mesma superfície (terra, relva ou relva artificial) (1) ou diferente do campo regular que utiliza (0)	1
32	Golos marcados	Número de golos marcados	1
33	Tipo de substituição	Neutra, defensiva e atacante com base na posição de jogo dos jogadores envolvidos na substituição	2
34	Variáveis táticas	Comprimento e largura da equipa, relação comprimento/largura, índice de alongamento. Todas as variáveis foram calculadas usando os 10	1

		jogadores, excluindo-se o guarda redes. Outra variável foi o controlo do espaço, utilizando diagramas de Voronoi, calculado e expresso como percentagem do espaço do campo controlado pela equipa que realiza a substituição. O espaço controlado foi obtido para 3 regiões: terço defensivo, terço médio e terço atacante	
35	Diferença de golos antes e depois de cada substituição	Diferença do resultado antes e depois de cada substituição	1
36	Tipo de competição	Qual a competição de Futebol a que as observações pertencem	1
37	Jogos rivais	Jogos entre equipas que estão com uma diferença pontual de ± 3 pontos	1
38	Jornada	Número da jornada do jogo em análise	1
39	Valor do 11 inicial	Valor do 11 inicial obtido através da média do valor dos defesas, médios e avançados respeitante à época 2014/2015	1

Associado ao jogo de Futebol, existe ainda o mercado das apostas desportivas que sofreu grandes mudanças nas últimas duas décadas. A introdução das apostas online permitiu aos apostadores não se terem de deslocar fisicamente para efetuar as suas apostas e comparar facilmente as várias probabilidades das diferentes casas de apostas existentes (Wikelmann *et al.*, 2020). Como consequência, e devido ao aumento da concorrência, as casas de apostas têm recorrido à análise de dados e à matemática para aumentarem a precisão das suas previsões e aumentar a lucratividade (Štrumbelj & Šikonja, 2010).

4 – Substituições e as suas conclusões

Selecionamos da análise anterior três estudos particularmente representativos e abrangentes do fenómeno das substituições em Futebol, que pela sua relevância apresenta-se de forma desenvolvida:

1. “A Proposed Decision Rule for the Timing of Soccer Substitutions” (Myers, 2011)

O estudo de Myers (2011) analisa o tempo das substituições recorrendo à técnica de *data mining*, árvores de decisão, para desenvolver uma regra de decisão que facilite a tarefa dos treinadores de Futebol.

Com base nas observações retiradas das substituições feitas nas três principais ligas do Futebol europeu foi realizada uma análise exploratória dos dados que inclui estatística descritiva e testes de hipótese.

Quanto à variável do tempo das substituições o autor conclui que:

1. A primeira substituição é feita em média ao minuto 57, sendo a sua distribuição enviesada pelas substituições que ocorrem na primeira parte, nas quais a maioria se deve a lesão.
2. A segunda substituição é em média feita aos 71 minutos de jogo, um pouco depois da metade da segunda parte.
3. A terceira substituição ocorre por volta dos 90 minutos.

Relativamente à variável do número de substituições utilizadas pelas equipas, ficou claro que a maioria das equipas utiliza as três substituições o que indica que valorizam o contributo dos suplentes no jogo.

Embora o tamanho da amostra seja pequeno, foram também analisados os desempenhos das equipas com o uso das três substituições *versus* duas substituições, concluindo que as equipas que utilizaram as três substituições parecem ter melhor desempenho do que aquelas que não o fazem, sobretudo quando se encontram a perder ou empatadas. Tal mostra que os jogadores substitutos podem ser valiosos na conquista da vitória da equipa.

Para obter mais informações sobre os dados recolhidos, foram realizados três testes de hipótese sobre as substituições e as variáveis em estudo, recorrendo à análise ANOVA.

O objetivo dos testes foi determinar quais os fatores que afetam os tempos de substituição em cada jogo.

No primeiro teste de hipótese o objetivo era perceber se o tempo de cada substituição varia consoante o resultado da equipa, isto é, se está a ganhar, a perder ou a empatar. Os resultados indicaram que os treinadores têm tendência a fazer as substituições mais tarde quando a equipa está a vencer e mais cedo quando se encontram empatados ou a perder. A conclusão retirada é que os treinadores valorizam mais os jogadores que iniciam a partida quando eles são capazes de produzir uma vantagem para a equipa, preferindo que estes joguem por mais tempo do que comprometer a *performance* e o fluxo da equipa ou de uma jogada com a mudança de jogadores.

No segundo teste de hipótese o objetivo era perceber se o tempo das substituições difere consoante a equipa jogue em casa ou fora. Os resultados indicam que não existe uma diferença significativa entre a primeira e a segunda substituição, mas que a terceira substituição foi realizada mais cedo nos jogos fora. A explicação poderá ser que os treinadores substituem mais cedo quando estão a perder e a equipa que joga fora tem maior probabilidade de estar nessa posição.

No último teste de hipótese pretendeu-se perceber se o tempo de substituição varia consoante o tipo de liga que se está a observar (inglesa, espanhola ou italiana). O resultado indica que existe uma diferença significativa na primeira substituição, mas não na segunda e na terceira. Esta diferença é sobretudo significativa entre a liga inglesa e a italiana, o que sugere que os treinadores em Itália são menos pacientes e crentes no seu onze inicial do que os treinadores das equipas inglesas.

Por fim, recorrendo ao *Data Mining*, foi proposta uma regra de decisão para os treinadores que consistia em:

1. Se a equipa está a perder, fazer a primeira substituição antes do minuto 58, a segunda antes do minuto 73 e a terceira antes do minuto 79.
2. Se está empatada ou a ganhar, fazer as substituições a qualquer tempo.
3. Esta regra não se aplica se: 1) existir uma substituição forçada por lesão na primeira parte, 2) qualquer equipa levar um cartão vermelho e 3) jogos que têm prolongamento se o encontro terminar empatado nos 90 minutos.

Esta regra de decisão mostrou, para um intervalo de confiança de 95%, uma eficácia de 38% a 47% e, quando não seguida, uma eficácia de 17% a 24%.

2. “Análise do contexto e efeito das substituições no Futebol de alto rendimento” (Russiano, 2016)

O estudo de Russiano corresponde a uma dissertação, desenvolvida em 2016 na Universidade de Lisboa – Faculdade de Motricidade Humana, que teve como objetivo analisar o contexto e o efeito das substituições.

Para tal, recorreu a uma amostra de 1584 substituições (556 primeiras, 548 segundas, e 480 terceiras), registadas em 279 jogos da fase de grupos da Liga dos Campeões (LC) das épocas 2012/13, 2013/14 e 2014/15. Os dados foram analisados recorrendo à estatística descritiva, ao teste do Qui-quadrado e à Regressão Logística Binomial.

O estudo foi realizado em duas etapas, sendo que, num primeiro momento caracterizou-se o contexto da solicitação das substituições (primeira, segunda e terceira) em função das variáveis contextuais (local do jogo, qualidade do adversário, período de jogo e resultado corrente). Já numa segunda fase foi feita a análise do efeito das substituições, realizadas num contexto de desvantagem mínima no marcador e na alteração do resultado corrente do jogo, considerando as variáveis temporais (tempo do golo sofrido, tempo da substituição, diferença do tempo entre o golo sofrido e a substituição, diferença do tempo entre a substituição e a alteração do resultado) e contextuais (local do jogo e qualidade do adversário).

