UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS - RIO CLARO

Educação Física

Victor Valvano Schimidt

Quais ações são significativas para uma vitória no futebol? Uma análise por regressão logística

Rio Claro – SP

2021

Victor Valvano Schimidt

Quais ações são significativas para uma vitória no futebol? Uma análise por regressão logística

Orientador: Prof. Dr. Fabio Augusto Barbieri

Coorientador: Prof. Ms. Luiz Henrique Palucci Vieira

Supervisor: Prof. Dr. José Ângelo Barela

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências da Universidade Estadual Paulista "Júlio de Mesquita Filho" - Câmpus de Rio Claro, para obtenção do grau de Licenciado em Educação Física.

Rio Claro – SP

2021

S335q

Schimidt, Victor Valvano

Quais ações são significativas para uma vitória no futebol? Uma análise por regressão logística / Victor Valvano Schimidt. -- Rio Claro, 2021

43 p.

Trabalho de conclusão de curso (Licenciatura - Educação Física) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Rio Claro

Orientador: Fabio Augusto Barbieri

Coorientador: Luiz Henrique Palucci Vieira

1. Futebol. 2. Desempenho. 3. Análise de regressão. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Agradecimentos

- Agradeço primeiramente meu pai, minha mãe, por sempre me apoiarem e confiarem em mim em tudo que eu fosse fazer. Amo muito vocês.

- À minha irmã, que muito me ensinou e ensina, que sempre me ajuda quando eu peço e sempre se faz presente. Obrigado por tudo.

- Ao meu coorientador Luiz, por toda a ajuda e disponibilidade para ensinar.

- Ao meu orientador Fabinho, por me aceitar nesse projeto e me ajudar sempre que precisei.

- Ao pessoal da CC de São Carlos, por todo ensinamento no período em que morei lá. Em especial ao Space e ao Ed, por toda disposição de sempre me ajudar e se manterem presentes nesse momento.

- À República Catorze e todos os moradores que por aqui passaram, sem vocês essa fase da vida teria sido muito mais difícil. Um obrigado especial ao Lasa e ao Buxa, meus irmãos que Rio Claro proporcionou, pelos vários momentos em que pudemos conversar sobre tudo e estarmos sempre nos ajudando em tudo.

- Aos Freaks, meus amigos de infância que já são família a um bom tempo.

Resumo

O futebol é um esporte que envolve uma gama de variáveis físicas, técnicas e táticas. Apesar do aumento de estudos sobre a importância das variáveis técnicas no desempenho do jogo nos últimos anos, ainda o estudo das variáveis físicas tem sobressaído. Além disso, a maioria dos estudos tem utilizado análises dos dados mais simples, que muitas vezes não conseguem responder precisamente algumas questões relacionadas ao desempenho esportivo. Análises estatísticas mais aprofundadas são importantes para melhor explorar as variáveis de desempenho e apresentar uma análise mais objetiva do jogo. Desta forma, o presente estudo teve a seguinte questão a ser respondida: é possível encontrar um modelo de características técnicas individuais de acordo com as posições de jogo que são determinantes para vencer no futebol? Assim, o objetivo deste trabalho foi analisar a influência das ações técnicas individuais de cada posição de jogo sobre o resultado da partida através de um modelo de regressão logística. Foram analisados 5396 jogadores das 10 principais ligas nacionais (Alemanha, Argentina, Brasil, Colômbia, Espanha, França, Holanda, Inglaterra, Itália e Portugal) nas partidas em que atuaram entre os anos de 2015 até 2019. Os indicadores foram obtidos através do site da *Wyscout*®, e então tratados em R® com o objetivo de compilar os dados em uma única planilha de Excel. Esses dados foram então importados para o Phyton®, onde foram manipulados para classificá-los de acordo com a posição de jogo. Finalmente, por meio de técnicas de scikit-learning, foi realizado um modelo estatístico de regressão logística para responder à pergunta do estudo. Os resultados revelaram individualidades entre cada posição, dentro de cada liga, sobre as ações que foram determinantes para obter vitórias. Os mais determinantes, e com grande recorrência foram: número de gols, assistências e passes certos. Outro fator comum na influência positiva para o resultado foi o local da partida, de modo que ser mandante demonstrou uma substancial vantagem. A partir dos resultados aqui apresentados treinadores poderiam utilizar as informações para ajustar treinamentos ou mesmo na busca por reforços específicos, de acordo com as ações significativas para a posição desejada.

Palavras-chave: Futebol, Habilidade técnica, Ações individuais, Posição de jogo, Regressão logística.

Abstract

Football is a sport that involves a range of physical, technical and tactical variables. Despite the increase in studies on the importance of technical variables in game performance in recent years, the study of physical variables still stands out. In addition, most studies have used simpler data analysis, which often fail to precisely answer some questions related to sports performance. Deeper statistical analysis is important to better explore performance variables and present a more objective analysis of the game. Thus, the present study had the following question to be answered: is it possible to find a model of individual technical characteristics according to the game positions that are decisive for winning in soccer? Thus, the objective of this work was to analyze the influence of the individual technical actions of each game position on the match result through a logistic regression model. A total of 5396 players from the top 10 national leagues (Germany, Argentina, Brazil, Colombia, Spain, France, Netherlands, England, Italy and Portugal) were analyzed in the matches they played from 2015 to 2019. The indicators were obtained from the website from Wyscout®, and then processed in R® with the aim of compiling the data in a single Excel spreadsheet. This data was then imported into Phyton®, where it was manipulated to classify it according to game position. Finally, using scikit-learning techniques, a statistical logistic regression model was performed to answer the study question. The results revealed individualities between each position, within each league, on the actions that were decisive for obtaining victories. The most decisive, and with great recurrence were: number of goals, assists and successful passes. Another common factor in positively influencing the result was the match location, so, playing as the home team demonstrated a substantial advantage. Based on the results presented here, coaches could use the information to adjust training or even in the search for specific reinforcements, according to the significant actions for the desired position.

Keywords: Football, Technical skill, Individual actions, Game position, Logistic regression.

Sumário

1. Introdução 5

2. Revisão de Literatura 7

2.1. Análise de Desempenho 7

2.2. Análise de jogo qualitativa / quantitativa 8

2.3. Estatística nos esportes 9

3. Métodos 11

4. Resultados 14

5. Discussão 28

5.1. Modelo utilizado no estudo 30

5.2. Aplicações práticas 31

6. Conclusão 32

7. Referências 33

1. Introdução

O futebol é um dos esportes mais populares do mundo e com isso seus seguidores comumente tentam predizer o desempenho dos atletas e resultados futuros tendo como base os resultados passados das partidas e competições. Porém, a maioria desse público geralmente não analisa de maneira objetiva os fatores que podem influenciar uma partida, tal como sua localização, adversário e variáveis situacionais de jogo, por exemplo (AQUINO et. al., 2017). Até mesmo as pesquisas científicas que levaram essas condições em consideração não conseguem geralmente obter um modelo de previsão com aceitável nível de precisão; apenas corroboram com a geral veracidade das influências do local da partida, status da partida e qualidade do oponente no desempenho físico dos atletas (LAGO-PEÑAS, 2012).

Inversamente, há um crescente interesse na área de estatística e análise de dados com foco em previsibilidade de eventos, sendo utilizadas ferramentas como derivadas de *machine learning* e regressão logística. Isso surgiu principalmente pelo ramo multibilionário de apostas esportivas presente no cenário global (FORREST et al., 2005). Nos primeiros estudos sobre análise do jogo de futebol foi revelado uma prevalência da aleatoriedade nos resultados finais (REEP e BENJAMIM, 1968), principalmente por conta da baixa frequência de gols em grande parte das partidas, algo que comprovadamente alterou-se dada a evolução dos atletas ao longo dos anos (WUNDERLICH, SECK & MEMMERT, 2021). Também existem outros fatores adicionais no jogo cuja previsibilidade é bastante incerta, como por exemplo, a ocorrência de lesões, penalidades máximas e cartões de advertência. Todos estes fatores mencionados, em isolamento ou em combinação, podem interferir diretamente no resultado de um jogo (DUBITZKY et. al., 2019). Levando em consideração essas informações, os modelos estatísticos de projeção têm sido desenvolvidos para revelar qual resultado deve ser esperado com base em dados retrospectivos. Segundo Dubitzky e colaboradores (2019), grandes bancos de dados com correspondências simples facilitam a previsibilidade de alguns jogos específicos ao longo do campeonato (exemplo, jogos da primeira rodada, ou jogos da décima rodada), porém com um número maior de variáveis, o modelo teria que ser atualizado a cada rodada para continuar com uma eficácia alta.

Os treinadores de futebol costumam utilizar dados mais simples, como o número de gols marcados por cada equipe, local da partida, quais equipes envolvidas, liga disputada e data do jogo, para se prepararem para uma partida futura. Isso facilita estudar o próximo oponente e assim planejar/adaptar uma estratégia adequada. Por exemplo, Gréhaine e Godbout (1995) indicam que as decisões tomadas pela equipe antes das partidas formam a estratégia, enquanto a tática se dá pelas interações contínuas entre a própria equipe, bem como entre as duas equipes oponentes. Essas interações ao longo do jogo podem ser classificadas em três níveis: espaço, tempo e ações (GARGANTA, 2009). Espaço são os locais onde ocorrem as ações do time no campo (i.e., posição), durante as fases de ataque ou defesa. O tempo refere-se ao momento em que ocorrem as ações, e a partir dele é possível obter indicativos de frequência e/ou rapidez do jogo. Por fim, as ações individuais (ex. número de passes realizados) revelam o envolvimento do jogador ao longo do jogo. Estas últimas podem ser classificadas como sendo com posse de bola (ex. passes, dribles, cruzamentos, finalizações) ou sem a posse de bola (ex. desarmes, faltas, interceptações, posicionamento (REIN e MEMMERT, 2016). Desta forma, entender variáveis individuais que podem determinar o resultado de uma partida pode ser de grande valia para a preparação das equipes e, assim, melhorar o desempenho no jogo. Ainda, a determinação destes fatores pode influenciar positivamente as contratações das equipes de futebol, procurando jogadores que apresentem desempenho individuais próximos a de equipes vencedoras/campeãs.

