

# Desigualdades regionais e futebol: uma análise da primeira divisão do campeonato brasileiro entre 2003 e 2019

*Regional inequalities and football: an analysis of the first division of the Brazilian championship between 2003 and 2019*

Luiz Ricardo Cavalcante<sup>1</sup>, ORCID: <https://orcid.org/0000-0003-1894-7238>

1. Doutor em Administração pela UFBA (2004) com estágio de pesquisa na University of Illinois at Urbana-Champaign (UIUC). Mestre em Administração pela UFBA. Consultoria Legislativa do Senado Federal / IDP - Brasília – Brasil. E-mail: [luiz.ricardo.teixeira.cavalcante@gmail.com](mailto:luiz.ricardo.teixeira.cavalcante@gmail.com)

## Resumo

As desigualdades regionais que marcam o país refletem-se também no desempenho esportivo e, em particular, no futebol. Neste trabalho, empregam-se dados relativos ao desempenho das unidades da federação e das macrorregiões brasileiras na série A do campeonato brasileiro, entre 2003 e 2019, para analisar essas desigualdades. Os resultados indicam que o desempenho das macrorregiões brasileiras mantém uma forte correlação com o produto interno bruto (PIB). No conjunto das 15 unidades da federação que estiveram representadas na série A do campeonato brasileiro também se observa, de modo geral, uma forte correlação entre desempenho e PIB. A presença de eventuais *outliers* (região Sul e Rio de Janeiro, por exemplo) sugere que fatores adicionais (como a tradição e o número de torcedores) podem contribuir para explicar o desempenho das regiões no campeonato brasileiro, embora dificuldades de aferição e problemas de colinearidade com outras variáveis imponham dificuldades adicionais para sua inclusão nos modelos.

**Palavras-chave:** desigualdades regionais, futebol, unidades da federação, Brasil

## Abstract

Regional inequalities in Brazil are also related to sport performance and, in particular, to football (soccer) performance. In this work, we use data related to the performance of the Brazilian federative units and macro-regions in the first division of the local football championship between 2003 and 2019 to analyze these inequalities. Results indicate a strong correlation between the performance of the Brazilian macro-regions and their gross domestic product (GDP). Considering the 15 federative units that participated in at least one edition of the Brazilian first division this correlation is reaffirmed. The presence of occasional outliers (Southern region and Rio de Janeiro, for example) suggests that additional factors (such as tradition and number of fans) may contribute to explaining regional performance in the Brazilian football championship. Measurement issues and the collinearity of these factors with other indicators, however, pose additional difficulties to this kind of analysis.

**Keywords:** regional inequalities, football soccer, federative units, Brazil

**Citation:** Cavalcante, L. R. (2023). Desigualdades regionais e futebol: uma análise da primeira divisão do campeonato brasileiro entre 2003 e 2019. *Gestão & Regionalidade*, 39, e20237996. <https://doi.org/10.13037/gr.vol39.e20237996>



## 1 Introdução

O Brasil é amplamente reconhecido por suas elevadas desigualdades regionais. Há mais de meio século Williamson (1965, pp. 111-115) já se referia ao país como aquele “com o mais largamente discutido problema Norte-Sul” e “universalmente citado pela sua severa cisão Norte-Sul”. Da mesma forma, Baer (2008, p. 243) destaca que a desigualdade da distribuição geográfica de renda tem sido uma característica da economia brasileira desde os tempos coloniais. Embora eventuais movimentos de desconcentração tenham sido observados ao longo das últimas décadas, as desigualdades regionais que marcam o país têm sido recorrentemente reafirmadas, tanto em publicações de caráter acadêmico como no debate político (Monteiro Neto, 2014; Figueiredo; Porto Junior, 2015; Gimenez, 2017).

As desigualdades regionais aferidas com base em indicadores econômicos estão fortemente associadas a outros indicadores sociais como educação e saúde. Embora não seja trivial estabelecer relações de causalidade entre essas variáveis, há recorrentes evidências de profundas desigualdades regionais na educação, na educação superior e nos indicadores de mortalidade infantil, por exemplo. Desse modo, o Índice de Desenvolvimento Humano (IDH) – que resulta de uma ponderação de indicadores de renda, educação e saúde – também se revela bastante desigual no país (Mendes *et al.*, 2020).

Contudo, análises a respeito das desigualdades regionais no Brasil dificilmente envolvem o desempenho esportivo. No caso específico do futebol, por exemplo, a intuição sugere que as equipes do Sul e do Sudeste tendem a colecionar mais títulos, mas não há uma análise sistemática dos indicadores relacionados ao desempenho das regiões – ou das equipes de cada macrorregião ou unidade da federação, por exemplo – e de sua eventual correlação com as desigualdades regionais que marcam o país. Trata-se de um tema ainda pouco explorado no Brasil, mas que tem sido objeto de publicações fora do país, seja sobre desigualdades no desempenho nos jogos olímpicos (Reiche, 2016), seja sobre o desempenho de equipes de futebol em diferentes regiões (Garcia *et al.*, 2007; Doran & Jordan, 2018; Jones & Jordan, 2019). De modo geral, essas análises têm um caráter exploratório – inclusive em função de sua contemporaneidade – e multidisciplinar. Nos termos de Garcia *et al.* (2007, p. 83), “a geografia do esporte, de modo geral, e a geografia do futebol, em particular, podem ser analisadas usando conceitos encontrados em uma ampla variedade de disciplinas, que incluem visões culturais, históricas, econômicas, demográficas, urbanas e políticas”.

O objetivo deste trabalho é analisar as desigualdades regionais no futebol do país com base nos dados relativos ao desempenho das unidades da federação e das macrorregiões brasileiras na série A do campeonato brasileiro entre 2003 e 2019. O trabalho está segmentado em mais quatro seções além desta introdução. Na seção 2, apresenta-se uma breve revisão bibliográfica da relação entre indicadores econômicos e desempenho esportivo. Na terceira seção, descrevem-se os dados e os métodos de análise utilizados para aferir as desigualdades regionais no futebol brasileiro. Os resultados são o objeto da seção 4. Finalmente, na quinta seção apresentam-se as principais conclusões do trabalho.

## 2 Referencial Teórico

Há uma ampla produção bibliográfica sobre desigualdades regionais no Brasil abordando aspectos econômicos como também sociais (Monteiro Neto, 2014; Figueiredo; Porto Junior, 2015; Gimenez, 2017, Mendes *et al.*, 2020). Além das análises dedicadas às diferenças de PIB *per capita*, há também trabalhos em que se analisam, por exemplo, as desigualdades



relacionadas ao acesso a serviços de saúde ou à internet (Albuquerque *et al.*, 2017; Barbosa, 2018). Contudo, no que diz respeito ao desempenho esportivo, de modo geral, e ao futebol, em particular, não foram identificadas publicações disponíveis. Análises preditivas do desempenho dos países (obviamente representados por uma única equipe) nas copas do mundo de futebol amparam-se, de modo geral, em características das próprias equipes (por exemplo, seu desempenho no período imediatamente anterior ao certame analisado ou sua indicação como cabeça de chave) e não em variáveis relacionadas aos países que representam (Monks & Husch, 2009).