Relativamente aos resultados obtidos neste estudo, verificou-se que:

1. A primeira substituição foi realizada em média ao minuto 60, a segunda ao minuto 72 e a terceira ao minuto 82.
2. Os treinadores realizaram as três substituições em 85,5% dos jogos analisados, em 12,5% dos jogos duas substituições e em 1,4% apenas foi feita uma substituição.
3. Nenhuma das variáveis contextuais (local do jogo e qualidade do adversário) mostraram associação significativa ($p \leq 0,001$) com a realização das substituições.
4. Existe uma associação significativa entre as interações das variáveis contextuais *local do jogo*resultado corrente, qualidade do adversário*período do jogo,*

*qualidade do adversário*resultado corrente, período do jogo*resultado corrente* e a frequência de realização das substituições.

5. A interação *local do jogo*resultado corrente* influenciou significativamente a realização da primeira, segunda e terceira substituições durante o jogo. Nos jogos em casa, quando as equipas se encontram a ganhar, os treinadores efetuam um maior número de substituições. Jogando fora, o maior número de substituições são realizadas quando as equipas se encontram a perder.
6. A interação *qualidade do adversário*período do jogo* teve efeito significativo apenas na frequência da primeira e terceira substituições realizadas. Quando jogam contra equipas mais fortes os treinadores realizam a primeira e a terceira substituições mais cedo, comparativamente com as situações em que defrontam adversários mais fracos ou do mesmo nível.
7. A interação *qualidade do adversário*resultado corrente* revelou-se significativa no número de todas as substituições realizadas (primeira, segunda e terceira). Jogando contra equipas mais fortes as três substituições são utilizadas com mais frequência quando a equipa está a perder, enquanto que quando defrontam equipas mais fracas as substituições são realizadas mais vezes quando a equipa está a ganhar.
8. A interação *período do jogo*resultado corrente* teve o efeito significativo na frequência de realização da primeira, segunda e terceira substituições. Quando a equipa estava a perder as substituições foram efetuadas mais cedo do que quando estavam a ganhar.
9. As variáveis “locais do jogo” e o “tempo da substituição” explicam o sucesso da equipa depois da realização da substituição em situação de desvantagem.
10. Nos jogos em casa depois da substituição realizada com desvantagem no marcador a probabilidade de sucesso relativo (obter o golo do empate ou não obter o golo do empate) aumenta (47,2%), comparativamente com as substituições realizadas nos jogos fora.
11. Nos jogos em casa, por cada minuto que a equipa passa em situação de desvantagem mínima no marcador sem realizar a substituição, a probabilidade de sucesso relativo (obter o golo do empate ou não obter o golo do empate) decresce (4,1%).

3. “What can coaches do? The relationship between substitution and results of professional football matches” (Flôres et al., 2019)

Este artigo, publicado em 2019 na Revista Brasileira de Futsal e Futebol, aborda a relação entre as substituições e o resultado de jogos de Futebol de campeonatos e taças nacionais.

Ao nível da amostra foram observados 258 jogos no total, 129 de taças e 129 de campeonatos que ocorreram durante a temporada 2016/17 em 16 países. No total foram feitas 737 substituições em campeonatos e 721 durante as taças nacionais.

Para efeitos estatísticos e de comparação foi utilizada a ANOVA. Os resultados mostraram que os jogadores que entraram na partida marcaram 10,76% dos golos, correspondendo a 26,19% dos golos marcados após a primeira substituição. Nas taças nacionais os substitutos marcaram 10% dos golos, representando 24,84% dos golos após a primeira substituição. As equipas que jogaram os campeonatos fizeram mais golos após a primeira substituição, após a terceira e após todas as substituições.

Para além disso, os dados mostraram que quando uma equipa joga em casa e vence, fazem uma melhor gestão das substituições, marcando mais golos com implicância direta dos substitutos.

Por fim, analisando cada substituição separadamente, concluíram que a primeira é mais importante quando a equipa joga em casa.

PARTE II – METODOLOGIA

5 – Caracterização da Amostra

A amostra deste estudo foi composta por um total de 192 jogos oficiais da Liga dos Campeões (LC), respeitante à fase de grupos da edição 2019/20 e 2020/21. Os dados foram recolhidos da plataforma de resultados de Futebol *SofaScore*, disponível no site www.sofascore.com.



Figura 1. Logotipo da Liga dos Campeões (UEFA, 2021)

As razões que estiveram na base da escolha desta Liga prenderam-se com o facto de ser uma competição internacional, onde estão presentes as melhores equipas europeias, os melhores treinadores e jogadores. Como tal, é uma competição que centra forte investimento na qualidade dos dados estatísticos para obter melhores desempenhos.

A decisão de a análise se centrar sobre a fase de grupos desta competição e não adicionalmente com as fases de eliminação tem que ver com o facto de estes serem exatamente os dois momentos em que os treinadores sabem desde o início da competição o número de substituições que têm ao seu dispor. Isto é, na fase de grupos da edição 2019/20, pré-covid, os treinadores sabiam que tinham a opção de realizar 3 substituições, enquanto na fase eliminatória dessa mesma edição, em pleno momento de pandemia, as regras foram alteradas e os treinadores já podiam realizar 5 substituições, em 3 momentos. Desta forma, não era possível a comparação entre a fase de eliminação das duas edições.

6 – Extração e registo dos dados

De forma a poder fazer um estudo completo e metódico sobre os dados recolhidos, foram criadas duas bases de dados em *Microsoft Excel*, uma para cada época estudada, com 40 variáveis. Os dados foram extraídos de forma manual e apenas foram seleccionadas algumas variáveis, com base na revisão de literatura das Tabelas 1 e 2 e as que se revelaram mais adequadas para o tema e propósito do trabalho.

Após a recolha dos dados procedeu-se ao seu tratamento através das ferramentas disponíveis no *Microsoft Excel* e na aplicação *PowerBI*, de forma a criar vários *dashboards* com informação que permita uma maior compreensão e análise visual.

7 – Variáveis

Conforme implícito anteriormente o registo de todas as variáveis fez-se posteriormente à realização dos jogos. Desta forma, as variáveis seleccionadas e estudadas foram:

1. **Local do jogo** – Indica se a equipa em causa joga em casa ou fora. Para classificar os jogos em função do local de jogo foi utilizada uma variável *dummy* com os seguintes valores: 1 – Casa; 2 – Fora;
2. **Resultado do jogo** – Designa se a equipa em estudo perdeu, empatou ou ganhou o jogo. A classificação usada foi: 2 – Vitória; 0 – Empate; 1 – Derrota;
3. **Número de substituições** – Representa o número total de substituições utilizadas pelos treinadores durante os jogos;
4. **Tempo da 1ª, 2ª, 3ª, 4ª e 5ª substituição** – 5 variáveis que indicam o tempo de jogo, em minutos, em que se realiza a 1ª, 2ª, 3ª, 4ª e 5ª substituição, respetivamente;
5. **Tipo de 1ª, 2ª, 3ª, 4ª e 5ª substituição** – 5 variáveis que representam se a substituição é de carácter neutro (0), defensiva (-1) e atacante (1) com base na posição de jogo dos jogadores envolvidos na substituição;
6. **Golos marcados** – Número total de golos marcados pela equipa observada numa partida;