Dentro desse contexto, o objetivo do presente estudo é analisar a importância das ações individuais de cada uma das posições de jogo e o quanto elas poderiam ser determinantes para obtenção de uma vitória no futebol através de um modelo de regressão logística. O estudo se faz relevante pelo fato de serem escassas as análises estatísticas nesse grau de especificidade objetivando identificar o impacto de cada ação, por posição, para o resultado no futebol. Considerando que, quanto maior especificidade e variáveis houver em um modelo estatístico, maior poderá ser a taxa de acerto do mesmo, o estudo atual propõe uma análise mais individualizada das ações de acordo com cada papel posicional de jogo, tendo em vista fornecer um melhor entendimento dos efeitos de suas ações no resultado final do jogo de futebol.

2. Revisão de Literatura

2.1. Análise de Desempenho

As demandas de jogo no futebol têm evoluído de forma ininterrupta (NASSIS et al., 2020). Desta forma, cada vez mais os estudos científicos se fazem importantes, com o objetivo de tentar identificar/preparar padrões, estruturas táticas e sistemas de treinamento, sempre com o a intenção primária de antecipar variáveis e situações de jogo (LUCAS, 2001). A análise de desempenho se destaca nesses estudos, pois fornece ferramentas objetivas para quantificar o jogo de maneira individual e coletiva durante situações variadas (ex. ataque, defesa, transição de ataque, transição de defesa) a fim de encontrar padrões nas equipes, mesmo adversárias (MONTANO, 2014). Segundo Toledo e Ferreira (2014) a análise de desempenho constitui-se no registro, organização, síntese e interpretação de dados obtidos nos jogos para caracterização dos atletas e com potencial de, posteriormente, serem utilizados nos treinamentos e partidas futuras. Quando são conhecidas as ações mais significativas para a otimização do rendimento esportivo, é possível planejar treinamentos que visem melhoras nas tomadas de decisões e desempenho dos jogadores e das equipes, de modo cada vez mais específico (AQUINO, 2019).

Treinadores necessitam conhecer os restritores da performance de seus atletas, para assim realizar de maneira plena o que é exigido para o cargo (SAMPAIO, 1994). Em virtude disto, a área de pesquisa relacionada com comportamento dos atletas e das equipes nos esportes coletivos vem apresentando crescente número de publicações científicas, destacando-se a análise do jogo (GIL, 2012). Segundo Garganta (1997) a observação e análise de jogo são imprescindíveis para identificar fatores influenciadores da performance competitiva, devendo ser considerados em orientações ao longo da partida e nos treinamentos.

Garganta (2001) afirma que a finalidade da análise de jogo abrange quatro tópicos: I) criação de modelos de atividade dos jogadores e das equipes; II) identificação das características das atividades cuja presença ou ausência implicam na eficácia dos processos e de obtenção de resultados positivos; III) o desenvolvimento de treinamentos mais fidedignos implicando maior transferência para situações de jogo e IV) indicação das tendências de evolução da modalidade.

Considerando esse modelo de análise de performance, é necessário levar em conta duas dimensões fundamentais, o processo e o resultado. No primeiro temos a descrição das ações dos jogadores e sua eficácia no tempo, enquanto o segundo consiste no produto final dessas ações (VOLOSSOVITCH, 2008). Combinando estas duas variáveis em modelos de predição, conseguimos incrementar os conhecimentos sobre ações que podem levar ao sucesso, caso seja encontrada uma associação entre ambas. Desta forma, o treinador poderá encontrar melhores estratégias para alcançar uma evolução do rendimento individual e coletivo e consequentemente melhores resultados (CALLIFARIS, MARELLA e INNOCENTI, 1990).

2.2. Análise de jogo qualitativa / quantitativa

As situações de jogo podem ser estudadas a partir de duas classificações, uma na dimensão quantitativa e outra dimensão qualitativa. Estas estão relacionadas tanto com a obtenção dos dados quanto com a interpretação de rendimento em competições e treinos (TOJO, 2011). A medição das ações realizadas ao longo do jogo é geralmente expressa em números, o que permite avaliações de performances individuais ou coletivas. Isso caracteriza a dimensão de análise quantitativa (GARGANTA, 1998; BOLT, 2000). Já as análises de caráter qualitativo costumam se abster da utilização de uma simples representação numérica. Franks et al. (1983) define a análise qualitativa como a impressão adquirida por meio da observação e julgamento introspectivo.

Há diversas situações no futebol em que interações entre os jogadores se fazem presentes, tornando a subjetividade algo inerente ao esporte. Isso tende a tornar o futebol um esporte extremamente complexo para analisarmos unicamente por *scouts* (HUGHES, 2002). Em virtude destes pontos, a troca entre análises qualitativas e quantitativas se torna imprescindível, e uma maneira simples de tornar isso possível é através de uma consideração das opiniões do técnico nos dados obtidos pelo método quantitativo somados com a análise do próprio treinador em relação a jogos e treinamentos (GIL, 2012).

Garganta (1998) afirma que uma boa análise de desempenho parte do princípio de que a observação vai além do número de ações realizadas, contudo leva em consideração a qualidade destas ações. No entanto, o ideal é que os

procedimentos qualitativos sejam combinados com resultados obtidos por meio de processos quantitativos (GARCÍA, 2000).

Com o objetivo de fornecer uma análise adequada, os dados quantitativos devem passar por processos estatísticos robustos para tentar, por exemplo, compensar a parte subjetiva do jogo, que é difícil de ser expressa somente em números, uma vez que ela assume critérios pré-definidos (JÚNIOR et. al., 2002). É possível que análises qualitativas impliquem em algum grau de subjetividade na interpretação do fenômeno em questão, podendo levar a uma potencial imprecisão da análise (LEES, 2002). Levando em consideração esta possível falta de objetividade, especificidade e sensibilidade quando utilizada apenas métricas qualitativas da análise de jogo, os métodos quantitativos surgem para suprirem parte destas lacunas (OLIVEIRA, 1993).

2.3. Estatística nos esportes

A relação entre esportes e as estatísticas relacionadas ao jogo pode ser considerada apenas recente. No futebol temos a primeira menção da importância de computar dados das partidas aproximadamente no ano de 1936 (GODIK, 1996). Nessa época, debatia-se a necessidade de anotar a quantidade de passes e outras técnicas de jogo, bem como sua efetividade (GODIK, 1996). Contudo, o futebol não foi pioneiro no emprego das análises de estatística relacionada ao jogo. O pioneiro nesse tipo de trabalho foi Bill James, no Basebol (ANDERSON, SALLY, 2013. p. 16-18). Posteriormente, outros esportes também aderiram a essa técnica, devido às suas notáveis vantagens (explorar pontos fracos dos adversários, analisar quais ações da equipe necessitam melhorar, etc.), entre eles o basquete e o futebol americano (ANDERSON, SALLY, 2013. p. 16-18). Os dados estatísticos sempre elucidam os avanços na qualidade desportiva (VENDITE, 2005).

O futebol foi um esporte de muita aversão às propostas científicas. Dentre esportes de maior visibilidade no cenário mundial, é certamente o que mais resistiu em aceitar a importância das análises estatísticas. Entretanto, após reconhecer a importância desta ferramenta analítica dentro do contexto prático deste esporte, o uso da estatística relacionada ao jogo de futebol vem crescendo de maneira exponencial nas últimas décadas (ANDERSON e SALLY, 2013). A empresa pioneira na análise de dados do futebol foi a empresa Inglesa Opta Sports

(https://www.optasports.com/), a qual surgiu apenas em meados da década de 90, entretanto apenas na década de 10 a análise em tempo real e com imensa quantidade de dados foi definitivamente obtida (OPTA, 2019). Estas análises nos provem uma forma mais otimizada de avaliar o desempenho no futebol (VENDITE, 2003).

Os modelos estatísticos utilizados variam muito de acordo com o que se pretende obter. Para uma previsão de resultados das partidas, costuma-se utilizar a regressão de Poison, uma distribuição com objetivo de verificar a probabilidade de ocorrer uma série de eventos em um certo período de tempo, sabendo que são independentes (LEE, 1997), enquanto para uma previsão de resultados em geral (da partida, de lesões, de classificação final em campeonatos) tem sido utilizado a técnica de *machine learning*. Este é um método onde o modelo começa a reconhecer padrões e então prever os próximos acontecimentos de acordo com estes padrões (SCHNEIDER, 2018). Contudo, conforme a quantidade de dados envolvidos, pode ser mais simples e eficaz utilizar uma regressão logística (FREITAS, 2019). Segundo Hosmer e Lemeshow (1997), caso uma análise envolva variáveis de respostas discretas, ou binárias, é possível construir um modelo de regressão logística. Isso parece bastante aplicável aos parâmetros técnicos da estatística relacionada ao jogo no futebol, onde geralmente há uma dicotomização das ações (i.e. certo/errado, vitória e derrota; (VIEIRA, 2019).

Nos livros ou artigos publicados (ver para exemplo ANDERSON, SALLY, 2013; GARGANTA, 2001; LEES, 2002) onde há explicação para a importância de se trabalhar com processos estatísticos, ainda é escasso o uso das análises apresentadas anteriormente. Há uma carência de estudos desenhados para explorarem novas e melhores técnicas de analisar os dados de estatística relacionada ao jogo, tanto para preparação nos treinos como para efetuar táticas nas partidas; GARGANTA, 2001).

3. Métodos

Para a realização deste trabalho foram recolhidos os dados de todos os jogadores de cada equipe presentes nas principais ligas dos seguintes países: Alemanha (Bundesliga®), Argentina (Superliga®), Brasil (Brasileirão®), Colômbia (Primera A®), Espanha (LaLiga®), França (League One®), Holanda (Eredivisie®), Inglaterra (Premier League®), Itália (Serie A®) e Portugal (Primeira Liga®) nas partidas em que atuaram, dos anos de 2015, 2016, 2017, 2018 e 2019. Todas as posições táticas foram consideradas, excetuando-se os goleiros, devido às diferenças evidentes deste último em relação aos atletas de linha (WHITE et al., 2018). Agrupados, houve um tamanho amostral de 5396 jogadores no total. Os procedimentos foram aprovados pelo Comitê de Ética em Pesquisa do Instituto de Biociências de Rio Claro, Universidade Estadual Paulista “Júlio de Mesquita Filho” (parecer nº 4.395.900; data do parecer: 12/11/2020; CAAE: 37562720.7.0000.5465).