Por essa razão, recorreu-se, para amparar as análises apresentadas neste trabalho, às publicações que apontam os fatores explicativos do desempenho esportivo dos países em olimpíadas e a uma bibliografia razoavelmente específica sobre os fatores associados ao desempenho das equipes de futebol. Essas duas abordagens – brevemente discutidas nas seções 2.1 e 2.2 a seguir – amparam a definição dos procedimentos metodológicos e a análise dos resultados do presente artigo.

## 2.1 Desempenho dos países em olimpíadas

Ainda que se enfatize, de modo geral, a relevância de competir, e não de ganhar, o sucesso de atletas nacionais em jogos olímpicos é obviamente uma fonte de prestígio para os países (Bian, 2005). Por essa razão, diversos autores têm buscado identificar as razões pelas quais alguns países têm desempenhos superiores aos de outros.

Há cerca de cinquenta anos, Novikov e Maximenko (1972) analisaram as correlações entre variáveis socioeconômicas e o desempenho dos países participantes dos jogos olímpicos de Tóquio em 1964. Os autores constataam que “as conquistas dos países socialistas nos jogos olímpicos são marcadamente superiores ao sucesso esportivo dos países capitalistas” (ibid., p. 38). Novikov e Maximenko (1972, p. 39) indicam as variáveis que exibiram correlações positivas com o desempenho esportivo. Essas variáveis, em ordem decrescente da magnitude dos coeficientes de correlação, são: *i)* renda *per capita*; *ii)* calorias ingeridas pela população; *iii)* expectativa de vida média; *iv)* taxa de analfabetismo; *v)* taxa de urbanização; *vi)* número de habitantes. Com relação ao número de habitantes, contudo, os autores não apontam a existência de “dependência significativa”. Embora ainda limitado do ponto de vista dos métodos de análise estatística que emprega, o trabalho de Novikov e Maximenko (1972) já aponta o PIB (que corresponde ao produto do PIB *per capita* pela população) como a principal variável associada ao desempenho em olimpíadas. Mais ou menos na mesma época, Grimes, Kelly e Rubin (1974) constroem um modelo socioeconômico de desempenho nos jogos olímpicos para verificar se países comunistas teriam resultados superiores aos dos demais países ao se levarem em conta aspectos econômicos e demográficos. Os autores concluem que os países comunistas efetivamente exibiam desempenhos superiores aos dos demais países mesmo após se controlar por outras variáveis.

Com base em uma lógica semelhante, Bernard e Busse (2004) desenvolveram um modelo para prever o número de medalhas obtidas por um país. Os autores consideram população, PIB *per capita*, uma *dummy* para o país sede, uma *dummy* para países sob a influência soviética e uma *dummy* para economias planejadas (eq. 1).

$$M_{it} = C + \alpha \ln N_{it} + \beta \ln \left( \frac{Y}{N} \right)_{it} + Host_{it} + Soviet_{it} + Planned_{it} + d_t + \tau_i + \varepsilon_{it} \text{ (eq. 1)}$$



Na equação 1,  $\tau_i$  é uma variável que captura as características idiossincráticas de cada país e  $\varepsilon_{it}$  é um termo de erro com distribuição normal.

Bernard e Busse (2004) concluem que todos esses fatores são significativos, mas destacam que “renda per capita e população têm efeitos idênticos na margem, sugerindo que o PIB total é o melhor preditor da performance olímpica” (Bernard & Busse, 2004, p. 417).

Johnson e Ali (2004) analisam a participação e o desempenho em jogos olímpicos de inverno e de verão. Suas conclusões são, em grande medida, intuitivas e semelhantes às de Bernard e Busse (2004). Contudo, ressaltam que a participação nos jogos de inverno é, em relação aos jogos de verão, mais afetada por indicadores de renda, menos afetada por indicadores de população, menos enviesada em favor do país sede e mais afetada pelo clima. Apesar dessas eventuais diferenças entre os dois tipos de eventos, Johnson e Ali (2004) concluem que variáveis socioeconômicas explicam bastante bem a participação e o desempenho dos países em jogos olímpicos.

Bian (2005, p. 43) reafirma as conclusões de Bernard e Busse (2004) e de Johnson e Ali (2004): “variáveis socioeconômicas, incluindo população, recursos econômicos, vantagem de sediar [o evento] e estrutura política têm impacto significativo no desempenho olímpico dos países”. Pfau (2006) usa um modelo semelhante (exceto pela inclusão de *dummies* para países escandinavos, germânicos, alpinos e norte-americanos) para prever o desempenho dos países nos jogos olímpicos de inverno de 2006 de Turim. Mais uma vez, confirma-se o papel central das variáveis demográficas e econômicas no desempenho em jogos olímpicos. Bernard (2008), em um trabalho mais simples, aplica o método desenvolvido por Bernard e Busse (2004) para prever o desempenho dos países nos jogos olímpicos de Pequim.

Rathke e Woitek (2008) usam um método de fronteira estocástica para avaliar a “eficiência” dos países produzirem medalhas levando em conta os múltiplos recursos de que dispõem (PIB, população e *dummies* para economias planejadas ou para países que sediaram os jogos, por exemplo). De maneira análoga, Wu, Liang e Yang (2009), embora limitando as variáveis explicativas ao PIB *per capita* e à população, aplicam um modelo *data envelopment analysis (DEA)* para aferir o desempenho dos países em seis edições dos jogos olímpicos entre 1984 e 2004. Com isso, os autores obtêm um ranking de países que, na prática, captura sua eficiência ao converter os “insumos” (isto é, as variáveis explicativas) em resultados (isto é, medalhas). Jablonsky (2018) também usa um modelo *DEA* em dois estágios, mas, em lugar de recorrer às variáveis “tradicionais” (“população, força econômica aferida pelo produto interno bruto e tradição esportiva”), busca avaliar o desempenho dos países “com respeito aos recursos que eles podem gastar”.

Reiche (2016) consolidou diversas análises dessa natureza em um livro intitulado “*Success and Failure of Countries at the Olympic Games*”. O autor segmenta os fatores explicativos em dois grandes blocos: características gerais e políticas públicas. As características gerais envolvem ideologia, riqueza, população e geografia, ao passo que as políticas públicas envolvem esportes femininos (“*women’s sports*”), institucionalização, especialização e aprendizado desde cedo (“*early learning*”). Ao longo da publicação, o autor refina várias análises mencionadas nesta seção. Por exemplo, em relação à riqueza, Reiche (2016, p. 11) indica que “muitos estudos enfatizam a correlação do PIB com o sucesso nas olimpíadas”, mas argumenta que “a importância da riqueza depende do tipo de esporte e do tipo dos jogos”, uma vez que “a riqueza do país é mais relevante nos jogos de inverno do que nos jogos de verão, porque os jogos de inverno são mais dependentes de infraestruturas mais caras”. Já a população se revela mais relevante na determinação do desempenho nos jogos de verão do que nos jogos de inverno.