7. **Tempo de golo** – Tempo, em minutos, em que ocorrem os golos marcados pela equipa observada (e.g., foram registados um máximo de 7 golos num jogo, pelo que existem 7 variáveis registadas);
8. **Tempo de golo depois da substituição** – Tempo, em minutos, do intervalo entre a realização da última substituição e o primeiro golo subsequente. Existem 5 variáveis registadas, uma para cada substituição;
9. **Lesões** – Número total de lesões por partida, da equipa observada;
10. **Jogadores que recuperaram para o jogo seguinte** – Número de jogadores lesionados no jogo observado da LC e que não foram convocados para o jogo seguinte da equipa em causa, independentemente da competição;
11. **Golos marcados** – Remates da equipa em análise que ultrapassam completamente a linha da baliza e que são considerados válidos;
12. **Golos sofridos** – Golos marcados pela equipa adversária;
13. **Falta** – Lance onde determinada irregularidade é marcada pelo árbitro;
14. **Passe** – Todas as formas de transferência da bola, dentro do campo, para um colega de equipa;
15. **Passes acertados** – Passes que chegam com sucesso ao destinatário;
16. **Cruzamentos** – Ação de pontapear a bola para uma zona ofensiva onde se encontra outro jogador da mesma equipa;
17. **Desarmes** – Movimento defensivo em que um jogador interseja a trajetória da bola, impedito que a equipa adversária continue a jogada;
18. **Duelos ganhos** – Disputa da bola entre jogadores ganha pela equipa em análise;
19. **Cartões amarelos** – Comunicação de uma advertência por uma infração considerada grave;
20. **Cartões vermelhos** – Comunicação de uma expulsão do jogo caso o jogador tenha cometido uma falta muito grave ou reincida em infração de gravidade semelhante à de um cartão amarelo;
21. **Percentagem de tempo de posse de bola** – Tempo que uma equipa possui com o controlo da bola;
22. **Defesas** – Interceções de remates efetuadas pelo guarda-redes que tinham como destino a baliza;

23. **Cantos** – É assinalado quando a bola ultrapassa completamente a linha de baliza, quer junto ao solo ou pelo ar, tocada em último lugar por um jogador da equipa defensora;
24. **Remates** – Tentativa de rematar à baliza de forma a marcar golo;
25. **Remates à baliza** – Remate que se dirige para o golo, isto é, remate que ou é defendido pelo guarda-redes ou outro jogador de campo ou que acaba em golo. Remates à barra ou aos postes não são considerados nesta variável.

8 – Procedimentos estatísticos

De forma a analisar e retirar conclusões dos dados recolhidos foi utilizada numa primeira fase a estatística descrita que permite descrever e resumir os dados de uma forma simples. As variáveis comparadas entre as duas edições da LC foram o número de substituições, o tempo de substituição, o tipo de substituição, o tempo de golo depois de cada substituição e o número de lesões. Para tal, foi utilizado o Excel como ferramenta que nos possibilita testar e apresentar os dados.

Numa outra fase, foram utilizados os testes estatísticos que permitem testar uma determinada assunção sobre uma característica da população. A regra é simples, se a probabilidade de ocorrer um determinado acontecimento for muito pequena, concluímos que a hipótese que o caracteriza não deve ser verdadeira (Abreu, 2007).

Os testes estatísticos podem ser divididos em dois conjuntos: os paramétricos e não-paramétricos. Os testes paramétricos baseiam-se num conjunto de pressupostos da distribuição dos dados, nomeadamente na distribuição normal, enquanto que os não paramétricos não são tão exigentes nos parâmetros da distribuição (Sedgwick, 2015).

Para esta análise utilizou-se o software IBM SPSS (versão 25) no sentido de aferir diferenças significativas nas duas edições da LC, para um nível de significância de 5%.

Os métodos e testes que pareceram mais adequados para analisar estes dados foram:

Regressão logística – É uma técnica que permite estimar a probabilidade de a variável dependente assumir um determinado valor em função dos valores conhecidos das outras variáveis testadas. Esta técnica é utilizada quando a variável dependente é de natureza binária ou multidimensional (Fernandes *et al.*, 2020). O objetivo da aplicação desta técnica é o de perceber quais as variáveis significativas na transição de uma edição para

a outra (e a sua força), assumindo-se como variável dependente a liga dos campeões, que se desdobra nas duas edições. Como variáveis independentes constam o número de lesões, cruzamentos, cortes, duelos ganhos, cartões amarelos, cartões vermelhos, posse de bola, defesas, cantos, remates, remates à baliza, *Index1* (número de substituições).

Teste de Mann Whitney – Pertence à categoria dos testes não paramétricos e é utilizado para apurar se as medianas de duas populações independentes são iguais. As duas amostras envolvidas não necessitam de ter o mesmo tamanho. Se o valor *p* observado for inferior ao valor de significância escolhido, aceita-se a hipótese da igualdade das medianas.

Nesta dissertação, o objetivo consiste em perceber se existe uma diferença significativa nos tempos das substituições nas duas edições da LC.

Effect Size (Dimensão do Efeito) – Em estatística é recorrentemente utilizada a análise da significância como um procedimento estatístico para avaliar a discrepância de uma hipótese estatística em relação aos resultados observados, utilizando uma medida padrão (valor-p) (Sirkin, 2005). Contudo, é importante avaliar o real significado, isto é, a importância prática dos resultados de eventuais diferenças encontradas entre duas ou mais médias ou variâncias (Fritz *et al.*, 2012).

A dimensão do efeito é uma medida quantitativa da magnitude do efeito experimental. Ou seja, não basta perceber que existe uma relação entre as variáveis, mas sim quão expressiva essa relação é. Quanto maior for o tamanho do efeito, mais forte é a relação entre as variáveis (Gignac & Szodorai, 2016).

Para tal, fez-se uso do teste de Cohen (*d*) que é utilizado quando o estudo abrange duas amostras independentes e se pretende perceber a diferença média padronizada entre dois grupos. Este teste é calculado através da seguinte expressão (Thalheimer & Cook, 2002):

$$\text{Effect Size (Teste d de Cohen)} = \frac{\bar{x}_1 - \bar{x}_2}{S}$$

Onde, \bar{x}_1 e \bar{x}_2 são a média de cada um dos grupos e *S* é o desvio padrão agrupado.

Neste teste é importante olhar primeiro para a dimensão do intervalo de confiança, sendo que, se o mesmo contiver o zero o efeito é nulo e quanto mais afastando do zero mais forte é o efeito. De seguida, deve-se avaliar o valor *d* seguindo o seguinte critério: efeito

reduzido ($d < 0,2$), efeito moderado ($0,2 \leq d \leq 0,8$) e elevado efeito ($d \geq 0,8$) (Clemente *et al.*, 2017).

Análise de sobrevivência – Seja T uma variável aleatória contínua que representa o tempo de substituição. A probabilidade do tempo para a substituição aumentar até um valor t é dado pela função de sobrevivência (Collett, 2003):

$$S(t) = \Pr(T \geq t) = \int_t^{\infty} f(x) dx = 1 - F(t)$$

Onde $f()$ e $F()$ são a função densidade e a função distribuição, respetivamente, funções de t . A probabilidade de existir uma substituição no tempo t , dado que ainda não houve, é definido como *hazard*:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log(S(t)))$$

O logaritmo da função de sobrevivência é dado por: $H(t) = -\ln S(t)$.