Esses dados foram obtidos por meio do site *Wyscout*® (https://wyscout.com/). Entre os dados baixados, estão as seguintes variáveis dependentes: Jogo; Competição; Data; Posição; Minutos jogados; Ações (totais e bem sucedidas); Gols; Assistências (passe que precede de imediato um gol); Finalizações (totais e no gol) (chute em direção a meta adversária, podendo ser na direção do gol ou para fora); Passes curtos e longos (totais e certos) (toque da bola de um jogador para outro do mesmo time); Cruzamentos (totais e certos) (bola alçada na área pela parte lateral do campo, podendo ser rasteiro ou aéreo); Dribles (totais e com sucesso) (tentativa de ludibriar o adversário por meio de movimentos corporais com objetivo de ultrapassá-lo com a bola); Duelos no solo e aéreos (totais e ganhos) (disputa pela posse da bola entre jogadores adversários); Interceptações (retomada da bola sem disputa com o adversário); Perdas de posse (no próprio campo ou do adversário) (todas as ações que acarretem em troca da posse de bola, bola para lateral, passe incompleto); Recuperações (no próprio campo ou do adversário) (Retomada da posse de bola); Cartão amarelo (minuto em que levou); Cartão vermelho (minuto em que levou). Considerou-se como variável independente o resultado final da partida: vitória, derrota ou empate.

Os dados foram tratados para a utilização de um modelo de classificação com o objetivo de identificar quais são as características, de cada posição de jogo (zagueiros, volantes, pontas (extremo), meio-campistas, meias-atacantes, laterais e

centroavantes), estas posições foram apresentadas pelo site *Wyscout*® no qual os jogadores já são designados nestas posições, relevantes para obter a vitória em uma partida de futebol (Figura 1). Para isso, todos os dados adquiridos passaram em um programa do software R® (R Studio v.1.2.5042; The R Foundation for Statistical Computing, Austria) que compilou as planilhas de todos os jogadores em uma única planilha de excel, resultado em 385059 linhas.

Figura 1. Exemplificação das posições utilizadas para análise das ações individuais.

Após esta etapa inicial, a planilha geral foi manipulada utilizando o pacote Pandas. Em sequência foi realizada uma análise exploratória utilizando NumPy junto ao Pandas (PEDREGOSA, 2011), estas primeiras etapas tiverem o objetivo de separar a planilha geral em blocos divididos por posição e liga jogada (ex. laterais da Premier League).

Com o uso do scikit-learning (PEDREGOSA, 2011) foi realizada uma regressão Lasso com validação cruzada interna (k=50) para a seleção de variáveis significantes para determinada posição e liga, dividindo-os em 8 categorias: Goleiros (Gol), Centroavantes (Cav), Laterais (Lat), Meia-atacantes (Mat), Meio-campistas (Mca), Pontas (Pta), Volantes (Vol) e Zagueiros (Zag). Por fim foi realizada a regressão logística com as variáveis selecionadas pelo Lasso, também utilizando de

validação interna (k=50), com o objetivo de quantificar a significância das variáveis pré-selecionadas.

Os resultados desta regressão logística foram apresentados em novas planilhas de Excel® (Microsoft® Excel® para Microsoft 365 MSO (16.0.14026.20202) 64 bits, versão 2105, Estados Unidos) com as especificações de nome da variável e o valor retornado da regressão, variando entre -1 e 1. Quanto mais próximos foram os valores–resultados–das extremidades da escala usada, maior foi a contribuição relativa do indicador de interesse para o resultado da partida. Valores positivos indicam que a ocorrência do evento foi diretamente relacionada com a vitória na partida enquanto valores negativos indicam uma relação inversa entre eles. Os resultados foram aproximados em até três casas decimais. Finalmente, foram construídos gráficos de barras, para cada posição de cada liga considerada, para facilitar a leitura e interpretação dos achados.

Visto que a literatura específica não apresenta parâmetros para classificar qualitativamente os resultados obtidos, adotaremos uma correlação arbitrária em três partes, sendo, |0 – 0,333| baixa, |0,334 – 0,666| moderada, |0,667 – 1| alta.

4. Resultados

No total foram analisados 5396 jogadores de 10 ligas diferentes ao longo dos jogos em que estes jogadores participaram. Cada liga nacional resultou em diferentes ações relevantes para a obtenção da vitória e ações que impactam de forma negativa no resultado.

Por não apresentarem dados em relação as ações propostas a serem analisadas, os resultados dos goleiros foram irrelevantes neste estudo. A Figura 2 apresenta os gráficos com os resultados da análise de regressão logística dos Laterais (Lat) de cada liga. Nos resultados dos laterais encontramos que o local, gol e assistência se mantêm aparecendo, porém nos laterais percebe-se que a assistência tem uma relevância maior que o gol. Outra estatística que apareceu forte foi a de passes certos, apresentado alta relevância no campeonato francês e outras três vezes apareceu com índice moderado, nos campeonatos Italiano, Português e Inglês. Por fim uma estatística comum em todas as ligas e de influência negativa foi a de cruzamento, com destaque maior para as ligas sul-americanas, onde chegou a alcançar níveis moderados de influência negativa.

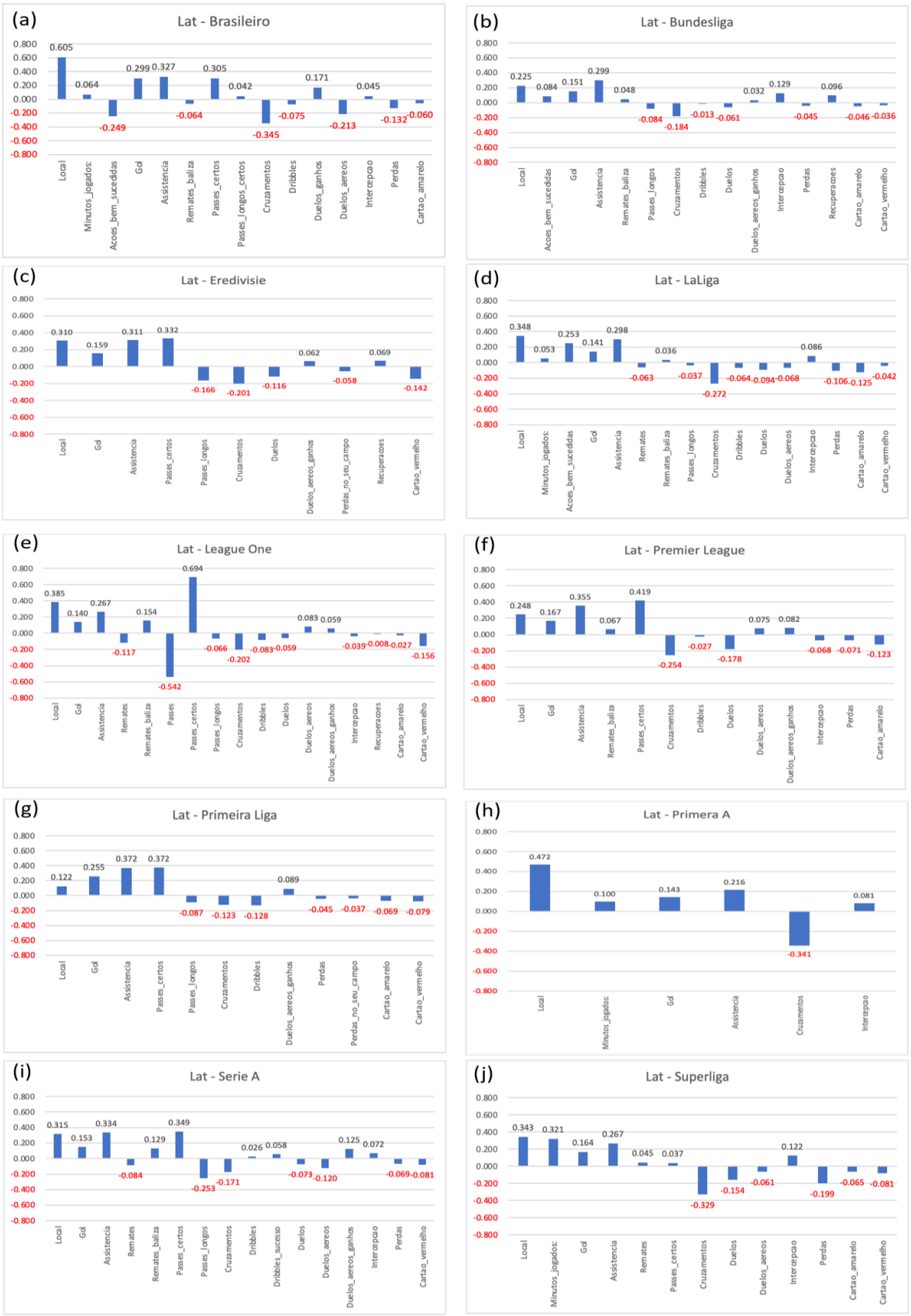

Figura 2. Valores de regressão logística dos laterais (Lat) dos campeonatos, brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

A Figura de 3 apresenta os gráficos com os resultados da análise de regressão logística dos Zagueiros (Zag) de cada liga. Com relação a posição de zagueiros, a estatística relacionada ao jogo que mais se destacou (chegando a alcançar julgamento moderado) foi uma com influência negativa no resultado, dependendo da liga, os duelos tiveram grande impacto, principalmente no campeonato italiano. Em outras palavras, um maior número de duelos realizados pelos zagueiros foi relacionado a menor chance de obtenção de vitória. Além disto local se manteve com aspecto positivo.