Assim, de modo geral, há bastante convergência entre os vários autores citados nesta seção quanto ao papel central das variáveis demográficas e econômicas no desempenho dos



países em olimpíadas. É claro que, nesse caso, ao contrário das equipes de futebol (que podem contratar jogadores de outras regiões, por exemplo) a dependência da população total (como uma espécie de *proxy* da capacidade de fornecimento de atletas) pode desempenhar um papel distinto nos resultados.

## 2.2 Desempenho das equipes de futebol

Ao contrário da seção precedente – cujo foco de análise recai sobre o desempenho de países envolvidos nas múltiplas competições dos jogos olímpicos – esta seção trata do desempenho das equipes de futebol individualmente. O propósito aqui é verificar se a bibliografia disponível apontaria alguma variável local capaz de explicar esse desempenho e de, assim, afetar eventuais desigualdades regionais relativas a esse aspecto.

Há uma literatura relativamente vasta sobre os fatores que influenciam o desempenho de equipes de futebol. O foco pode recair tanto sobre seleções nacionais como sobre agremiações.

Em uma parte das análises sobre seleções nacionais, as variáveis explicativas privilegiam o desempenho precedente da equipe. Robberechts e Davis (2018), por exemplo, usam o desempenho das seleções nacionais nas copas do mundo entre 2002 e 2014 para prever os resultados da copa do mundo de 2018. Outros autores recorrem a parâmetros análogos àqueles usados na análise do desempenho nos jogos olímpicos. Hoffmann, Ging, Ramasamy, (2002), por exemplo, buscaram identificar as variáveis explicativas para as pontuações obtidas pelas seleções nacionais no ranking da *Fédération Internationale de Football Association (FIFA)* em janeiro de 2001 e constataram que aspectos econômicos, demográficos, culturais e climáticos exercem forte influência nos resultados. Em particular, o produto nacional bruto (PNB) *per capita*, a temperatura média anual, a participação da população do país na população mundial, uma *dummy* para países sede desde a copa do mundo de 1930 e a combinação de uma *dummy* para países latinos com a população se revelaram significativos nas regressões estimadas.

No caso dos clubes (que, em geral, podem ter acesso a jogadores de diferentes nacionalidades), a disponibilidade de recursos para contratações, para a formação de jogadores e para treinamentos é também uma variável explicativa do desempenho. Frick (2007), ao analisar o mercado de trabalho de jogadores de futebol dos principais campeonatos europeus, registra os valores de mercado das equipes da primeira divisão na Inglaterra, na Itália, na Espanha, na Alemanha e na França nas temporadas de 2005/2006 e de 2006/2007. Os valores de transferência agregados situam-se entre € 2,33 bilhões (Inglaterra, relativo à temporada 2006/2007) e € 1,08 bilhão (França, relativo à temporada anterior). Esses valores sugerem que, em tese, regiões mais ricas (ou, pelo menos, equipes que mobilizam mais recursos de seus torcedores ou dos interessados em suas partidas) tenderiam, *ceteris paribus*, a contar com melhores jogadores e com estruturas mais robustas e a apresentar desempenhos superiores. Ao analisarem as equipes da “série A” do campeonato italiano entre as temporadas de 2007-2008 e 2013-2014, Ferri *et al.* (2017) concluem que o desempenho é positivamente associado com o salário dos jogadores. Os autores, porém, não identificam um efeito significativo dos valores das transferências no desempenho das equipes.

A análise da eficiência dos clubes ao converterem recursos em resultados foi objeto de alguns trabalhos. Ribeiro e Lima (2012) usam um modelo *DEA* para aferir a eficiência das equipes portuguesas na primeira divisão e Zambom-Ferraresi, García-Cebrián e Lera-López (2017) analisam a *UEFA Champions League* ao longo de múltiplas temporadas. Embora o foco dessas análises seja a gestão dos clubes (isto é, sua capacidade de converter insumos em



resultados), é evidente que, como regra, equipes de regiões mais ricas e mais populosas tendem a mobilizar mais recursos e a obter melhores desempenhos.

Em linha com essa percepção, Garcia *et al.* (2007, p. 83) testam “a influência relativa dos principais fatores demográficos e econômicos que determinam o sucesso internacional no futebol de uma amostra de cidades europeias”. Sua hipótese central é que a capacidade das cidades apoiarem as equipes locais é uma função do tamanho e da riqueza de sua população. Omondi-Ochieng (2015) examina a associação entre PNB *per capita*, contratação de profissionais de futebol (inclusive jogadores) e desempenho (aferido com base na classificação para campeonatos continentais). Os resultados indicam que o PNB *per capita* é positivo e significativamente associado à contratação de profissionais de futebol, mas não é um bom preditor da performance. Omondi-Ochieng (2015) argumenta que “gestores e profissionais do futebol se mostraram os preditores centrais da performance no futebol, e não o PNB *per capita*”. Desse modo, o PNB *per capita* influenciaria a performance de forma indireta.

O conceito de economias de aglomeração (Combes, Duranton & Gobillon, 2011) tem também sido usado por alguns autores para analisar o desempenho das equipes de futebol. De fato, Bäker, Mechtel e Vetter (2012) indicam que *derbies* entre equipes da mesma cidade ou da mesma região atraem mais atenção do público do que partidas “normais”, mesmo que seus indicadores objetivos (por exemplo, número de gols marcados) não sejam estatisticamente distintos. Doran e Jordan (2018) analisam os efeitos da proximidade e da rivalidade (regional) no desempenho das equipes da primeira divisão do campeonato inglês. Os autores analisam as equipes participantes em 21 edições com base em técnicas de econometria espacial e concluem que há evidências de dependência espacial positiva entre os desempenhos. Isso implica dizer que a presença de equipes melhores em uma determinada região termina afetando positivamente o desempenho das demais equipes daquela região. Da mesma forma, autores como Jones e Jordan (2019, p. 432) estendem o conceito de economias de aglomerações ao desempenho esportivo. Os autores argumentam que há uma vasta literatura que mostra que “as empresas se aglomeram mais do que seria esperado e se beneficiam ao fazê-lo”. Da mesma forma, o desempenho das equipes da *English Premier League (EPL)* é afetado pela escala urbana. Jones e Jordan (2019) argumentam, assim, que “equipes em grandes conurbações têm uma performance relativamente superior, e sugere-se que esses benefícios provavelmente têm origem no novo mercado de trabalho esportivo”. Desse modo, a presença de rivais locais (viabilizados pela escala urbana ou metropolitana, por exemplo) pode contribuir para o desempenho de uma equipe na medida em que contribui para a oferta de mão de obra (inclusive de jogadores) e estabelece padrões locais de competição que acabam beneficiando as equipes em certames nacionais.

### 3 Metodologia

Nesta seção, descrevem-se os dados e os métodos de análise utilizados para aferir as desigualdades regionais no futebol brasileiro.

#### 3.1 Dados utilizados

Os dados de PIB e de população dos municípios, unidades da federação e macrorregiões foram calculados para o ano de 2018, que corresponde ao último disponível nesse nível de agregação no site do Instituto Brasileiro de Geografia e Estatística (IBGE).