A regressão de Cox ou também chamada modelo dos *hazards* proporcionais é dada pela seguinte função: $h(t, x, \beta, h_0) = h_0(t) \varphi(x, \beta)$, onde $h_0(t)$ é o *hazard* base caracterizando a evolução do *hazard* em relação ao tempo assumindo $\varphi(x, \beta) = 1$; $\varphi(x, \beta)$ é um factor que depende das variáveis (X) e dos coeficientes (β) da equipa (como numa regressão ordinária) (Cox & Oakes, 1984).

A importância deste método relaciona-se com os tempos para a substituição, isto é, podem ser interpretados como os tempos para um determinado evento, tal como na análise de sobrevivência, poderemos assim analisar se há alguma variável que influencie este tempo (para mais ou para menos) atrasando ou acelerando o tempo para uma determinada ocorrência (substituição).

Coefficiente de variação (CV) – analisa a dispersão em termos relativos. Isto é, quanto menor for o valor do coeficiente de variação, mais homogêneos serão os dados, ou seja, menor será a dispersão em torno da média. O CV é calculado pelo coeficiente do desvio padrão com a média

PARTE III – ANÁLISE E DISCUSSÃO DOS RESULTADOS

9 – Introdução

A presente dissertação teve como finalidade analisar o impacto que a introdução das 5 substituições acarretam no Futebol de elite comparativamente com as tradicionais 3 substituições. Esta análise advém da comparação entre as variáveis escolhidas e retiradas em cada jogo.

As equipas analisadas foram praticamente as mesmas nas duas edições da fase de grupos da Liga dos Campeões, sendo que existem apenas 5 que não coincidem, de um universo de 24 equipas por edição. Contudo, a nível do país de origem das equipas, Portugal é o único país que apenas está presente numa edição, conforme a Figura 2 e 3 ilustram.

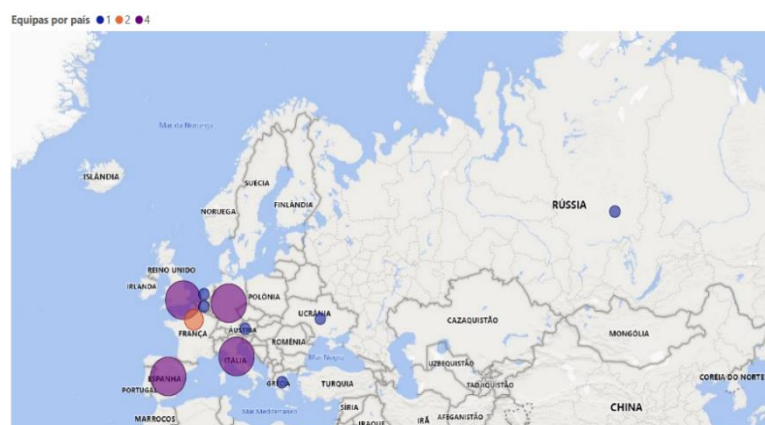


Figura 2. Distribuição geográfica das equipas por país edição LC 2019/20 (elaboração própria)

Dos onze países com equipas a competir, quatro países tiveram 4 equipas, sendo que a maioria dos presentes foram representados apenas por uma equipa.

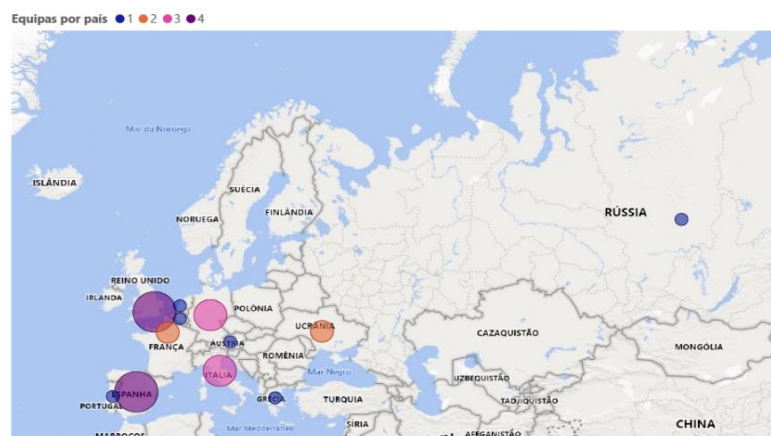


Figura 3. Distribuição geográfica das equipas por país edição LC 2020/21 (elaboração própria)

Na edição de 2020/21 participaram equipas de 12 países, sendo que apenas a Espanha e o Reino Unido tiveram 4 equipas a competir.

10 – Estatística descritiva

10.1 – ANÁLISE DESCRITIVA DAS VARIÁVEIS LOCAL DE JOGO E RESULTADO

Monitorizando as variáveis local de jogo e resultado foi possível analisar:

Tabela 3. Análise do valor percentual das variáveis local de jogo e resultado nas duas edições da LC (LC 2019/20 e 2020/21) (elaboração própria)

		LC 2019/20	LC 2020/21
Local de Jogo	Casa (1)	50%	50%
	Fora (2)	50%	50%
Resultado	Derrota (1)	14,58%	17,71%
	Empate (0)	22,92%	20,83%
	Vitória (2)	62,50%	61,46%

Conforme a Tabela 3, ao nível da variável local de jogo ambas as edições da LC estão distribuídas na mesma proporção de jogos casa e jogos fora, pelo que nos 96 jogos analisados em cada temporada 48 dizem respeito a jogos jogados em casa e a mesma quantidade de jogos fora.

No que diz respeito à variável resultado, observa-se que nas duas edições existe uma homogeneidade dos resultados, pelo que se verifica praticamente a mesma proporção de vitórias, empates e derrotas e que em cada edição da LC existe uma predominância muito acentuada das vitórias.

Outro output gerado no *IMB SPSS* consistiu na relação entre as variáveis resultado final e local de jogo:

Tabela 4. Análise do valor percentual da relação entre as variáveis local de jogo e resultado nas duas edições da LC (LC 2019/20 e 2020/21) (elaboração própria)

	LC 2019/20		LC 2020/21	
	Casa (1)	Fora (2)	Casa (1)	Fora (2)
Derrota (1)	10,42%	18,75%	16,33%	19,15%
Empate (0)	25,00%	20,83%	22,45%	19,15%
Vitória (2)	64,58%	60,42%	61,22%	61,70%

Na Tabela 4 podemos observar que parece existir uma certa coerência no que diz respeito à percentagem de vitórias, empates e derrotas em função da equipa em análise jogar em casa ou fora entre as duas edições da LC.

10.2 – ANÁLISE DESCRITIVA DAS VARIÁVEIS DE JOGO NAS DUAS EDIÇÕES DA LC (2019/20 E 2020/21)

As variáveis de jogo são relevantes para compreender o comportamento técnico-tático de uma equipa de Futebol. Analisando a estatística descritiva das variáveis de jogo, já definidas no ponto 7, foi feita uma comparação entre os resultados das duas edições.

Tabela 5. Comparação dos valores da estatística descritiva das variáveis de jogo nas duas edições da LC (2019/20 e 2020/21) (elaboração própria)

Variáveis de jogo	LC 2019/20	LC 2020/21
Golos marcados	207	191
Golos sofridos	104	98
Nº de faltas	1203	1216
Nº de passes	55372	54517
Nº de passes acertados	47562	46283
Cruzamentos	411	381
Cortes	1356	1404
Duelos ganhos	5432	4882
Cartões amarelos	181	180
Cartões vermelhos	10	9

% Posse de bola	57,97%	55,77%
Defesas	257	258
Cantos	551	491
Remates	1538	1301
Remates à baliza	587	525

Através da Tabela 5 é possível observar que a nível absoluto e percentual, no que diz respeito à posse de bola, não existem diferenças significativas entre os vários valores apresentados, excetuando as variáveis duelos ganhos, cantos, remates e remates à baliza que evidenciarem ter um maior valor na edição da liga dos campeões da época 2019/20. Isto parece mostrar que apesar da diferença da possibilidade de efetuar mais duas substituições nesta presente época, as variáveis de jogo são muito homogêneas.