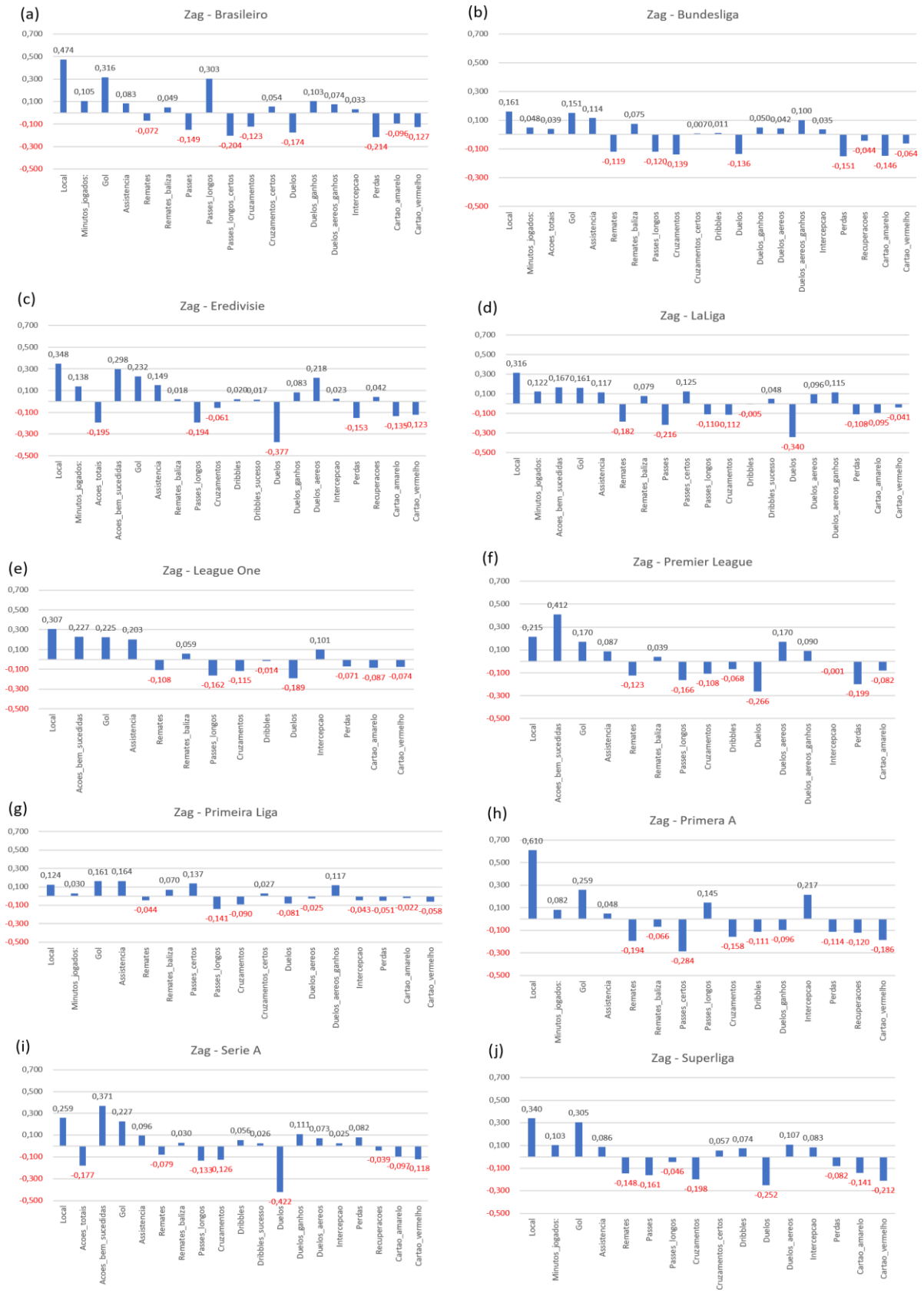

Figura 3. Valores de regressão logística dos zagueiros (Zag) dos campeonatos: brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

A Figura 4 apresenta os gráficos com os resultados da análise de regressão logística dos volantes (Vol) de cada liga. Relativamente aos volantes, observou-se uma redução na influência do gol e da assistência para o resultado da partida, de modo que na maioria dos casos um nível baixo de coeficiente retornados foi identificado, diferente do local que continuou com grande influência. Pelo contrário, houve impacto negativo das perdas de posse, as quais atingiram em nível moderado no campeonato colombiano. Duelos ganhos apareceu com destaque, apesar da baixa influência por si só no resultado da partida, em algumas ligas como da França e Holanda.

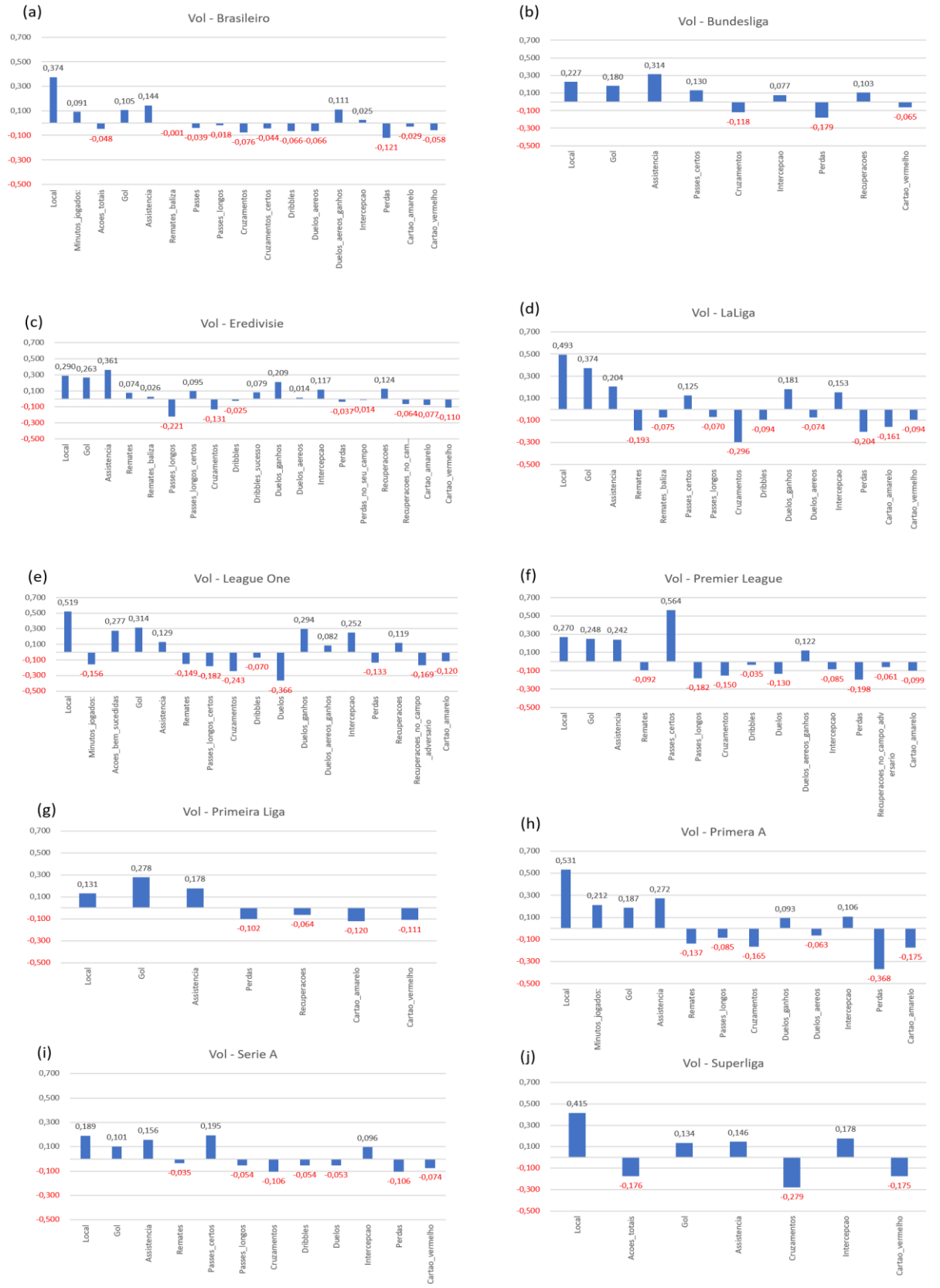

Figura 4. Valores de regressão logística dos volantes (Vol) dos campeonatos: brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

A Figura 5 apresenta os gráficos com os resultados da análise de regressão logística dos Meio-campistas (Mca) de cada liga. Em se tratando dos meio campistas, local, gol, assistência e passes certos mantiveram-se como estatísticas relacionadas ao diretamente favoráveis a obtenção de vitória. Contudo, outra estatística (i.e. defensiva) apresentou indícios de influência que merecem destaque. A intercepção, embora tenha sido classificada como de baixa influência no resultado, se destaca entre as estatísticas dos meio campistas como algo positivo.