Os dados relativos ao desempenho (afetados com base nas participações e na pontuação acumulada) das equipes da série A do campeonato brasileiro foram tabulados ano a ano para o período em que os certames foram disputados no formato de pontos corridos (entre 2003 e 2019). Em seguida, os dados foram consolidados para se obterem as participações e as pontuações acumuladas no período (tabela 1). É claro que a pontuação acumulada pode ser afetada pelo comportamento das equipes que, nas últimas rodadas não têm incentivos para acumular mais pontos (quando suas chances de título, de acesso a outras competições ou de rebaixamento já estão definidas). Uma alternativa seria usar, em lugar da pontuação acumulada, indicadores ordinais (isto é, a colocação obtida no conjunto das equipes). Essa opção, contudo, não seria capaz de capturar desempenhos excepcionalmente bons ou ruins. Por essa razão, optou-se por empregar, neste trabalho, as participações ao lado da pontuação acumulada na série A do campeonato brasileiro.

Conforme se evidenciou na seção 2, outros fatores podem explicar o desempenho das equipes e das regiões onde estão sediadas em campeonatos de futebol. Esses fatores podem envolver:

- Tradição, aferida por uma medida do desempenho progresso da equipe ou da região em campeonatos de futebol.
- Número de torcedores, que afeta a capacidade de mobilizar recursos financeiros e, desse modo, de contratar recursos humanos (inclusive jogadores).
- Rivalidades regionais, que podem, por meio das economias de aglomeração, afetar positivamente o desempenho das equipes locais.
- Orçamento disponível, decorrente tanto da capacidade de captação de recursos (sócios, público pagante em estádios, cotas de televisão e patrocínios, por exemplo) como das despesas (folha salarial, despesas com infraestrutura de treinamento e de campos próprios, por exemplo).
- Recursos públicos investidos em esportes e políticas de incentivos (por exemplo, equipamentos públicos como estádios e renúncias fiscais).

Contudo, a aferição desses dados é, de modo geral, controversa. No caso da tradição, ainda que se trate de um conceito em certa medida intuitivo para aqueles que acompanham o futebol – Macmillan e Smith (2007) chegaram a apontar o papel de uma história longa no futebol internacional na determinação dos rankings –, não se identificaram bases quantitativas consolidadas que pudessem ser usadas nas regressões. Quanto ao número de torcedores, institutos de pesquisa (como o Ibope ou o Datafolha) divulgam apenas séries ocasionais e baseadas em amostragens. Não há, desse modo, uma metodologia unificada e os rankings podem exibir diferenças significativas. Questões como o perfil dos torcedores de equipes de abrangência nacional (distinto daquele de torcedores de equipes regionais) precisariam ser também levadas em consideração para a eventual inclusão dessa variável em modelos de regressão. Isso ocorre porque esse perfil está associado tanto aos níveis de renda como ao grau de envolvimento dos torcedores com a equipe e afeta a forma como o número de torcedores interfere na capacidade de captação de recursos. No caso das rivalidades regionais, embora o conceito seja também intuitivo, sua definição, na prática, requeria o enfrentamento de aspectos como a própria definição de “região” (estados, regiões metropolitanas ou municípios). Definições desse tipo afetariam, por exemplo, em que medida equipes da baixada santista teriam rivais regionais na cidade de São Paulo. Os orçamentos dos clubes, por sua vez, não são amplamente disponíveis e, mesmo que fossem, seriam também sujeitos a controvérsias quanto a suas formas de contabilização. Finalmente, não há dados consolidados sobre recursos públicos



investidos em esportes e políticas de incentivos em nível subnacional. De qualquer forma, se disponíveis, esses dados dificilmente divergiriam significativamente de outros indicadores econômicos mais gerais (como o PIB ou o PIB *per capita*).

Ainda que fosse possível obter dados sobre esses fatores adicionais, sua inclusão traria, em vários casos, problemas de colinearidade com o próprio desempenho que limitaria suas possibilidades de uso em regressões econométricas. Por exemplo, a tradição é, em alguma medida, também uma métrica do desempenho progressivo. De qualquer modo, a ausência desses dados obviamente limita as possibilidades da análise das desigualdades regionais no futebol brasileiro e sua eventual disponibilidade futura permitiria um refinamento dos modelos usados neste trabalho. Neste momento, porém, em virtude das limitações apontadas, optou-se por empregar apenas dados demográficos e econômicos. Esse procedimento, apesar das ressalvas apresentadas, alinha-se com diversas referências apontadas na seção 2, conforme se detalha a seguir.

### 3.2 Métodos de análise

Entre os artigos mencionados na seção 2.1, o método mais frequente de análise é baseado em regressões lineares em que uma variável dependente (participações em eventos esportivos ou desempenho nesses eventos) é associada a um conjunto de variáveis independentes, com ênfase em variáveis demográficas e econômicas. Da mesma forma, alguns trabalhos citados na seção 2.2 – como Garcia *et al.* (2007) e Omondi-Ochieng (2015), por exemplo – correlacionam variáveis demográficas e econômicas de jurisdições subnacionais ao desempenho de equipes de futebol nelas sediadas.

Ao se levarem em conta o desempenho, as variáveis demográficas (população), econômicas (PIB) e os demais aspectos mencionados na seção 3.1 (tradição, número de torcedores, presença de rivalidades regionais, orçamento e recursos públicos), um modelo geral (que corresponderia à “base teórica” do presente trabalho) poderia ser:

$$Desempenho_{it} = \beta_{02} + \beta_1 \ln(população)_{it} + \beta_2 \ln(PIB)_{it} + \beta_3 tradição_{it} + \beta_4 \ln(torcedores)_{it} + \beta_5 rivalidades_{it} + \beta_6 \ln(orçamento público)_{it} + \tau_i + \varepsilon_{it} \quad (eq. 2)$$

Nesse modelo, os subscritos *i* e *t* representam a jurisdição (macrorregião ou unidade da federação, por exemplo) e o período (temporada ou temporadas acumuladas). As características idiossincráticas de cada jurisdição são capturadas por  $\tau_i$  e  $\varepsilon_{it}$  é o termo de erro com distribuição normal.

Uma série de dificuldades práticas se interpõe à aplicação de um modelo dessa natureza às macrorregiões ou às unidades da federação no Brasil:

- O número reduzido de observações (cinco, no primeiro caso, e 27, no segundo), de modo que as regressões teriam reduzida significância estatística.
- Esse problema é agravado pela presença de um número razoavelmente elevado de variáveis explicativas (que reduzem os graus de liberdade);
- A presença de um número significativo de observações para as quais a variável dependente é zero (no caso dos estados que não estiveram representados na série A do campeonato brasileiro no período considerado).



- Conforme se argumentou acima, a inclusão de variáveis explicativas adicionais implicaria problemas relacionados à qualidade dos dados e a sua colinearidade uns com os outros e com o próprio desempenho.