10.3 – ANÁLISE DESCRITIVA DO NÚMERO E TEMPO DE REALIZAÇÃO DAS SUBSTITUIÇÕES NAS DUAS EDIÇÕES DA LC (2019/20 E 2020/21)

A análise do número médio de substituições nos jogos observados, demonstrou:

Tabela 6. Análise do valor absoluto, médio e coeficiente de variação de cada substituição nas duas competições (LC 2019/20 e 2020/21) (elaboração própria)

Substituição	N – LC 2019/20	Tempo \bar{X} – LC 2019/20	CV – LC 2019/20	N – LC 2020/21	Tempo \bar{X} – LC 2020/21	CV – LC 2019/21
1ª	96	57,13	26,83%	96	59,58	24,37%
2ª	96	71,63	13,58%	95	67,4	15,98%
3ª	90	82,64	8,91%	90	74,99	13,16%
4ª	-	-	-	70	79,84	11,55%
5ª	-	-	-	46	83,83	7,01%

Na fase de grupos da LC da edição 2019/20, em que apenas foi permitido realizar 3 substituições, conforme a Tabela 6, observa-se que na maioria dos jogos (93,75%), os treinadores utilizam todas as substituições, um valor ligeiramente superior ao registrado no trabalho de Russiano (cerca de 85,5%). No que diz respeito à fase de grupos da LC 2020/21, a percentagem de realização de três primeiras substituições é praticamente a

mesma que na edição anterior, sendo que a quarta substituição é realizada pelos treinadores em 72,93% dos jogos e a quinta em apenas 46,88%. Desta forma, no que diz respeito à frequência de utilização das substituições pode-se concluir que com a possibilidade de realizar 5 substituições, apenas em menos de metade dos jogos é que tal foi usado. Parece ser mais relevante para os treinadores a utilização de uma substituição adicional, isto é, a quarta substituição.

Relativamente ao tempo das substituições, na época 2019/20 a primeira substituição foi realizada em média aos 58 minutos, a segunda aos 72 e a terceira aos 83. Estes resultados são praticamente coincidentes com os observados por Russiano (2016), na sua dissertação e exposto no capítulo 4.1 deste trabalho. Observando a edição deste presente ano, a primeira substituição foi em média aos 60 minutos, a segunda aos 68, a terceira aos 75, a quarta aos 80 e a quinta aos 84.

Analisando o CV percebemos que em ambas as edições, os CVs da primeira substituição aproximam-se, 26,83% em 2019/20 e 24,37% em 2020/21, pelo que parece existir uma dispersão algo acentuada, o que indica que não existe um tempo estratégico claro definido pelos treinadores para realizar a primeira substituição. Em relação às restantes parecem ter uma dispersão mais baixa, sendo essas substituições efetuadas num tempo mais fixo por todos os treinadores. Outro padrão observado, é que à medida que o número de substituições aumenta o CV diminui, o que é justificado pelo menor tempo de jogo que resta para a partida terminar e o número de substituições que faltam efetuar.

Em ambas as edições, na primeira parte praticamente não ocorrem substituições, 13 na primeira edição e 9 na segunda, num universo de 679 substituições observadas, a maioria delas devido a lesão e de carácter neutro, e o tempo da terceira substituição da LC da época 2019/20 praticamente coincide com o tempo médio da quinta e última substituição da época 2020/21. Mais podemos constatar observando este indicador e o tempo médio de cada substituição que a possibilidade de ter cinco substituições não retarda o tempo em que a última é executada. Isto significa, que existe uma maior oportunidade para alterar a dinâmica do jogo, via substituições de jogadores e, assim poder renovar a equipa com jogadores sem níveis de fadiga física acentuada.

10.4 – ANÁLISE DESCRITIVA DO TIPO DE SUBSTITUIÇÕES

Para classificar o tipo de substituições foi atribuída uma variável multinomial, em que o 0 representa uma substituição de carácter neutro, 1 de carácter defensivo e 2 de carácter ofensivo. De forma a clarificar melhor o tipo de análise utilizada, uma substituição neutra significa que ocorreu por motivo de lesão ou existe troca entre jogadores da mesma posição, a substituição de carácter defensivo ocorre quando o jogador que entra tem uma posição mais recuada no campo do que o jogador que sai e uma substituição atacante traduz uma alteração de um jogador com uma posição mais ofensiva, permitindo chegar mais rapidamente ao golo, do que o jogador que sai.

A análise do tipo de substituições está expressa na Tabela 7.

Tabela 7. Valores absolutos e percentuais de substituições neutras, ofensivas e defensivas nas duas competições (LC 2019/20 e 2020/21) (elaboração própria)

Substituição	Tipo de substituição					
	Neutra (0)		Defensiva (1)		Atacante (2)	
	LC 2019-2020	LC 2020-2021	LC 2019-2020	LC 2020-2021	LC 2019-2020	LC 2020-2021
1ª	82 (36,94%)	82 (24,05%)	4 (15,38%)	5 (17,24%)	4 (11,76%)	9 (33,33%)
2ª	74 (33,33%)	86 (25,22%)	8 (30,77%)	3 (10,34%)	8 (23,53%)	6 (22,22%)
3ª	66 (29,73%)	76 (22,29%)	14 (53,85%)	7 (24,14%)	14 (41,18%)	7 (25,93%)
4ª	-	60 (17,60%)	-	6 (20,69%)	-	4 (14,81%)
5ª	-	37 (10,85%)	-	8 (27,59%)	-	1 (3,70%)
Total	222 (78,72%)	341 (89,89%)	26 (9,22%)	29 (7,31%)	34 (12,06%)	27 (6,80%)

Pela leitura da Tabela 7 podemos constatar que em ambas as edições a maioria das substituições tiveram um carácter neutro, na LC de 2019/20 78,72% das substituições foram neutras, sendo que em 2020/21 houve um aumento e a percentagem passou para 89,89. Parece existir uma crescente tendência para as substituições de carácter neutro com o aumento da possibilidade de realizar mais substituições, sendo que se procura mais substituir os jogadores das mesmas posições do que fazer alterações mais táticas que previsivelmente acarretam riscos mais elevados.

10.5 – ANÁLISE DESCRITIVA DOS GOLOS APÓS AS SUBSTITUIÇÕES

Um indicador de análise importante é a capacidade que a equipa tem em marcar depois de ser realizada uma substituição (Tabela 8). Desta forma, foi monitorizado a diferença entre o tempo do primeiro golo após a realização de cada substituição e o tempo dessa substituição. Por exemplo, se uma equipa faz uma substituição aos 60 minutos e marca aos 65 minutos a diferença entre o tempo que ocorreu a substituição e o golo foram de 5 minutos.