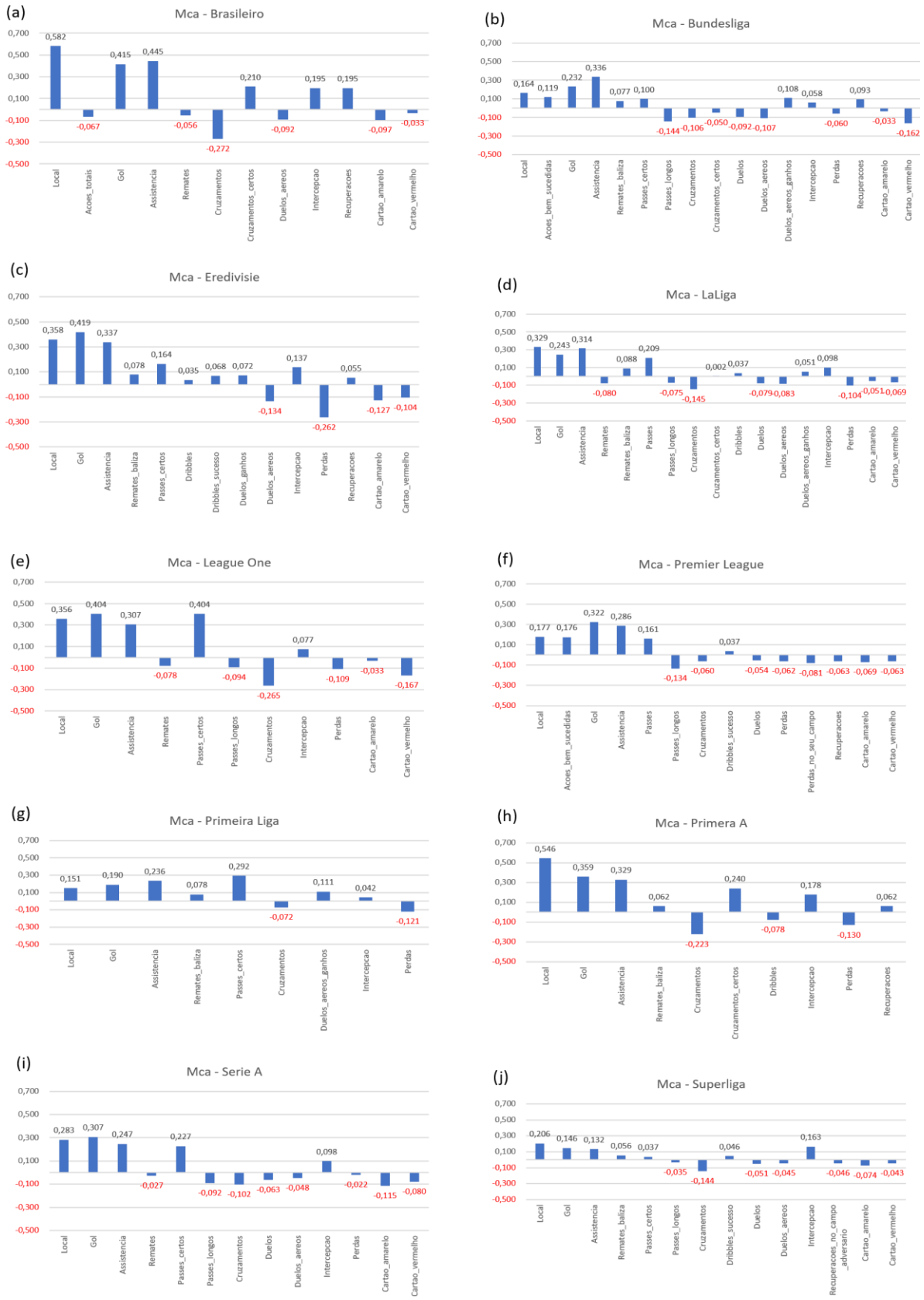

Figura 5. Valores de regressão logística dos meio campistas (Mca) dos campeonatos: brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

A Figura 6 apresenta os gráficos com os resultados da análise de regressão logística dos Meias-atacantes (Mat) de cada liga. Com os meias atacantes foi possível identificar que local, gol e assistência aparecerem diversas vezes com níveis moderados de influência. No caso específico da Premier League a influência do gol e assistência foi próxima a alta. Passes certos também se apresentou importante na busca da vitória, chegando a níveis moderados no campeonato italiano e colombiano.

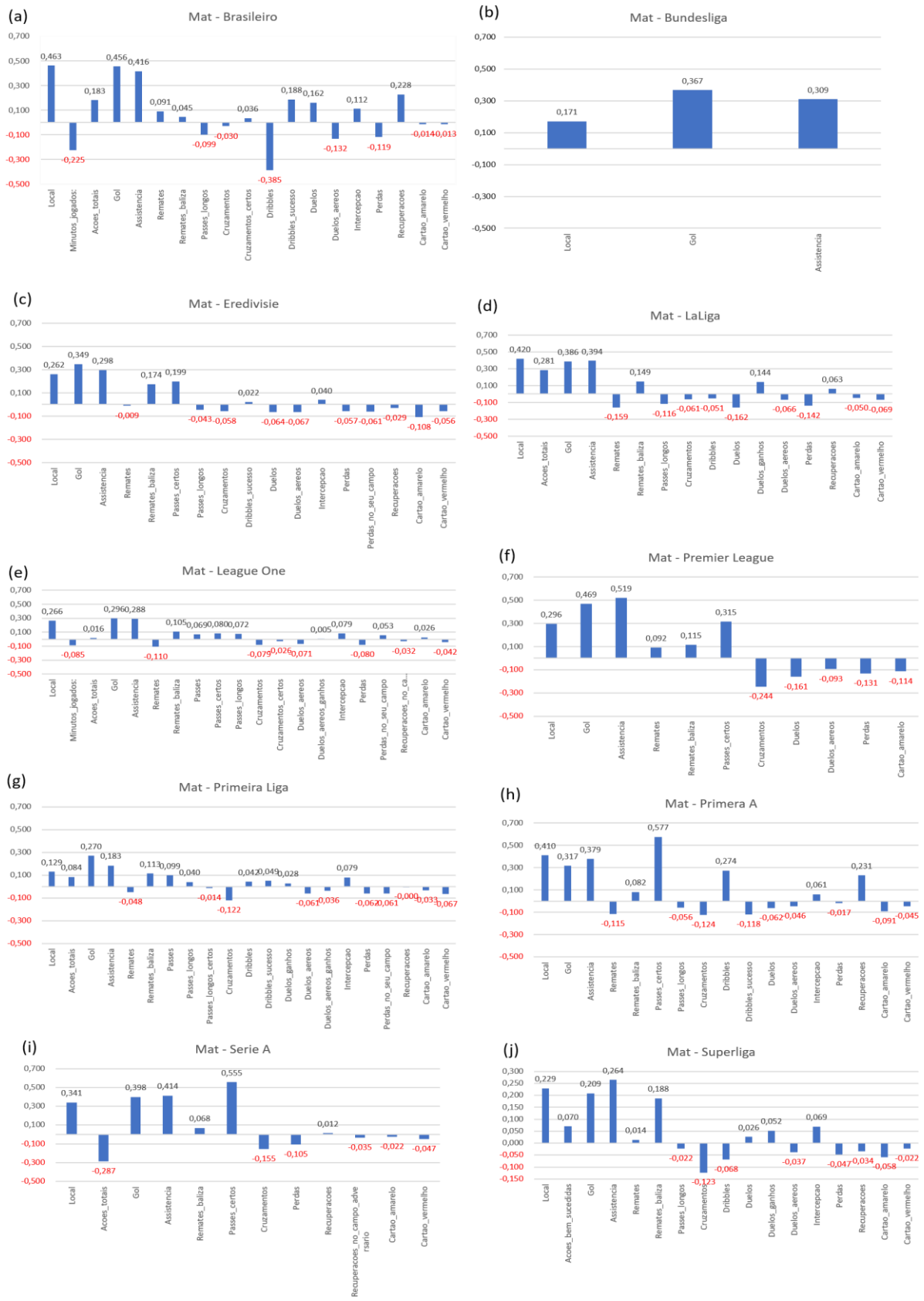

Figura 6. Valores de regressão logística dos meias-atacantes (Mat) dos campeonatos: brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

A Figura 7 apresenta os gráficos com os resultados da análise de regressão logística dos pontas (Pta) de cada liga. Os pontas apresentaram cinco principais estatísticas com notável influência no resultado final do jogo: local, gol, assistência, passes certos (mantiveram-se como importantes), remates a baliza (voltou a aparecer com certo destaque entre as estatísticas retornadas para os pontas de cada liga) e, finalmente, o cruzamento exibiu mais uma vez um impacto negativo para obtenção de vitórias.

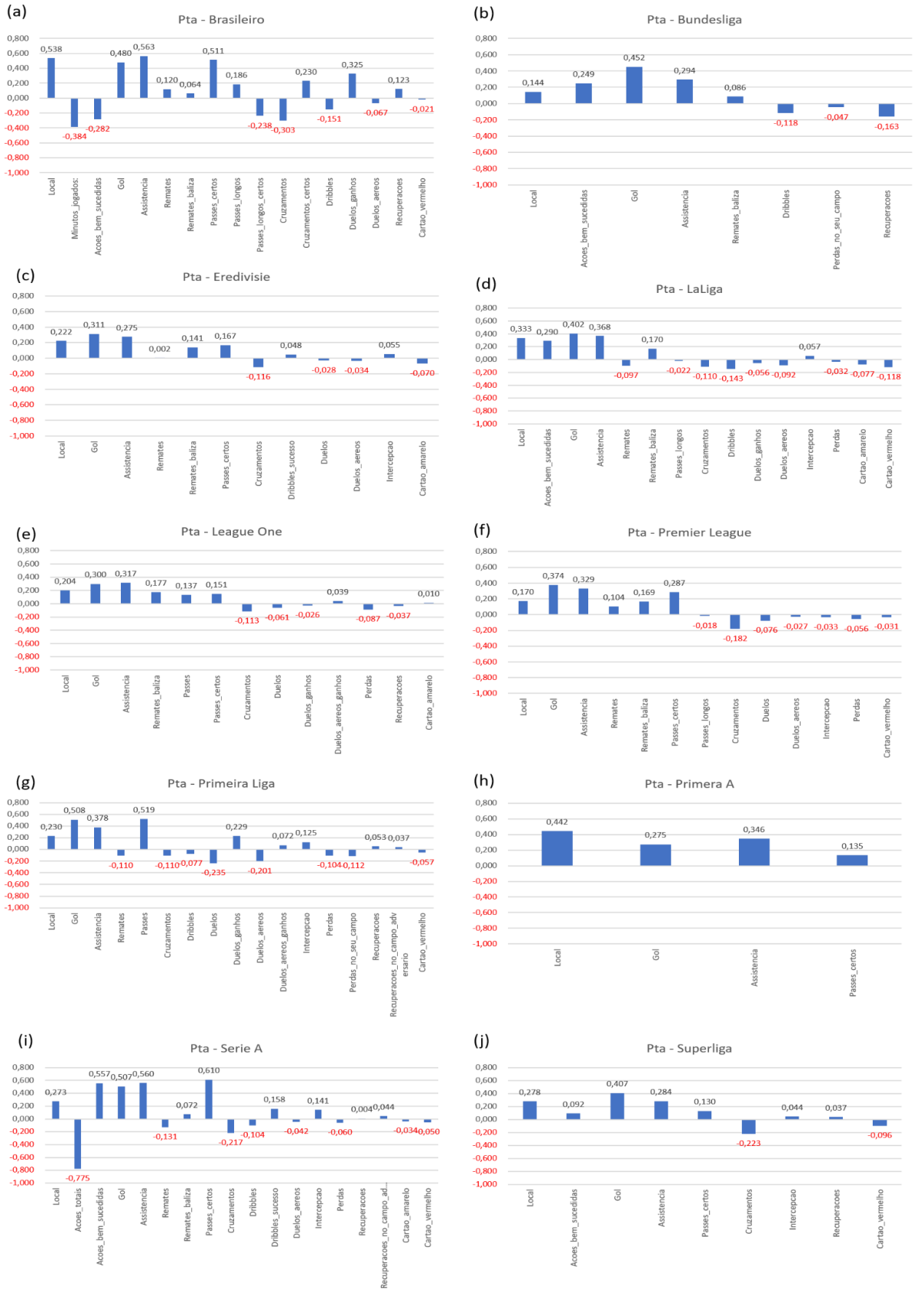

Figura 7. Valores de regressão logística dos pontas (Pta) dos campeonatos: brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

A Figura 8 apresenta os gráficos com os resultados da análise de regressão logística dos Centroavantes (Cav) de cada liga. Para os centroavantes, local, gol e assistência foram as ações técnicas mais relevantes, porém uma terceira estatística que aparece muito frequentemente é a de remates_baliza, a qual se destacou pela importância principalmente no campeonato francês e no alemão.