Por essas razões, optou-se, neste trabalho, por privilegiar análises mais simples nas quais as correlações entre as variáveis são exploradas por meio de um conjunto de gráficos, tendo-se evitado regressões cuja significância estatística poderia ser questionada. Nesse sentido, embora a equação 2 sintetize a base teórica adotada neste trabalho, as regressões aqui adotadas são regressões simples, e não regressões múltiplas.

É evidente que essas análises relativamente simples podem ser refinadas em trabalhos futuros, tanto pela inclusão de fatores explicativos adicionais como pelo uso de unidades de análise que aumentem o número de observações (por exemplo, municípios ou regiões metropolitanas).

## 4 Resultados

As participações e a pontuação acumulada pelas equipes da série A do campeonato brasileiro apuradas conforme os procedimentos descritos na seção 3.1 estão indicadas na tabela 1.

**Tabela 1** – Desempenho (participações e pontuação acumulada) das equipes da série A do campeonato brasileiro (2003-2019)

Equipe	Participações (2003-2019)	Pontuação acumulada (2003-2019)	Município
América de Natal	1	17	Natal
América Mineiro	3	105	Belo Horizonte
Athletico Paranaense	16	902	Curitiba
Atlético Goianiense	4	156	Goiânia
Atlético Mineiro	16	904	Belo Horizonte
Avaí	6	236	Florianópolis
Bahia	8	371	Salvador
Botafogo	15	787	Rio de Janeiro
Brasiliense	1	41	Brasília
Ceará	4	169	Fortaleza
Chapecoense	6	272	Chapecó
Corinthians	16	986	São Paulo
Coritiba	12	615	Curitiba
Criciúma	4	188	Criciúma
Cruzeiro	17	1.030	Belo Horizonte
CSA	1	32	Maceió
Figueirense	11	550	Florianópolis
Flamengo	17	1.009	Rio de Janeiro
Fluminense	17	937	Rio de Janeiro
Fortaleza	4	195	Fortaleza
Goiás	12	648	Goiânia
Grêmio	16	963	Porto Alegre
Grêmio Barueri (1)	1	49	Barueri
Grêmio Prudente (1)	1	28	Presidente Prudente
Guarani	3	147	Campinas
Internacional	16	975	Porto Alegre



Equipe	Participações (2003-2019)	Pontuação acumulada (2003-2019)	Município
Ipatinga	1	35	Ipatinga
Joinville	1	31	Joinville
Juventude	5	266	Caxias do Sul
Náutico	5	200	Recife
Palmeiras	15	902	São Paulo
Paraná	6	304	Curitiba
Paysandu	3	146	Belém
Ponte Preta	9	432	Campinas
Portuguesa	3	127	São Paulo
Santa Cruz	2	56	Recife
Santo André	1	41	Santo André
Santos	17	1.043	Santos
São Caetano	4	215	São Caetano do Sul
São Paulo	17	1.103	São Paulo
Sport	9	417	Recife
Vasco da Gama	14	726	Rio de Janeiro
Vitória	10	468	Salvador

Nota 1: o Grêmio Barueri e o Grêmio Prudente foram tabulados como equipes distintas.

Fonte: elaborado pelo autor.

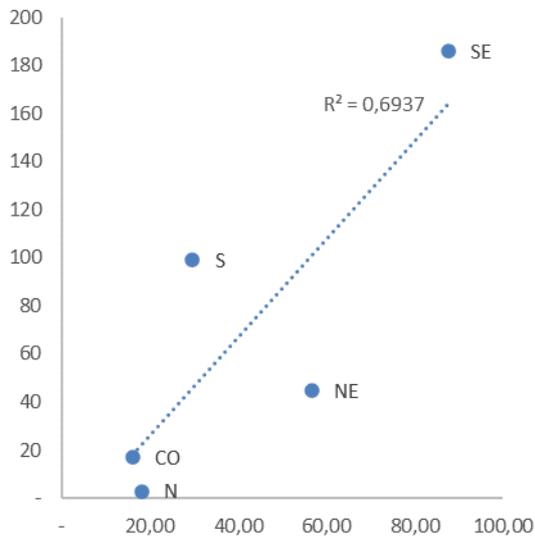
Conforme se pode observar, considerando as participações no período, contabilizaram-se 43 equipes sediadas em 25 municípios e em 15 unidades da federação. Todas as macrorregiões brasileiras estão representadas na amostra.

Conforme se indicou na seção precedente, regressões com apenas cinco macrorregiões não seriam minimamente robustas. Foram feitos também exercícios com regressões para as 27 unidades da federação usando apenas o PIB e a população como variáveis explicativas (uma vez que as demais variáveis indicadas na equação 2 não estavam disponíveis). Testaram-se modelos de regressão usando mínimos quadrados ordinários (MQO) e modelos lineares generalizados (MLG) com distribuições de Poisson e binomial negativa. Embora em alguns casos coeficientes positivos e significativos tenham sido obtidos (especialmente no caso da população), pelas razões expostas no final da seção 3.2 optou-se por privilegiar análises mais simples nas quais as correlações entre as variáveis são exploradas em gráficos.

As figuras 1, 2 e 3 relacionam o número total de participações das cinco macrorregiões brasileiras na série A do campeonato brasileiro entre 2003 e 2019 com a população, o PIB *per capita* e o PIB, respectivamente. A figura 4, por sua vez, relaciona a pontuação acumulada no mesmo período com o PIB.

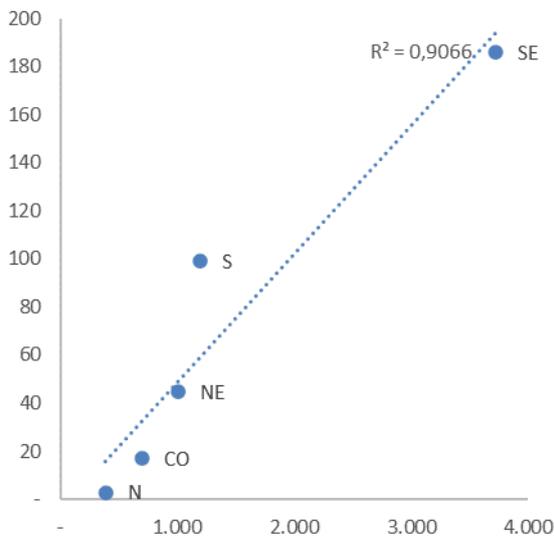


**Figura 1** – Participações na série A do campeonato brasileiro (2003-2019) e população (milhões de habitantes em 2018), macrorregiões brasileiras



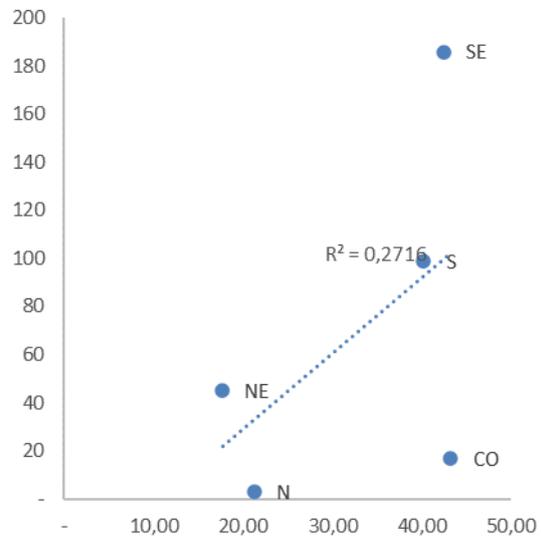
Fonte: elaborado pelo autor.