Tabela 8. Análise do golo após a substituição (elaboração própria)

Substituição	LC 2019/20		LC 2020/21	
	Golo após substituição			
	N	\bar{X} do tempo (em minutos)	N	\bar{X} do tempo (em minutos)
1 ^a	28	8,36	7	9
2 ^a	24	8,25	24	7,54
3 ^a	23	9,09	14	5,43
4 ^a	-	-	10	8,2
5 ^a	-	-	12	7,67

Através da análise da tabela 8 podemos constatar que na temporada 2019/20 todas as substituições parecem ter a mesma influência, não só no que toca ao número de vezes em que é marcado um golo, mas também ao tempo em que se demora a marcar esse golo, uma vez que são muito coincidentes. Já na temporada 2020/21, existe uma clara tendência para se marcar um golo após a segunda substituição face às outras substituições.

10.6 – ANÁLISE DESCRITIVA DO NÚMERO DE LESÕES

As lesões são um problema muito comum em jogadores profissionais de Futebol de alta competição. Conforme já referido nesta dissertação, comparativamente com outros jogos desportivos coletivos, sobretudo por causa das regras impostas no Futebol, existe uma exigência física muito elevada e superior aos outros desportos, o que acaba por levar a um maior número de lesões (Mota *et al.*, 2020). O estudo citado afirma que mesmo com a mudança de três para cinco substituições, as diferenças de exigência física continuam a ser extremamente grandes.

Através da monitorização do número de lesões das equipas observadas por jogo, foi possível observar os valores da Tabela 9:

Tabela 9. Análise do número de lesões (elaboração própria)

	LC 2019/20	LC 2020/21
Nº lesões	32	20
Nº jogadores que recuperaram para o jogo seguinte	34,38%	25%

Na época 2019/20 ocorreram mais 12 lesões do que na época 2020/21, isto é, mais 60% de lesões. Isto evidencia que o aumento da possibilidade de realizar mais substituições contribui para refrescar a equipa e prevenir a ocorrência de lesões.

Contudo, a percentagem de jogadores que se lesionaram e estiveram disponíveis para o jogo seguinte, isto é, convocados, é efetivamente mais baixa nesta presente época, 25%, em comparação com a edição do ano anterior, que apresenta 34,38%. Isto permite observar que apesar do número de lesões ter diminuído o impacto e a gravidade das lesões parece ter sido maior.

11 – Estatística inferencial

11.1 – REGRESSÃO LOGÍSTICA

A análise de Regressão Logística está expressa no output dado pelo *IMB SPSS* (Tabela 10).

Tabela 10. Output regressão logística SPSS (elaboração própria)

Variáveis independentes	B	S.E.	Wald	df	Valor p	Exp(B)
Nº de Lesões	,495	,170	8,456	1	,004	1,641
Cruzamentos	-,062	,038	2,591	1	,107	,940
Cortes	,011	,014	,617	1	,432	1,011
Duelos Ganhos	,052	,009	34,824	1	≥,001	1,054
Cartões Amarelos	,033	,072	,213	1	,644	1,034

Cartões Vermelhos	,377	,283	1,778	1	,182	1,458
Posse de bola	,018	,014	1,792	1	,181	1,018
Defesas	,042	,049	,750	1	,386	1,043
Cantos	,004	,037	,014	1	,905	1,004
Remates	,092	,027	11,465	1	,001	1,097
Remates à baliza	-,040	,045	,778	1	,378	,961
Index1	-,689	,076	82,365	1	≥,001	,502
Constante	-4,069	,990	16,881	1	≥,001	,017

Através da Tabela 10, e observando o valor $p < 0,05$, é possível perceber que as variáveis significantes entre as duas edições da LC foram as variáveis número de lesões, duelos ganhos, remates e número de substituições (*Index1*).

11.2 – TESTE U DE MANN-WHITNEY

Através do valor p expresso na Tabela 11, conclui-se que na segunda e terceira substituição existem diferenças significativas no tempo de realização das substituições nas duas edições da LC, ao invés da primeira substituição.

Tabela 11. Output teste Mann-Whitney SPSS (elaboração própria)

	Tempo 1ª substituição	Tempo 2ª substituição	Tempo 3ª substituição
U de Mann-Whitney	4175	3414	2150,5
Valor P	0,259	0,003	≥0,001

11.3 – EFFECT SIZE (TESTE d DE COHEN)

A estatística d (Tabela 12) mostra que o efeito entre os tempos da segunda substituição é moderado e alto entre os tempos da terceira substituição para ambos os grupos.

Tabela 12. Output teste d de Cohen SPSS (elaboração própria)

Substituição	Valor d de Cohen	Intervalo de confiança de 95%	
		Inferior	Superior
1ª (ambas edições)	0,165	-0,119	0,448
2ª (ambas edições)	-0,412	-0,698	-0,125
3ª (ambas edições)	-0,880	-1,185	-0,572
3ª (primeira edição) e 5ª (segunda edição)	0,171	-0,185	0,527

11.4 – ANÁLISE DE SOBREVIVÊNCIA

As Figuras 4 e 5 correspondem aos outputs relativos à análise de sobrevivência.

Pela análise da Figura 4 (as linhas cruzam-se) e através do valor p igual a 0,287 verifica-se que não existem diferenças entre os tempos da primeira substituição para as duas edições.

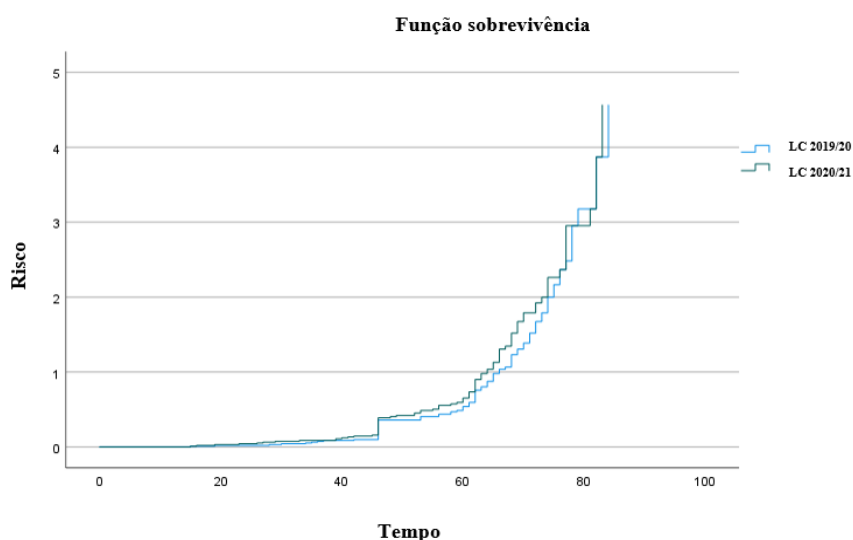


Figura 4. Análise de sobrevivência: curvas do tempo da 1ª substituição para ambas as edições da LC

Pela análise da Figura 5 (as linhas cruzam-se) e como o valor p é igual a 0,556 verifica-se que não existem diferenças entre os tempos das últimas substituições de cada uma das edições.