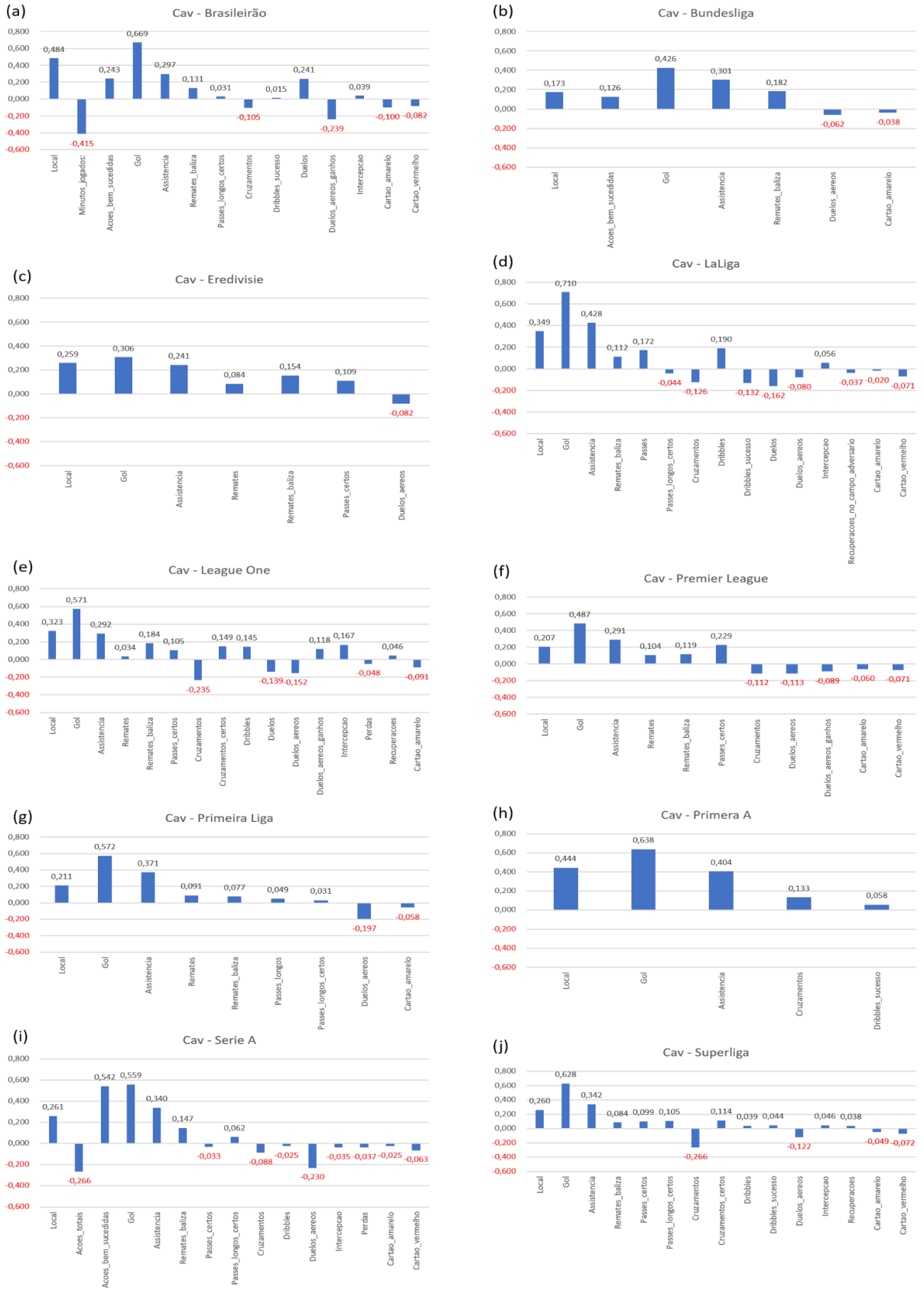

Figura 8. Valores de regressão logística dos centroavantes (Cav) dos campeonatos: brasileiro (a), alemão (b), holandês (c), espanhol (d), francês (e), inglês (f), português (g), colombiano (h), italiano (i), argentino (j).

5. Discussão

O objetivo do presente estudo foi de analisar a importância de uma variedade de ações técnicas individuais, de acordo com cada uma das posições de um time, e o quanto elas podem ser determinantes para a obtenção de uma vitória no futebol. Com a divisão feita em três partes para classificarmos o resultado (baixa, moderada e alta), foi possível atingir um maior grau de objetividade na discussão dos resultados e assim qualificar adequadamente suas respectivas relevâncias. De modo geral, entre os principais resultados, destacou-se em primeiro momento, a importância das seguintes ações de jogo: a ocorrência de gols, assistência e passes certos. Além disso, o local da partida (mandante ou visitante) foi um fator importante. Estes foram os índices associados ao alcance de vitórias. Inesperadamente, o número de cruzamentos realizados pelos jogadores de todas as posições, entre os 70 grupos estudados o cruzamento somente não apareceu de maneira negativa em 9 casos. Este foi um fator negativamente ligado a ocasiões em que as equipes saíram vencedoras.

O local da partida, mandante ou visitante, foi um fator importante para o resultado da partida. Este resultado foi similar a estudos anteriores (SILVA, 2008). Aparente em todos os grupos estudados, o fato de ser mandante em uma partida de futebol foi comprovado novamente como algo vantajoso. Segundo Nevill (1999) o efeito da multidão da torcida local tem um peso grande no resultado, impactando em dois aspectos, primeiro podendo elevar o desempenho dos jogadores da equipe mandante e segundo, de maneira inconsciente, os árbitros podem tomar decisões mais favoráveis ao time da casa. Então, ao pegarmos o primeiro fator citado, a melhora no desempenho do atleta faz com que suas estatísticas sejam melhores ao longo da partida, e ao processarmos estas estatísticas no modelo de regressão logística, encontramos essa correlação. Mesmo assim, foi possível verificar diferenças entre os valores nas posições de mesma liga. Um dos fatores que pode explicar essa variação na única estatística independente do jogador são as diferentes formações táticas de cada equipe (ex. 4-4-2; 4-3-3; 3-4-3), entre outras. Isso poderia ocasionar uma variação no número de jogadores analisados para determinada posição dentro de uma mesma liga, exemplo centroavantes premier league, onde há times (exemplo: Manchester City) que jogam sem um centroavante e consequentemente isso altera o peso de suas ações no resultado obtido.

Algumas outras estatísticas, além do local, foram mais comuns de serem encontradas entre os destaques de cada posição e independente da liga. Gol, assistência e passes certos foram as ações individuais mais frequentemente destacadas como de grande influência positiva. O objetivo de uma partida é a vitória, sendo assim, indiscutivelmente gols e assistências (passes que precedem o gol) são algo de extrema importância para o resultado. O sucesso na partida é determinado pelo balanço entre o número de gols marcados e sofridos na partida (GOMES e SOUZA, 2008), tornando mais evidente a importância destas duas estatísticas. Portanto, já era esperado seu valor correspondente ser positivo e alto, todavia há posições em que seu peso para obtenção da vitória não foi substancial. Isso ocorreu, como por exemplo, nos dados dos zagueiros, onde o coeficiente de influência atinge no máximo 0,316 na liga Brasileirão (Figura 3a). Por outro lado, os passes certos não apresentam uma relação tão lógica para o resultado da partida. Todavia uma das maneiras de não se tomar gol é com a manutenção da posse de bola (ANDERSON e SALLY, 2013). Um alto número de passes certos é um bom indicativo de uma maior posse de bola, privando o adversário de poder fazer o gol, algo primordial na busca da vitória.

No que diz respeito a influência negativa, vale destacar os cruzamentos, que apareceram em mais de uma posição exercendo impacto negativo no resultado dos jogos. Neste sentido, especula-se que isso poderia ser em decorrência de um elevado número de cruzamentos que acabam sendo interceptados e que geram contra-ataques e perda da posse de bola, em uma tentativa das equipes “forçarem” jogadas de ataque aéreas. Isso corrobora com indicações anteriores (ANDERSON e SALLY, 2013), que sugerem, por exemplo, que os escanteios curtos podem ser mais efetivos que cruzamentos diretos na área adversária. Importante destacar cinco grupos nos quais houve uma diferença moderada entre o valor de cruzamentos e o de cruzamentos certos (ver Figuras 5a, 5h, 7a, 8e, e 8j), mostrando situações em que o cruzamento em geral é uma influência negativa, porém os cruzamentos certos apresentam dados positivos na busca pelo sucesso. Isso nos chamam à atenção para analisarmos a importância de não desperdiçar a posse de bola com excessivos cruzamentos mal planejados, mas sim utilizar desta ação quando portarmos uma certeza maior da sua eficácia, seja por uma situação mais favorável no momento durante o jogo, ou por um preparo melhor ao longo dos treinamentos.