**Figura 3** – Participações na série A do campeonato brasileiro (2003-2019) e PIB (R\$ bilhões correntes, 2018), macrorregiões brasileiras



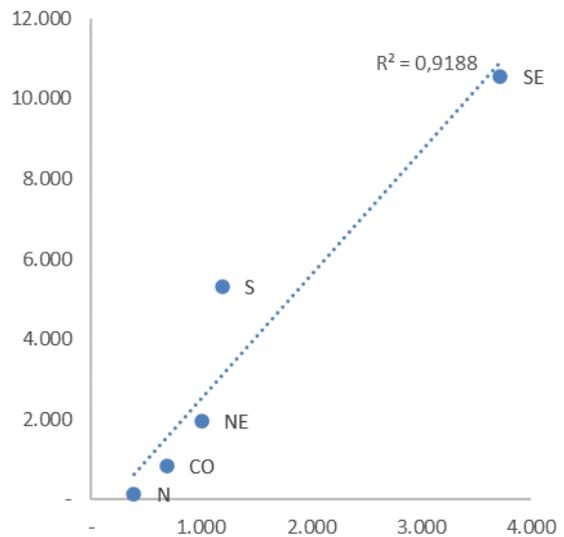
Fonte: elaborado pelo autor.

**Figura 2** – Participações na série A do campeonato brasileiro (2003-2019) e PIB per capita (R\$ mil correntes por habitante, 2018), macrorregiões brasileiras



Fonte: elaborado pelo autor.

**Figura 4** – Pontuação acumulada na série A do campeonato brasileiro (2003-2019) e PIB (R\$ bilhões correntes, 2018), macrorregiões brasileiras



Fonte: elaborado pelo autor.

Embora obviamente simplificadas, as figuras já apontam para algumas conclusões de natureza metodológica e relativas às desigualdades regionais no futebol brasileiro.

As figuras indicam que, ao se considerarem as participações, a população (figura 1) e o PIB *per capita* (figura 2) exibem ajustes inferiores àquele que se obtém ao se usar o PIB (figura 3). Além disso, a pontuação acumulada (figura 4) exibe um ajuste um pouco superior àquele decorrente do uso das participações acumuladas (figura 3). Os valores de  $R^2$  obtidos nesses dois casos (0,9066 e 0,9188) confirmam essa percepção. A correlação entre desempenho e PIB é consistente com as regressões mencionadas na seção 2.1, segundo as quais os melhores ajustes para a performance em jogos olímpicos foram obtidos ao se combinarem população e PIB *per*

*capita* (uma vez que, obviamente,  $\text{PIB} = \text{população} \times \text{PIB per capita}$ ). Conforme se evidenciou naquela seção, diversos autores apontam para a existência de uma correlação positiva entre PIB e desempenho nas olimpíadas (Novikov & Maximenko, 1972; Bernard & Busse, 2004; Bian, 2005; Reiche, 2016; Jablonsky, 2018). Os resultados reportados neste trabalho são também consistentes com alguns trabalhos mencionados na seção 2.2, que trata do desempenho das equipes de futebol. Em particular, destaca-se sua convergência com os trabalhos de Garcia *et al.* (2007) e com aqueles que estendem o conceito de economias de aglomeração ao desempenho no futebol.

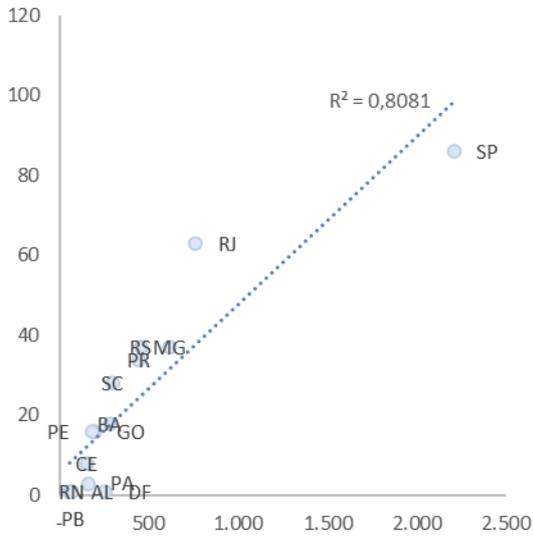
Além das conclusões de natureza metodológica, as figuras confirmam a percepção de que há uma correlação entre o PIB e as participações (figura 3) ou a pontuação (figura 4) acumuladas das macrorregiões brasileiras na série A do campeonato brasileiro. Esse resultado, em linhas gerais, é consistente com os trabalhos mencionados na seção 2 deste artigo. É claro que essa é a tendência geral. Ao se analisarem as macrorregiões individualmente, observa-se que quatro delas estão “sub-representadas” (isto é, situadas abaixo da linha de tendência), ao passo que a região Sul exibe um desempenho cerca de 40% superior ao que seria esperado considerando seu PIB. Isso sugere que outros fatores podem explicar seu desempenho na série A do campeonato brasileiro. Esses fatores poderiam envolver, por exemplo, a tradição, o número de torcedores e rivalidades regionais (que, conforme se argumentou na seção 2.2, podem afetar positivamente o desempenho das equipes locais). Esses aspectos poderiam servir, de igual modo, para explicar a sub-representação das regiões Norte e Centro-Oeste. Trata-se, nos dois casos, de regiões que, de modo geral, têm equipes menos tradicionais, com menos torcedores e que, em muitos casos, carecem de rivais regionais que estimulem um desempenho superior. Assim, de modo geral, os desvios da tendência apontada nas figuras 3 e 4 poderiam ser atribuídos a uma espécie de causalidade circular e cumulativa (Myrdal, 1957): regiões que têm melhor desempenho no futebol tendem a se tornar mais tradicionais, a mobilizar mais torcedores e a favorecer rivalidades locais e, com isso, a melhorar seu desempenho. Por outro lado, nas regiões onde esses fatores estão ausentes, o desempenho tende a se reduzir e, com isso, a limitar a ocorrência dos fatores que poderiam melhorá-lo.

A eventual inclusão desses fatores, contudo, é limitada porque *i)* não há métricas consagradas de tradição, do número de torcedores e das rivalidades regionais que permitiriam seu uso em regressões econométricas; e *ii)* ainda que os dados estivessem imediatamente disponíveis, seria preciso lidar com problemas de colinearidade. Por essas razões, não se buscou, neste trabalho, incluir variáveis adicionais no modelo proposto.

Ao se estender a análise as 15 unidades da federação representadas na amostra, obtêm-se os gráficos 5 e 6, em que se correlacionam as participações e a pontuação acumulada com o PIB.

