

O IMPACTO DO FATOR CASA NO DESEMPENHO DO CORINTHIANS NO BRASILEIRÃO (2003-2024): UMA ANÁLISE ESTATÍSTICA

MARIANA COSTA DE MELLO (FACULDADE DE TECNOLOGIA DE JUNDIAI
DEPUTADO ARY FOSSEN)

mariana.mello1@fatec.sp.gov.br

Resumo

Este estudo analisou o desempenho do Corinthians em jogos do Brasileirão (2003-2024), excluindo o ano de 2020 devido à pandemia, realizados em casa e fora, considerando gols marcados (GM) e saldo de gols (SG). Utilizou-se o teste de Shapiro-Wilk para verificar a normalidade dos dados e o teste não paramétrico de Mann-Whitney para comparar os grupos. Os resultados indicam que o desempenho do Corinthians difere significativamente entre jogos em casa e fora de casa. Os achados confirmam a influência do local do jogo no desempenho do Corinthians, destacando a vantagem do mando de campo.

PALAVRAS-CHAVE: desempenho esportivo, mandante, teste de Mann-Whitney, análise estatística, Corinthians.

Abstract

This study analyzed the performance of Corinthians in home and away matches, considering goals scored and goal difference. Data from Corinthians' matches were analyzed, excluding the year 2020 due to the pandemic. The Shapiro-Wilk test was used to verify data normality, and the non-parametric Mann-Whitney test was applied to compare the groups. The results indicate that Corinthians' performance differs significantly between home and away games. The findings confirm the influence of the match location on Corinthians' performance, highlighting the advantage of playing at home.

Keywords: sports performance, home team, Mann-Whitney test, statistical analysis, Corinthians

1. Introdução

A expressão “fator casa” (FC) é amplamente conhecida no contexto dos esportes coletivos, principalmente no futebol e em países da América do Sul, em que o esporte faz parte da cultura da maioria dessas nações. Essa expressão é utilizada para descrever a vantagem que um time tem ao jogar em seu próprio estádio ou arena, cercado por sua torcida [1].

No dia 27 de março de 2025, foi realizada a final do Paulistão 2025, confronto decidido entre Corinthians e Palmeiras (conhecido como “*derby*”). O Corinthians conquistou o título ao empatar contra o Palmeiras em sua casa, tendo vencido o primeiro jogo pelo placar de 1-0 na casa do rival, e a sua torcida foi um fator crucial para essa conquista, conduzindo o clima do jogo [2].

Levando em conta essa característica do time, através dos dados disponíveis no site basedosdados.org, essa pesquisa propõe-se a investigar a relação entre o mando de campo e o desempenho do Corinthians no Brasileirão 2003 a 2024 (desconsiderando o ano de 2020, por conta da pandemia), com base nos gols marcados (GM) e no saldo de gols (SG).[3]. Serão realizadas análises referente às frequências e as estatísticas descritivas, sendo verificado os valores das médias, desvio padrão, mínimo, máximo, mediana, 1º quartil e 3º quartil. Além disso, serão utilizados testes estatísticos, como o de Shapiro-Wilk e o de Mann-Whitney, para verificar a normalidade dos dados e comparar os grupos, respectivamente. Para uma análise mais visual será utilizado o recurso de gráficos como boxplot, violino e estimativa de densidade Kernel.

2. Embasamento Teórico

2.1. Revisão da literatura

Diversos estudos apontam uma vantagem em se jogar em casa, Courneya e Carron verificaram a relação entre o FC e o percentual de vitórias como mandante para diferentes esportes, os resultados apontaram taxas de 53,5% para beisebol, 57,3% para futebol americano, 61,1% para hóquei no gelo, 64,4% para o basquete e 64,5% para o futebol. Esses dados sugerem que o futebol, dentre as modalidades estudadas, é a que o fator mais se evidencia [4]. Silva (2004) também investigou essa questão, tendo como foco os campeonatos de futebol brasileiros da série A e B. Em sua pesquisa, foram obtidos percentuais ainda mais altos: 68,71% e 68,46%, respectivamente [5]. De acordo com o psicólogo esportivo Maurício Marques:

“A diferença de jogar em casa ou como visitante é enorme. Há muito apoio e pouca pressão contra – excluindo casos em que a equipe mandante passa por momentos ruins e a torcida “joga contra”. Em casa, por exemplo, a equipe conhece o campo, a grama, a velocidade

da bola e tem pontos de referência no gramado. A torcida, em geral, é uma segurança psicológica para os atletas.” [6]

Com base nisso, podemos dizer que o fato do time jogar em casa influencia significativamente as decisões dos jogadores, seja pelo maior conhecimento do local ou pelo clima hostil da arquibancada, a favor ou contra. Para Machado (1997), o emocional do atleta pode, sim, ser afetado pela pressão da torcida, podendo ter um aspecto positivo ou negativo. Segundo ele, jogadores mais experientes tendem a lidar melhor com essa pressão do que atletas em início de carreira. No entanto, esse impacto pode alcançar até mesmo as decisões da comissão técnica [7].

2.2. Teste de Normalidade (Shapiro-Wilk)

O teste de normalidade Shapiro-Wilk é um teste estatístico utilizado para avaliar a normalidade de uma amostra de dados. Para esse teste é considerado como hipótese nula (H_0) quando $p > 0.05$, indicando que a amostra possui distribuição normal, enquanto a hipótese alternativa (H_1) é aceita quando $p \leq 0.05$, indicando que os dados não seguem uma distribuição normal [8].

2.3. Estimativa de densidade de kernel (KDE)

A estimativa de densidade de kernel (KDE) é um método para visualizar a distribuição em um conjunto de dados, análogo a um histograma, o KDE usa uma curva suave para representar a densidade de probabilidade [9].

2.4. Teste de Mann-Whitney

Inicialmente desenvolvido por F. Wilcoxon em 1945, o teste de Mann Whitney foi idealizado para comparar as tendências centrais de duas amostras independentes de tamanhos iguais. Em 1947, H.B. Mann e D.R. Whitney generalizaram a técnica para amostras de tamanhos diferentes. Análogo do teste t de Student paramétrico, ele compara as médias entre dois grupos independentes, considerando que os dados não estão em uma distribuição normal [10-11].

3. Metodologia

3.1. Obtenção dos Dados

Os dados utilizados foram extraídos da plataforma basedosdados.org, no conjunto de dados referente ao Brasileirão Série A, com uma cobertura temporal entre 2003 e 2024 [3].

3.2. Importação das bibliotecas

Utilizando o aplicativo VSCode, foi criado um arquivo .ipynb, disponível no repositório [12] e as bibliotecas pandas, seaborn, matplotlib e scipy.stats foram importadas para manipulação do arquivo CSV, criação do gráfico, plotagem do gráfico e realização das medidas estatísticas, respectivamente.

3.3. Carregamento do arquivo

A função “read_csv” do pandas foi empregada para carregar o arquivo, identificando “;” como o delimitador.

3.4. Tratamento dos Dados

Referente ao tratamento dos dados, utilizando-se de funções da biblioteca pandas, foram selecionados apenas os jogos em que o Corinthians era mandante ou visitante, foram descartados os dados referentes ao ano de 2020, devido à pandemia. Para ajudar na análise foram criadas as colunas gols_corinthians e gols_adversário a partir da coluna gols_mandante e gols_visitante, a coluna saldo_gols a partir das colunas anteriormente criadas e a coluna em_casa com valores lógicos, assumindo como True quando o Corinthians estava como mandante e False caso contrário. A partir da coluna em_casa foi criada a coluna local com os valores “Fora” ou “Casa”.

3.5. Tabelas Frequência (Absoluta e Relativa)

Com objetivo de analisar a distribuição de frequências dos gols marcados e do SG do Corinthians, novamente utilizando-se das funções da biblioteca pandas, foi gerada a tabela de frequências, mostrando as frequências absolutas (número de jogos com determinado número de GM ou SG) e as frequências relativas (percentual de ocorrência de cada valor).