Os resultados das ações consideradas relevantes para obtenção da vitória e as de impacto negativo neste objetivo tiveram variações entre as ligas. Isso possivelmente ocorre devido às diferenças culturais pertinentes a cada uma das ligas/países (DELLAL et al., 2011), uma vez que todos os jogadores apresentavam as mesmas ações para serem analisadas, porém o retorno não foi constante de acordo com a significância da ação encontrada pelo modelo estatístico em cada grupo (liga).

5.1. Modelo utilizado no estudo

O modelo utilizado no estudo recebeu os dados com todas as ações de todos os jogadores e então, após os jogadores foram designados para seus respectivos grupos (posição e liga). Os dados foram processados para retornar em cada grupo as ações que mais tivessem impacto no objetivo, sendo este positivo ou negativo. Ações não relevantes para determinado grupo não foram retornadas nos dados (excluídas a partir do “lasso”), e devido a isso, foram observadas estatísticas que aparecem em alguns grupos posicionais enquanto isso não foi verdadeiro em outros. Para exemplificar, o indicador de cartão vermelho aparece em 9/10 ligas na posição de zagueiros (i.e., exceto na Premier League, figura 3, imagem 3f), mas em apenas 5/10 isso se manteve no caso dos centroavantes, mostrando a diferença de impacto entre estas duas posições que uma expulsão causa na obtenção da vitória.

De modo geral, houve um baixo número de ações com influência de grau alto sobre o resultado final das partidas. Apenas 4 (Figuras 2e, 7i, 8a e 68d) no total entre as 70 análises desenvolvidas obtiveram os níveis mais elevados de importância. É possível relacionar esta ocorrência com a ligação direta entre o resultado da partida e a combinação de indicadores de desempenho, instrumentos de identificação das adaptações dos jogadores na partida e o componente de imprevisibilidade do jogo (HUGHES; BARTLETT, 2002). Não obstante, considerou-se aqui apenas os indicadores técnicos, participação e advertências recebidas pelos jogadores enquanto o desempenho físico (AQUINO et al., 2017, 2020; HOPPE et al., 2015) e a tática das equipes (GARGANTA, 2001; TAYLOR et al., 2010; GONÇALVES et al., 2021) também são componentes bem estabelecidos como sendo de fundamental relevância para a construção de vitórias no futebol, além de outro fator que poderia interferir na influência da importância dos gols é o momento

em que foram marcados (ex. primeiro gol em uma partida “apertada” ou quarto gol em uma com placar “amplo”) (ANDERSON e SALLY, 2013), e isso deve ser considerado em pesquisas futuras, sendo uma das limitações do trabalho atual.

5.2. Aplicações práticas

A partir dos resultados fornecidos pelo trabalho atual, é possível extrair algumas possíveis aplicações práticas. Isso fica evidente em cada gráfico demonstrando os pontos positivos e negativos das ações de jogo, ao nível de cada papel posicional e em cada liga nacional de futebol. Com base nestes dados, as equipes poderiam ter um embasamento para efetuar buscas por reforços enquadrados nas exigências específicas (i.e. com nível adequado de desempenho em determinado componente técnico essencial para obtenção de vitórias), comparar as estatísticas do jogador para antecipar a adaptação necessária à liga e assim auxiliar em potenciais resultados positivos. Outra questão possível de ser trabalhada de posse dos resultados é realizar um trabalho individualizado com os jogadores da equipe para melhorar as ações que retornaram valor negativo moderado/alto, e assim diminuir os prováveis prejuízos que determinada posição possa ocasionar ao desempenho durante jogos oficiais de futebol.

6. Conclusão

Por meio dos resultados obtidos podemos concluir que cada posição em cada liga tem um estilo diferente com ações significativas para obtenção de vitórias que podem divergir entre elas. Porém alguns parâmetros apresentam potencial de serem parcialmente generalizados, como o fato de ser mandante ou não no local da partida, a ocorrência de gols, assistências e passes certos, os quais foram os determinantes técnicos mais comuns de apresentarem impacto positivo para obtenção de uma vitória. Outro ponto é o fato de os cruzamentos em geral acarretarem uma influência negativa na busca pela vitória, se mostrando não tão interessantes de serem feitos ao longo da partida. Os dados aqui apresentados poderiam servir como base para permitir individualizar o treinamento de acordo com a liga e posição do atleta, almejando uma evolução individual em aspectos que em conjunto com os componentes físicos e táticos, poderia ajudar no alcance de mais vitórias no contexto de futebol.

7. Referências

ANDERSON, Chris; SALLY, David. Os números do jogo: por que tudo o que você sabe sobre futebol está errado. **São Paulo: Paralela**, 2013.

Aquino, R., Carling, C., Vieira, L. H. P., Martins, G., Jabor, G., Machado, J., ... & Puggina, E. (2020). Influence of situational variables, team formation, and playing position on match running performance and social network analysis in brazilian professional soccer players. **The Journal of Strength & Conditioning Research**, *34*(3), 808-817.

AQUINO, Rodrigo et al. Influence of match location, quality of opponents, and match status on movement patterns in Brazilian professional football players. **The Journal of Strength & Conditioning Research**, v. 31, n. 8, p. 2155-2161, 2017.

AQUINO, Rodrigo Leal de Queiroz Thomaz et al. **Observação, análise e interpretação do desempenho em jogo no futebol: implicações para a avaliação e treinamento**. Tese de Doutorado. Universidade de São Paulo, 2019.

BARROS, Mauricio. **Qual o melhor campeonato nacional do mundo? Vejamos o que dizem os números**. 2017. Disponível em: http://www.espn.com.br/blogs/mauriciobarros/724979_qual-o-melhor-campeonato-nacional-do-mundo-vejamos-o-que-dizem-os-numeros. Acesso em: 30 abr. 2021.

Bolt, B. (2000). Using computers for qualitative analysis of movement. Journal of Physical Education, Recreation and Dance, 71 (3), 15 - 18.

CALLIGARIS, Alfredo; MARELLA, Mario; INNOCENTI, A. **Il calcio al computer da Mexico'86 verso Italia'90: analisi sulle nazione: Argentina, Brasile, Francia, Italia, Marocco, Olanda, Spagna, URSS**. Società stampa sportiva, 1990.

Dellal, A., Chamari, K., Wong, D. P., Ahmaidi, S., Keller, D., Barros, R., ... & Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European journal of sport science*, *11*(1), 51-59.

DUBITZKY, Werner et al. The open international soccer database for machine learning. **Machine Learning**, v. 108, n. 1, p. 9-28, 2019.

FORREST, David; GODDARD, John; SIMMONS, Robert. Odds-setters as forecasters: The case of English football. **International journal of forecasting**, v. 21, n. 3, p. 551-564, 2005.

FRANKS, I.; GOODMAN, D.; MILLER, G. Analyse de la performance: qualitative ou quantitative? **Science du Sport**, [S.l.], GY-1, p.1-7, 1983.

FREITAS, Pablo Henrique de. Regressão Logística na modelagem da probabilidade de vitória em jogos de Futebol Americano. 2019.

GARCÍA, J. L. Metodologia de Análises y Observation de equipos en la alta competición. **Balonmano: Perfeccionamiento e investigation**, p. 234-237, 2000.

GARGANTA, Júlio Manuel. Modelação táctica do jogo de futebol: Estudo da organização da fase ofensiva em equipas de alto rendimento. 1997.

GARGANTA, Julio Manuel. O ensino dos jogos desportivos colectivos. Perspectivas e tendências. **Movimento (ESEFID/UFRGS)**, v. 4, n. 8, p. 19-27, 1998.

GARGANTA, Júlio. A análise da performance nos jogos desportivos. Revisão acerca da análise do jogo. **Revista portuguesa de ciências do desporto**, v. 1, n. 1, p. 57-64, 2001.

GIL, André Filipe Pereira et al. **Futebol: análise do resultado final: estudo de quatro ligas profissionais europeias**. 2012. Dissertação de Mestrado.

GODIK, Mak Alexandrovich; GOMES, Antonio Carlos; MANTOVANI, Marcelo. **Futebol: preparação dos futebolistas de alto nível**. Ed. Grupo de Palestra Sport, 1996.

Gonçalves, L. G. C., Clemente, F. M., Vieira, L. H. P., Bedo, B., Puggina, E. F., Moura, F., ... & Aquino, R. (2021). Effects of match location, quality of opposition, match outcome, and playing position on load parameters and players' prominence during official matches in professional soccer players. **Human Moviment**, *22*(3), 35-44.

GRÉHAIGNE, Jean-Francis; GODBOUT, Paul. Tactical knowledge in team sports from a constructivist and cognitivist perspective. **Quest**, v. 47, n. 4, p. 490-505, 1995.

Hoppe, M. W., Slomka, M., Baumgart, C., Weber, H., & Freiwald, J. (2015). Match Running Performance and Success Across a Season in German Bundesliga Soccer Teams. International Journal of Sports Medicine, 36(7), 563–566. https://doi.org/10.1055/s-0034-1398578

HOSMER, David W. et al. A comparison of goodness-of-fit tests for the logistic regression model. **Statistics in medicine**, v. 16, n. 9, p. 965-980, 1997.

HUGHES, Mike D.; BARTLETT, Roger M. The use of performance indicators in performance analysis. **Journal of sports sciences**, v. 20, n. 10, p. 739-754, 2002.

JÚNIOR, Dante; GASPAR, Alexandre; SINISCALCHI, Marcello. Análise estatística do desempenho técnico coletivo no basquetebol. **Lecturas en Educación Física y Deportes, Buenos Aires**, v. 8, n. 49, 2002.

LAGO-PEÑAS, Carlos. The role of situational variables in analysing physical performance in soccer. **Journal of human kinetics**, v. 35, n. 1, p. 89-95, 2012.

LEE, Alan J. Modeling scores in the Premier League: is Manchester United really the best?. **Chance**, v. 10, n. 1, p. 15-19, 1997.

LEES, Adrian. Technique analysis in sports: a critical review. **Journal of sports sciences**, v. 20, n. 10, p. 813-828, 2002.