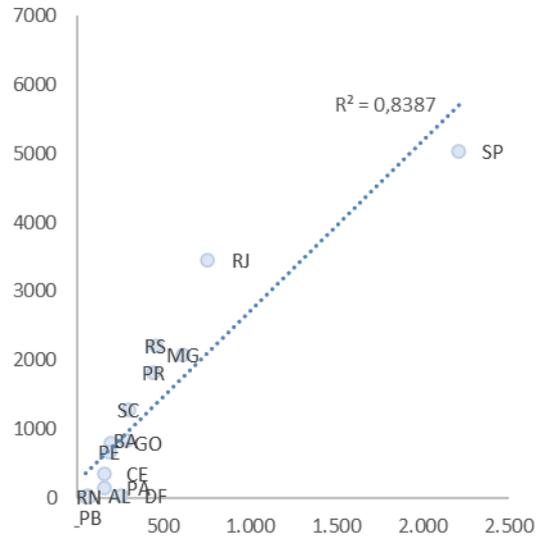


**Figura 5** – Participações na série A do campeonato brasileiro (2003-2019) e PIB (R\$ bilhões correntes, 2018), unidades da federação representadas na série A no período



Fonte: elaborado pelo autor.

**Figura 6** – Pontuação acumulada na série A do campeonato brasileiro (2003-2019) e PIB (R\$ bilhões correntes, 2018), unidades da federação representadas na série A no período



Fonte: elaborado pelo autor.

Os resultados indicados nas figuras 5 e 6 são em grande medida intuitivos. Em primeiro lugar, reafirmam a percepção de que o PIB é, de modo geral, um bom preditor para a participação e o desempenho na série A do campeonato brasileiro. Essa proposição é obviamente limitada ao conjunto das 15 unidades da federação representadas na amostra. Trata-se de um procedimento semelhante àquele adotado por Bernand e Busse (2004), que limitaram sua análise aos países que participaram dos jogos olímpicos. De todo modo, esse conjunto de 15 unidades da federação representou 88,60% do PIB do país em 2018. Em ambos os gráficos, destacam-se os estados da região Sudeste (com exceção do Espírito Santo) e os três estados da região Sul. As demais unidades da federação aparecem na parte inferior à esquerda do gráfico. De qualquer modo, a correlação entre desempenho das unidades da federação na série A do campeonato brasileiro e PIB é consistente – assim como no caso das macrorregiões – com diversos trabalhos analisados na seção 2 deste artigo.

Em termos absolutos, o principal *outlier* dos gráficos 5 e 6 é o Rio de Janeiro, cujos indicadores são superiores àqueles que seriam preditos com base em seu PIB. Os mesmos fatores adicionais mencionados anteriormente concorrem para explicar esse comportamento. Além disso, por ter abrigado a capital do país até 1960 (isto é, no período em que o futebol se tornou mais popular), o Rio de Janeiro acabou sediando equipes que mobilizaram o interesse de torcedores também de fora do estado, o que pode contribuir para a amplificação dos fatores extraeconômicos que concorrem para seu desempenho. Curiosamente, em São Paulo – onde muitos fatores extraeconômicos estariam presentes no caso de suas equipes mais bem posicionadas – o desempenho observado é inferior àquele que seria esperado, considerando seu PIB. Nesse caso, a elevada concentração econômica naquele estado não se traduziu plenamente em desempenho na série A do campeonato brasileiro entre 2003 e 2019.

## 5 Considerações Finais

Neste trabalho, exploraram-se as desigualdades regionais no futebol do país com base nos dados relativos ao desempenho das unidades da federação e das macrorregiões brasileiras na série A do campeonato brasileiro entre 2003 e 2019. Uma breve revisão bibliográfica indicou que o desempenho esportivo em geral é fortemente correlacionado com indicadores demográficos e econômicos. Essa percepção foi confirmada em trabalhos que analisam o desempenho dos países em olimpíadas. No caso do desempenho no futebol, muitas análises amparam-se no desempenho precedente das equipes e não em indicadores das regiões onde estão sediadas. Ainda assim, foi possível observar que regiões mais ricas tendem a contar com estruturas mais robustas e a apresentar desempenhos superiores. No caso do futebol europeu, reportaram-se ainda evidências de uma associação entre indicadores econômicos regionais, capacidade de contratação de profissionais de futebol e desempenho aferido com base na classificação para campeonatos continentais. Além disso, a revisão bibliográfica mostrou que a presença de rivais locais (viabilizados por maiores escalas urbanas) pode contribuir para o desempenho das equipes.

Os dados relativos ao desempenho das equipes da série A do campeonato brasileiro foram tabulados para o período em que os certames foram disputados no formato de pontos corridos. Trabalhos futuros poderão incluir também dados da série B a partir de 2006. Com isso, 30 novas equipes e mais três unidades da federação passariam a compor a amostra. Contudo, a eventual inclusão da série B requereria que se definisse algum tipo de critério para ajustar o desempenho observado com o desempenho das equipes na série A. Trata-se, por ora, de um procedimento *ad hoc* que precisaria ser claramente definido antes que fosse possível estender a análise também à série B a partir de 2006.

Os resultados indicam que o desempenho das macrorregiões brasileiras na série A do campeonato brasileiro entre 2003 e 2019 mantém uma forte correlação com o PIB. Esse resultado é consistente com várias referências mencionadas na seção 2, segundo as quais os melhores ajustes foram obtidos ao se combinarem população e PIB *per capita* (uma vez que  $PIB = população \times PIB \text{ per capita}$ ). A exceção é a região Sul, cujo desempenho é cerca de 40% superior ao que seria esperado considerando seu PIB. A correlação entre desempenho e PIB foi reafirmada ao se analisarem as 15 unidades da federação (correspondentes a 88,60% do PIB do país) que estiveram representadas em ao menos uma ocasião na série A do campeonato brasileiro. A principal exceção, nesse caso, é o Rio de Janeiro, que tem um desempenho superior ao que seria predito com base em seus indicadores econômicos. A presença de eventuais *outliers* sugere que outros fatores podem explicar o desempenho na série A do campeonato brasileiro. Esses fatores poderiam envolver, conforme se apontou na seção 4, a tradição, o número de torcedores e a presença de rivalidades regionais que configurariam uma espécie de causalção circular e cumulativa: regiões que têm melhor desempenho no futebol tendem a se tornar mais tradicionais, a mobilizar novos torcedores e a favorecer rivalidades locais e, com isso, a melhorar seu desempenho. Embora não seja possível, neste momento, incluir esses fatores adicionais nas regressões (em virtude das dificuldades de aferição e de sua colinearidade com os indicadores econômicos), não se pode descartar sua correlação com o desempenho das equipes no campeonato brasileiro. Desse modo, o presente trabalho pode ser o ponto de partida para uma série de aprimoramentos futuros para investigar a relação entre desigualdades regionais e desempenho no futebol que envolvem, inclusive, a inclusão de fatores explicativos adicionais e o uso de unidades de análise que aumentem o número de observações (municípios ou regiões metropolitanas).