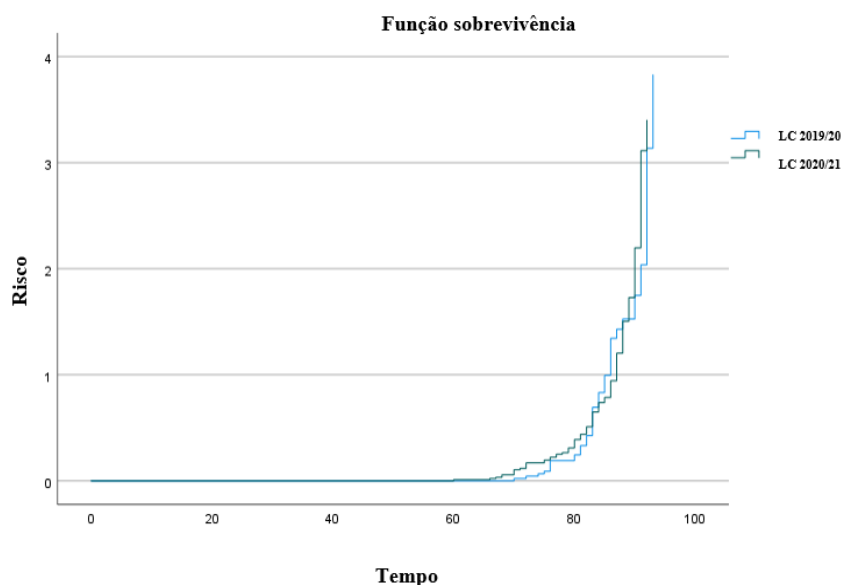


Figura 5. Análise de sobrevivência: curvas do tempo da 3ª substituição e da 5ª substituição para ambas as edições da LC

Por fim, a regressão de Cox, cujo output está representado na Tabela 13, demonstra que não existem diferenças significativas ao nível das variáveis independentes das duas edições da LC. Isto é, as substituições não parecem depender de determinantes do próprio jogo mas da manutenção da tática do jogo.

Tabela 13. Regressão de Cox

Variáveis independentes	B	SE	Wald	gl	Valor p	Exp(B)
Golos sofridos	,113	,098	1,326	1	,250	1,119
Nº de faltas	,048	,028	3,100	1	,078	1,050
Nº de passes	-,002	,003	,265	1	,607	,998
Nº de passes acertados	,004	,003	2,130	1	,144	1,004
Nº de lesões	-,167	,161	1,076	1	,300	,846
Cruzamentos	,055	,044	1,573	1	,210	1,056
Cortes	,020	,016	1,595	1	,207	1,020
Duelos ganhos	-,019	,012	2,591	1	,108	,982
Cartões amarelos	,071	,089	,639	1	,424	1,074
Cartões vermelhos	-,441	,363	1,478	1	,224	,643

Substituições no jogo de Futebol – Análise de dados em época de competição sem e com COVID-19

Porcentagem de posse de bola	,006	,024	,075	1	,784	1,007
Defesas	-,065	,053	1,531	1	,216	,937
Cantos	-,067	,042	2,551	1	,110	,935
Remates	,007	,032	,046	1	,830	1,007
Remates à baliza	,042	,055	,582	1	,445	1,043

CONCLUSÃO

A realização deste trabalho tinha como objetivo compreender os impactos da possibilidade das equipas de Futebol efetuarem cinco substituições, em consequência das alterações provenientes da pandemia COVID-19, ao invés das tradicionais três substituições.

Após uma revisão de literatura foi possível perceber os trabalhos realizados nesta área, os seus principais propósitos, as técnicas utilizadas e as variáveis mais estudadas.

No capítulo da introdução desta dissertação, foram levantadas várias hipóteses. Assim, através da aplicação da estatística descritiva e inferencial foi possível concluir que existe uma alteração significativa na passagem de uma época para a outra nas variáveis número de lesões, duelos ganhos, remates e número de substituições efetuadas.

De realçar que a pandemia do COVID-19 provocou um confinamento com uma duração de cerca de dois meses (março a maio de 2020), o que obrigou a recalendarizar as provas desportivas e a que os jogadores tivessem de realizar mais jogos num menor período de tempo, no caso, no final da época desportiva 2019/20.

Os resultados comprovam que existe um menor número de lesões com a introdução das 5 substituições passíveis de serem usadas num jogo. Também parece existir uma quebra de rendimento geral na *performance* dos jogadores e da equipa, sobretudo com as variáveis duelos ganhos e remates que decresceram significativamente da temporada 2019/20 para a 2020/21.

Mais, se pode enaltecer, ainda que sem dados estatísticos evidentes, que poderá existir uma maior rotatividade dos jogadores da equipa com o aumento do número de substituições, dando possibilidade a jogadores menos utilizados de jogarem mais vezes. Tal poderá manifestar-se no aumento motivacional de mais jogadores no quadro de competitividade da equipa.

No que diz respeito ao tempo das substituições conclui-se que na segunda e terceira substituição existem diferenças significativas no tempo de realização das substituições nas duas edições da LC, ao invés da primeira substituição. Mais se observa que os treinadores parecem privilegiar mais a realização de quatro substituições e que a possibilidade de dispor de 5 substituições não retarda o tempo em que a última é

executada. Isto significa, que existe uma maior oportunidade de alterar a dinâmica do jogo e renovar os jogadores que eventualmente apresentem níveis de fadiga mais elevados e ou níveis de rendimento mais baixo.

Por fim, é evidente que a primeira e a última substituição de cada edição ocorrem praticamente ao mesmo tempo, já nas do meio não existe um tempo definido. Este facto parece mostrar que os treinadores não mudam a estratégia de jogo de uma edição para a outra, optando por um *continuum* de opção tática, mesmo quando dispõem de mais possibilidade de substituição de jogadores.

O jogo de Futebol, sobretudo na LC, é um desporto de elite que capta a atenção de milhões de espetadores por todo o mundo. É de facto fascinante a forma como as pessoas vivem este desporto, fruto de um fenómeno social e cultural, que não cabe neste trabalho aprofundar, sendo que no Futebol tudo acontece muito rápido e o que hoje é certo e elogiado, em pouquíssimo tempo pode tornar-se errado e como tal criticado.

É um negócio, cuja gestão deve considerar variáveis racionais, mas que se centra numa atividade coletiva, aberta, com interferência do adversário e com elevados níveis de variabilidade, alternando entre o sucesso e o insucesso dos jogadores e ou das equipas. Todos os dados possíveis de considerar para gerir uma equipa são relevantes, quer para identificar padrões coletivos da equipa, quer para explorar padrões de comportamento ou assinaturas de *performance* individual, muitas vezes recorrendo a métodos matemáticos para o efeito considerados (Martins *et al.*, 2021).