3.6. Estatística Descritiva

Para realização das medidas descritivas foi utilizada a função describe() do pandas para obtenção, tanto para GM quanto para SG, do número de jogos analisados, média, mediana, desvio padrão, primeiro quartil, terceiro quartil, mínimo e máximo. Essas medidas servem para entender melhor sobre a distribuição dos dados, tendência central e a dispersão.

3.7. Teste de Normalidade

Para verificar a normalidade do conjunto de dados, verificar se ele segue a distribuição normal (Gaussiana), foi utilizado Teste de Shapiro-Wilk através da função `shapiro()` da biblioteca `scipy.stats` e a partir do resultado foi decidido qual método estatístico utilizar.

Foi constatado que os dados, tanto para gols quanto para SG não possuíam uma distribuição normal, por isso foi escolhido o Teste de Mann-Whitney, teste estatístico não paramétrico.

3.8. Testes de Mann-Whitney

Para o teste de Mann-Whitney foi utilizada a função `mannwhitneyu()` da biblioteca `scipy.stats`, usando como parâmetro `casa['gols_corinthians']` e `fora['gols_corinthians']` para análise dos GM e `casa['saldo_gols']` e `fora['saldo_gols']` para SG.

3.9. Plotagem dos Gráficos

Para uma análise mais visual dos resultados da mediana, quartis, mínimo e máximo foi utilizado o gráfico de boxplot (função `boxplot()` da biblioteca `seaborn`) e para visualizar a densidade da distribuição dos dados foi utilizado o gráfico de violino (função `violinplot()` da biblioteca `seaborn`) e o gráfico de estimativa densidade Kernel (função `kdeplot()` da biblioteca `seaborn`).

4. Resultados e Discussão

Na Tabelas 1 e 2 estão dispostas as frequências absolutas e relativas para GM e SG do Corinthians em casa e fora, respectivamente. Na tabela 3 estão dispostas as estatísticas descritivas dos dados como a quantidade de jogos (qntd), a média (med), o desvio padrão (DP), o primeiro quartil (25%), a mediana (50%), o terceiro quartil (75%), o valor mínimo (min) e o valor máximo (max).

Tabela 1. Tabela de Frequência: Gols do Corinthians.

Gols	Frequência Absoluta		Frequência Relativa (%)	
	Casa	Fora	Casa	Fora

0	75	135	19,69	35,53
1	143	139	37,53	36,58
2	91	72	23,88	18,95
3	51	23	13,39	6,05
4	11	9	2,89	2,37
5	7	2	1,84	0,53
6	2	0	0,52	0,00
7	1	0	0,26	0,00

Fonte: Próprio autor.

Tabela 2. Tabela de Frequência: Gols do Corinthians

Saldo de Gols	Frequência Absoluta		Frequência Relativa	
	Casa	Fora	Casa (%)	Fora (%)
-5	1	1	0,26	0,26
-4	3	5	0,79	1,32
-3	8	13	2,1	3,42
-2	16	34	4,2	8,95
-1	38	100	9,97	26,32
0	109	118	28,61	31,05
1	120	70	31,5	18,42
2	50	28	13,12	7,37
3	25	9	6,56	2,37
4	6	2	1,57	0,53
5	4	0	1,05	0
6	1	0	0,26	0

Fonte: Próprio autor.

Tabela 3: Estatísticas descritivas dos gols e saldo de gols do Corinthians nos jogos Fora e Em Casa.