Apesar de suas limitações, as conclusões aqui apresentadas têm uma série de implicações para formuladores de políticas de esporte e para gestores de equipes de futebol. Em



primeiro lugar, é improvável que regiões periféricas consigam exibir desempenhos acumulados excepcionais. Ainda que ocasionalmente isso possa acontecer, aspectos econômicos tenderão a fazer o desempenho regredir à média (isto é, a ajustar-se, em linhas gerais, ao PIB). Ainda assim, a combinação de fatores complementares (tradição e número de torcedores, por exemplo) pode concorrer para melhorar o desempenho. Embora esses fatores requeiram tempo para serem construídos por uma equipe determinada, é possível que estratégias de comunicação mais ajustadas ao novo contexto social possam contribuir para a obtenção de melhores desempenhos. Por fim, contrariamente à reação emocional das torcidas (que tendem a desejar o fracasso de rivais regionais), a presença de antagonistas com desempenho compatível tende a concorrer para o desempenho agregado das equipes e das regiões.

## Referências

- Albuquerque, M. V., Viana, A. L. A., Lima, L. D., Ferreira, M. P., Fusaro, E. R., & Iozzi, F. L. (2017). Desigualdades regionais na saúde: mudanças observadas no Brasil de 2000 a 2016. *Ciência & Saúde Coletiva*, 22(4), 1055-1064.
- Baer, W. (2008). *The Brazilian Economy: Growth and Development* (6ª ed.). Lynne Rienner Publishers.
- Bäker, A., Mechtel, M., & Vetter, K. (2012). Beating thy neighbor: derby effects in German professional soccer. In R. H. Koning & W. Maennig (Eds.), *Sports Economics: Present and Future Impact on General Economics*. Lucius & Lucius Verlagsgesellschaft mbH.
- Barbosa, J. R. A. (2018). Reflexões sobre desigualdades regionais referentes ao uso da internet no território brasileiro. *Revista contexto geográfico*, 3(5), 75-81.
- Bernard, A. B. (2008). *Going for the gold: who will win the 2008 Olympic Games in Beijing*. Tuck School of Business.
- Bernard, A. B., & Busse, M. R. (2004). Who Wins the Olympic Games: Economic Resources and Medal Totals. *Review of Economics and Statistics*, 86(1), 413-417.
- Bian, X. (2005). *Predicting Olympic medal counts: the effects of economic development on Olympic performance*. *The Park Place Economist*, 13.



- Combes, P.-P., Duranton, G., & Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2), 253–266.
- Doran, J., & Jordan, D. (2018). The effect of geographical proximity and rivalry on performance: evidence from the English Football League. *Regional Studies*, 52(11), 1559–1569.
- Ferri, L., et al. (2017). Financial Versus Sports Performance: The Missing Link. *International Journal of Business and Management*, 12(3), 36.
- Figueiredo, E., & Porto Junior, S. (2015). Persistência das desigualdades regionais no Brasil: polarização e divergência. *Nova Economia*, 25(1), 195-208.
- Frick, B. (2007). The football players' labor market: empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, 54(3), 422–446.
- García, P. C., Castro, J. A. D., & Santos, J. M. S. (2007). *The economic geography of football success: empirical evidence from European cities*. *Rivista diritto ed economia dello sport*, III(2).
- Gimenez, D. M. (2017). *Política social e desequilíbrios regionais no Brasil em tempos de crise*. Campinas: Unicamp / Instituto de Economia (texto para discussão n. 301).
- Grimes, A. R., Kelly, W. J., & Rubin, P. H. (1974). A socioeconomic model of national olympic performance. *Social Science Quarterly*, 55(3), 777–783.
- Hoffmann, R., Ging, L. C., & Ramasamy, B. (2002). The Socio-Economic Determinants of International Soccer Performance. *Journal of Applied Economics*, 5(2), 253–272.
- Jablonsky, J. (2016). Ranking of countries in sporting events using two-stage data envelopment analysis models: a case of Summer Olympic Games 2016. *Central European Journal of Operations Research*, 26(4), 1-16.



- Johnson, D. K. N., & Ali, A. (2004). A tale of two seasons: participation and medal counts at the Summer and Winter Olympic Games. *Social Science Quarterly*, 85(4).
- Jones, C., & Jordan, D. (2019). Agglomeration, urbanization, and competitive performance: the natural experiment of English football. *Regional Studies, Regional Science*, 6(1), 421-438.
- Macmillan, P., & Smith, I. (2007). Explaining International Soccer Rankings. *Journal of Sports Economics*, 8(2), 202–213.
- Mendes, W. A., et al. (2020). Desenvolvimento humano e desigualdades regionais nos municípios brasileiros. *Latin American Research Review*, 55(4), 742–758.
- Monks, J., & Husch, J. (2009). The Impact of Seeding, Home Continent, and Hosting on FIFA World Cup Results. *Journal of Sports Economics*, 10(4), 391-408.
- Monteiro Neto, A. (2014). Desigualdades regionais no Brasil: características e tendências recentes. *Boletim regional, urbano e ambiental*, 9, jan.-jun.
- Myrdal, G. (1960). *Teoria econômica e regiões subdesenvolvidas*. Rio de Janeiro: UFMG Biblioteca Universitária (Edição original: 1957).
- Novikov, A. D., & Maximenko, A. M. (1972). The Influence of Selected Socio-Economic Factors on the Level of Sports Achievements in the Various Countries: (using as an example the 18th Olympic Games in Tokyo). *International Review of Sport Sociology*, 7(1), 27–44.
- Omondi-Ochieng, P. (2015). Gross national income, football workers, and national football team performances: a logistic regression analysis. *Team Performance Management*, 21(7/8), 405-420.
- Pfau, W. D. (2006). Predicting the medal wins at the 2006 Winter Olympics. *The Korea economic review*, 22(2).
- Rathke, A., & Woitek, U. (2008). Economics and the Summer Olympics: An Efficiency Analysis. *Journal of Sports Economics*, 9(5), 520–537.



Reiche, D. T. (2016). *Success and failure of countries at the Olympic Games*. Abingdon, Oxon; New York, NY: Routledge.

Ribeiro, A. S., & Lima, F. (2012). Portuguese football league efficiency and players' wages. *Applied Economics Letters*, 19(6), 599–602.

Robberechts, P., & Davis. (2018). Forecasting the FIFA World Cup: combining result- and goal-based team ability parameters. In U. Brefeld, J. Davis, J. Haaren, & A. Zimmermann (Eds.), *Machine Learning and Data Mining for Sports Analytics*. 5th International Workshop, MLSA 2018.

Williamson, J. G. (1968). Regional Inequality and the Process of National Development. In L. Needleman (Ed.), *Regional Analysis: Selected Readings*. Penguin Books, p. 99-158.

Wu, J., Liang, L., & Yang, F. (2009). Achievement and benchmarking of countries at the Summer Olympics using cross efficiency evaluation method. *European Journal of Operational Research*, 197(2), 722–730.

Zambom-Ferraresi, F., García-Cebrián, L. I., & Lera-López, F. (2017). Sports Results measurement and efficiency in UEFA Champions League. *Athens Journal of Sports*, 4(4).

*et al.*