Com a quantidade de dados gerados nesta área acreditamos que este trabalho seja um ponto de partida e de continuidade para investigações futuras e de reflexão para todos os intervenientes associados ao jogo de Futebol, para que melhor possam identificar, compreender e avaliar quais as medidas que se podem tomar e as variáveis que se podem mudar para aperfeiçoar e potenciar a qualidade do jogo e o nível de rendimento desportivo que se espera do mesmo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abreu, A. M. (2007). *Testes de Hipóteses Estatísticas Resenha Hipótese nula e hipótese alternativa*. Faculdade de Ciências Exatas e da Engenharia da Universidade da Madeira.
- Aguiar, C. (2020^a, December 3). A ciência dos dados está a ser decisiva para gerir a pandemia - DN. *Diário de Notícias*.
- Aguiar, N. (2018). *Visão | A matemática já nos ensina a vencer no Futebol*.
- Aguiar, N. (2020^b, July 27). *Visão | A Covid-19 também mudou como o Futebol é jogado?*
- Bradley, P. S., Lago-Peñas, C., & Rey, E. (2014). Evaluation of the match performances of substitution players in elite soccer. *International Journal of Sports Physiology and Performance*, 9(3), 415–424. <https://doi.org/10.1123/IJSP.2013-0304>
- Caballero, P., García, J., & Ibáñez, S. (2017). Influence of situational variables on the U'18 soccer performance analysis. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, 32, 224–227.
- Carling, C., Espié, V., Le Gall, F., Bloomfield, J., & Jullien, H. (2010). Work-rate of substitutes in elite soccer: A preliminary study. *Journal of Science and Medicine in Sport*, 13(2), 253–255. <https://doi.org/10.1016/j.jsams.2009.02.012>
- Carvalho, R. D., Bueno UTP, M., & Alves Neto UTP, W. (2003). Pré-processamento de dados em aprendizado de máquina supervisionado. Doctoral Thesis, Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos. doi:10.11606/T.55.2003.tde-06102003-160219. Retrieved 2021-09-27, from www.teses.usp.br
- Clemente, F. M., Martins, F. M. L., Nikolaidis, P. T., & Mendes, R. S. (2017). Association between physical activity patterns and anthropometric characteristics of adults: An issue of public health? *Biomedical Human Kinetics*, 9(1), 124–132. <https://doi.org/10.1515/bhk-2017-0018>
- Coelho, D. B., Coelho, L. G. M., Morandi, R. F., Junior, J. B. F., Marins, J. C. B., Prado, L. S., Soares, D. D., & Silami-Garcia, E. (2012). O efeito das substituições realizadas no segundo tempo da partida na intensidade de jogo de futebol. *Revista Brasileira de*

- Cineantropometria e Desempenho Humano*, 14(2), 183–191.
<https://doi.org/10.5007/1980-0037.2012v14n2p183>
- Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman and Hall / CRC.
- Cox, D.R., & Oakes, D. (1984). *Analysis of Survival Data* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315137438>
- Direção Geral da Saúde. (2020). *Perguntas Frequentes Categoria - COVID-19*. <https://covid19.min-saude.pt/category/perguntas-frequentes/?t=como-se-transmite-2#como-se-transmite-2>
- Emmanuel, I., & Stanier, C. (2016). Defining Big Data. *American University, Bulgaria*. <https://doi.org/10.1145/12345.67890>
- Fernandes, A., Filho, D., Rocha, E., & Nascimento, W. (2020, May 16). Leia este artigo se você quiser aprender regressão logística | Fernandes | Revista de Sociologia e Política. *Revista de Sociologia e Política*.
- Flôres, F. S., Santos, D. L. dos, Carlson, G. R., & Gelain, E. Z. (2019). What can coaches do? The relationship between substitution and results of professional soccer matches. *RBFF-Revista Brasileira de Futsal e Futebol*, 215–222.
- Frankenfield, J. (2020, July 1). *Data Analytics Definition*. <https://www.investopedia.com/terms/d/data-analytics.asp>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Gomez, M. A., Lago-Peñas, C., & Owen, A. L. (2016). The influence of substitutions on elite soccer teams' performance. *International Journal of Performance Analysis in Sport*, 16(2), 553–568. <https://doi.org/10.1080/24748668.2016.11868908>
- Grant, A. (2020, January 3). *What Is Data Analysis and Why Is It Important?*.

<https://www.makeuseof.com/tag/what-is-data-analysis/>

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.

Hills, S. P., Barwood, M. J., Radcliffe, J. N., Cooke, C. B., Kilduff, L. P., Cook, C. J., & Russell, M. (2018). Profiling the Responses of Soccer Substitutes: A Review of Current Literature. In *Sports Medicine* (Vol. 48, Issue 10, pp. 2255–2269). Springer International Publishing. <https://doi.org/10.1007/s40279-018-0962-9>

Hinshelwood, S. (2018, November 12). *5 Reasons Why Data Analysis is Important for Every Business*. <https://businesspartnermagazine.com/5-reasons-why-data-analysis-is-important-for-every-business/>

Laja, R. (2021, February 2). *Pandemia provoca queda histórica de 7,6% do PIB em 2020 / TVI24*.

Lorenzo-Martínez, M., Rein, R., Garnica-Caparrós, M., Memmert, D., & Rey, E. (2020). The Effect of Substitutions on Team Tactical Behavior in Professional Soccer. *Research Quarterly for Exercise and Sport*, 1–9. <https://doi.org/10.1080/02701367.2020.1828563>

Martins, F., Gomes, R., Lopes, V., Silva, F., & Mendes, R. (2021). *Mathematical models to measure the variability of nodes and networks in team sports*. *Entropy*, 23, 1072. <https://doi.org/10.3390/e23081072>

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.

Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), 213–222. <https://doi.org/10.1007/s41060-017-0093-7>

Mota, G. R., Santos, I. A. dos, Arriel, R. A., & Marocolo, M. (2020). Is It High Time to Increase Elite Soccer Substitutions Permanently? *International Journal of Environmental Research and Public Health*, 17(19), 7008. <https://doi.org/10.3390/ijerph17197008>

Myers, B. R. (2011). A proposed decision rule for the timing of soccer substitutions. *Journal of Quantitative Analysis in Sports*, 8(1). <https://doi.org/10.1515/1559->

0410.1349

Schumaker, R., K. Solieman, O., & Chen, H. (2010). *Sports Data Mining* (S. Vob & R. Sharda (eds.); SPRINGER).

Russiano, P. (2016). *Análise do contexto e efeito das substituições no Futebol de alto rendimento*. Faculdade de Motricidade Humana da Universidade de Lisboa. <http://hdl.handle.net/10400.5/12732>

Sagioglu, S., & Sinanc, D. (2013). Big data: A review. *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, 42–47. <https://doi.org/10.1109/CTS.2013.6567202>

Sedgwick, P. (2015, April 17). A comparison of parametric and non-parametric statistical tests. *BMJ*, 350(apr17 1), h2053–h2053. <https://doi.org/10.1136/bmj.h2053>

Sirkin, R. M. (2005). Two-sample t tests. *Statistics for the Social Sciences 3rd Ed.*, 271–316. <https://doi.org/10.1038/nmeth.2698>

Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, 26(3), 482–488. <https://doi.org/10.1016/j.ijforecast.2009.10.005>

Thalheimer, W., & Cook, S. (2002). *How to calculate effect sizes from published research: A simplified methodology*.

UEFA. (2021). *UEFA Champions League*. <https://pt.uefa.com/uefachampionsleague/>

Varela-Quintana, C., Cuervo, J., & Prieto-Rodríguez, J. (2016). The effect of an additional substitution in association football. Evidence from the Italian Serie A. *Revista de Psicología del Deporte. Revista de Psicología Del Deporte*, 25(1), 101–105.

Weil, E., Rollin, J., Joy, B., Alegi, P., & Giulianotti, R. (2020, August 26). *football / History, Rules, & Significant Players | Britannica*. Encyclopedia Britannica. <https://www.britannica.com/sports/football-soccer>

Wikelmann, D., Oetting, M., & Deutscher, C. (2020). Short and Long-Term Biases in European Football Betting Markets. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3672233>

Wu, Y., Xia, Z., Wu, T., Yi, Q., Yu, R., & Wang, J. (2020). Characteristics and optimization of core local network: Big data analysis of football matches. *Chaos, Solitons and Fractals*, 138, 110136. <https://doi.org/10.1016/j.chaos.2020.110136>