LUCAS, João Manuel da Costa. Comparação entre a concepção do treinador e a percepção dos jogadores, face à prestação táctica, individual e colectiva: Um Estudo de caso numa equipa de futebol de Juniores A. 2001.

MONTANO, A. S. G. **Análise de desempenho no futebol:** Uma análise do panorama atual desta área de atuação no futebol brasileiro, 2014. Disponível em: < https://universidadedofutebol.com.br/analise-de-desempenho-no-futebol/>. Acesso em: 23 jun. 2020.

Nassis GP, Massey A, Jacobsen P, Brito J, Randers MB, Castagna C, Mohr M, Krustrup P. Elite football of 2030 will not be the same as that of 2020: Preparing players, coaches, and support staff for the evolution. Scand J Med Sci Sports. 2020 Jun;30(6):962-964. doi: 10.1111/sms.13681.

NEVILL, A. M.; HOLDER, R. L. Home advantage in sport. **Sports Medicine**, v. 28, n. 4, p. 221-236, 1999. https://doi.org/10.2165/00007256-199928040-00001.

OKAZAKI, V. H. A.; DASCAL, J. B.; OKAZAKI, F. H. A.; TEIXEIRA, L. A. Ciência e tecnologia aplicada à melhoria do desempenho esportivo. **Revista Mackenzie de Educação Física e Esporte**, v. 11, n. 1, p. 143-157, 2012.

OLIVEIRA, J. A análise do jogo em basquetebol. **A ciência do desporto, a cultura e o Homem**, p. 297-306, 1993.

OPTA SPORTS. **Driving the sports data agenda since 1996**. London, 2019. Disponível em: <https://www.optasports.com/about/our-history>. Acesso em: 07 ago. 2020.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, v. 12, p. 2825-2830, 2011.

REEP, Charles; BENJAMIN, Bernard. Skill and chance in association football. **Journal of the Royal Statistical Society. Series A (General)**, v. 131, n. 4, p. 581-585, 1968.

REIN, Robert; MEMMERT, Daniel. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. **SpringerPlus**, v. 5, n. 1, p. 1-13, 2016.

SAMPAIO, J. O Scouting na preparação das competições da Divisão A-1 do Basquetebol português. **UTAD: Monografia não publicada**, 1994.

SCHNEIDER, Cristian Felipe. Machine learning aplicado na previsão de resultados de partidas de futebol: um estudo de caso para comparação de diferentes classificadores. 2018.

SILVA, Cristiano Diniz da; MOREIRA, Danilo Gomes. A vantagem em casa no futebol: comparação entre o Campeonato Brasileiro e as principais ligas nacionais do mundo. 2008.

Taylor, B. J., Mellalieu, D. S., James, N., & Barter, P. (2010). Situation variable effects and tactical performance in professional association football. *International Journal of Performance Analysis in Sport*, *10*(3), 255-269.

TOJO, Óscar Miguel Farias Fialho. **Factores de sucesso nas transições defesa-ataque em diferentes contextos competitivos: estudo realizado com quatro equipas melhor classificadas no campeonato do mundo 2010 de futebol**. 2011. Tese de Doutorado.

TOLEDO, H. C.; FERREIRA, G. N. P.; BRAZIL, G. P. P. Desenvolvimento da análise de desempenho esportivo no voleibol de alto rendimento no contexto da gestão do conhecimento. **Podium Sport, Leisure and Tourism**, v. 3, n. 3, p. 25-35, 2014.

VENDITE, Caroline Colucio; VENDITE, Laércio Luis; MORAES, Antonio Carlos de. Scout No Futebol: Uma Ferramenta Para a Imprensa Esportiva “In:”, CONGRESSO BRASILEIRO DE CIENCIAS DA COMUNICAÇÃO., 2005, Rio de Janeiro, UERJ. **Trabalho apresentado ao NP**, p. 1-10.

VENDITE, Laércio Luiz; DE MORAES, Antonio Carlos; VENDITE, Carolina Coluccio. Scout no futebol: uma análise estatística. **Conexões**, v. 1, n. 2, p. 183-194, 2003.

VIEIRA, Luiz H. Palucci et al. Team dynamics, running, and skill-related performances of Brazilian U11 to professional soccer players during official

matches. **The Journal of Strength & Conditioning Research**, v. 33, n. 8, p. 2202-2216, 2019.

VOLOSSOVITCH, A. Análise dinâmica do jogo de andebol. Estudo dos factores que influenciam a probabilidade de marcar golo. **Unpublished Tese Doutorado, Universidade Técnica de Lisboa, Lisboa**, 2008.

White A, Hills SP, Cooke CB, Batten T, Kilduff LP, Cook CJ, Roberts C, Russell M. Match-Play and Performance Test Responses of Soccer Goalkeepers: A Review of Current Literature. Sports Med. 2018 Nov;48(11):2497-2516. doi: 10.1007/s40279-018-0977-2. PMID: 30144021.

Wunderlich F, Seck A, Memmert D. The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. J Sports Sci. 2021 May 24:1-16. doi: 10.1080/02640414.2021.1930685.

UNESP - INSTITUTO DE BIOCIÊNCIAS DE RIO CLARO DA UNIVERSIDADE ESTADUAL PAULISTA

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Quais ações são significativas para uma vitória no futebol? Uma análise por regressão logística

Pesquisador: Fabio Augusto Barbieri

Área Temática:

Versão: 1

CAAE: 37562720.7.0000.5465

Instituição Proponente: Instituto de Biociências de Rio Claro/ Universidade Estadual Paulista -

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.395.900

Apresentação do Projeto:

Trata-se de uma pesquisa de TCC do aluno Victor Valvano Schimidt, sob orientação do Prof. Dr. Fabio Augusto Barbieri e conta com co-orientação de Luiz Henrique Palucci Vieira. projeto de pesquisa tem como temática "Quais ações são significativas para uma vitória no futebol? Uma análise por regressão logística".

Objetivo da Pesquisa:

"O objetivo do presente estudo é analisar a importância das ações individuais de cada uma das posições de um time e o quanto elas são determinantes para obtenção de uma vitória no futebol."

Avaliação dos Riscos e Benefícios:

"Riscos:

Não traz riscos aos participantes, pois os dados já estavam prontos e disponibilizados na internet.

Benefícios:

Caso a hipótese venha a ser verdadeira, este estudo pode trazer benefícios na maneira como os jogadores se preparam, podendo aprimorar os pontos determinantes na busca do sucesso nas partidas. Além disso, técnicos poderão trabalhar em estratégias melhores para a preparação de suas equipes."

Comentários e Considerações sobre a Pesquisa:

"Para a realização desse trabalho, foram recolhidos os dados de todos os jogadores de cada

equipe presentes nas principais ligas dos seguintes países: Alemanha, Argentina, Brasil, Colômbia, Espanha, França, Holanda, Inglaterra, Itália e Portugal. Dando um n de aproximadamente 5000 jogadores. Esses dados foram obtidos por meio do site Wyscout®. Entre os dados baixados, estão as seguintes variáveis: Jogo; Competição; Data; Posição; Minutos jogados; Ações totais – Ações bem sucedidas; Gols; Assistências; Finalizações – Finalizações certas; Passes – Passes certos; Passes longos – Passes longos certos; Cruzamentos – Cruzamentos certos; Dribles – Dribles com sucessos; Duelos – Duelos ganhos; Duelos aéreos – Duelos aéreos ganhos; Interceptações; Perdas – Perdas no seu meio campo; Recuperações – Recuperações no meio de campo do adversário; Cartão amarelo (minuto em que levou); Cartão vermelho (minuto em que levou). Os dados serão tratados para a utilização de um modelo de classificação com o objetivo de identificar quais são as características, de cada posição de jogo, relevantes para obter a vitória em uma partida de futebol. O modelo terá sua base em regressão logística, pois dessa maneira conseguimos retornar ao modelo e observar quais foram as características importantes para a classificação do time ganhador, de acordo com as variáveis utilizadas. Os softwares utilizados nessa pesquisa serão o Excel e o R®. Os dados adquiridos no site Wyscout® vieram em planilhas de Excel, porém o tratamento, manipulação e análise serão feitos no software estatístico R®. A partir dessas análises, teremos uma visão sobre a relevância de cada ação em cada uma das posições de jogo no futebol. Por termos os dados de dez ligas diferentes, teremos base para comparação entre elas e descoberta de semelhanças ou diferenças da significância das ações de acordo com cada liga.

Considerações sobre os Termos de apresentação obrigatória:

Os pesquisadores propõe dispensa do TCLE, pois "Pois os atletas não serão abordados. Os dados foram obtidos pelo site wyscout.com o qual já tem autorização para extrair os dados estatísticos dos jogadores e disponibilizam para quem estiver interessado."

Recomendações:

não há.

Conclusões ou Pendências e Lista de Inadequações:

O CEP APROVA O PROTOCOLO DE PESQUISA.

Considerações Finais a critério do CEP:

O projeto encontra-se APROVADO para execução. Pedimos atenção aos seguintes itens:

1) De acordo com a Resolução CNS nº 466/12, o pesquisador deverá apresentar relatório final.

2) Eventuais emendas (modificações) ao protocolo devem ser apresentadas, com justificativa, ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada.

3) Sobre o TCLE: caso o termo tenha DUAS páginas ou mais, lembramos que no momento da sua assinatura, tanto o participante da pesquisa (ou seu representante legal) quanto o pesquisador responsável deverão RUBRICAR todas as folhas , colocando as assinaturas na última página.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1619063.pdf	08/09/2020 11:02:42		Aceito
Folha de Rosto	Folha_de_Rosto_Victor_Valvano.pdf	08/09/2020 11:01:42	Fabio Augusto Barbieri	Aceito
Projeto Detalhado / Brochura Investigador	ProjetoTCC_Victor_Valvano.docx	26/08/2020 12:58:20	Fabio Augusto Barbieri	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

RIO CLARO, 12 de Novembro de 2020

Assinado por:
Flávio Soares Alves
(Coordenador(a))

Victor Valvano Schimidt

Fabio Augusto Barbieri

Luiz Henrique Palucci Vieira

José Ângelo Barela