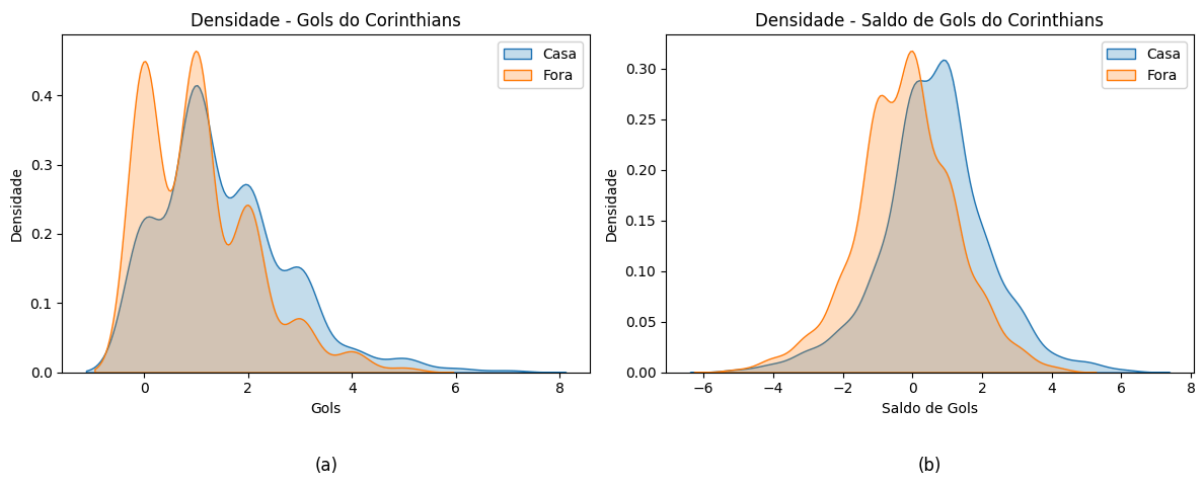
Estatísticas Descritivas				
	Gols		Saldo Gols	
	Casa	Fora	Casa	Fora
qntd	381	380	381	380
med	1,51	1,05	0,6	-0,19
DP	1,22	1,04	1,52	1,4
min	0	0	-5	-5
25%	1	0	0	-1
50%	1	1	1	0
75%	2	2	1	1
max	7	5	6	4

Fonte: Próprio autor.

O Gráfico 1 apresenta a estimativa de densidade de kernel (KDE), que permite visualizar a distribuição dos dados de forma suavizada.

Gráfico 1. (a) Gráfico da estimativa da densidade de gols do Corinthians no Brasileirão (2003-2024).

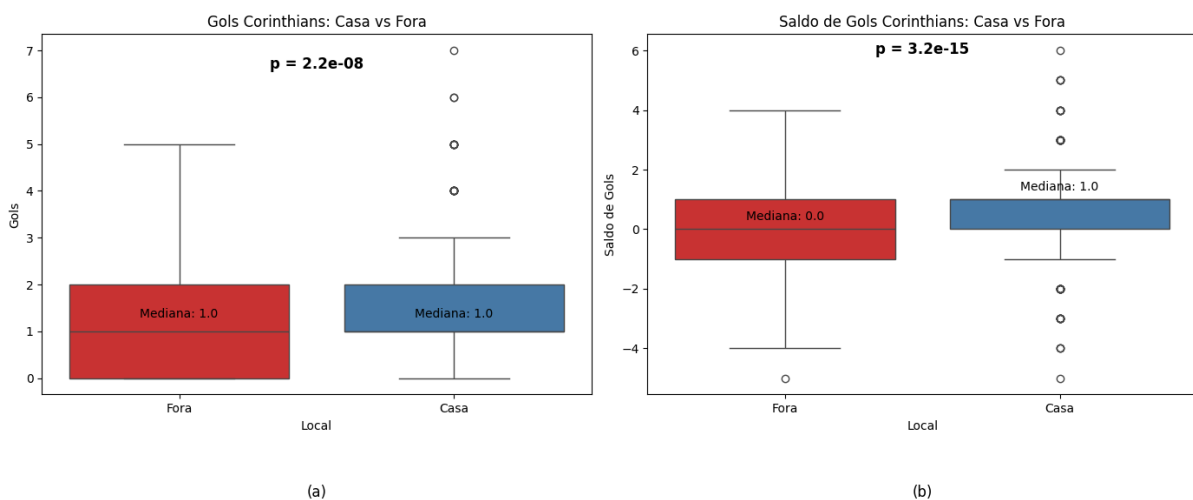
(b) Gráfico da estimativa da densidade de saldo de gols do Corinthians no Brasileirão (2003-2024).



Fonte: Próprio autor.

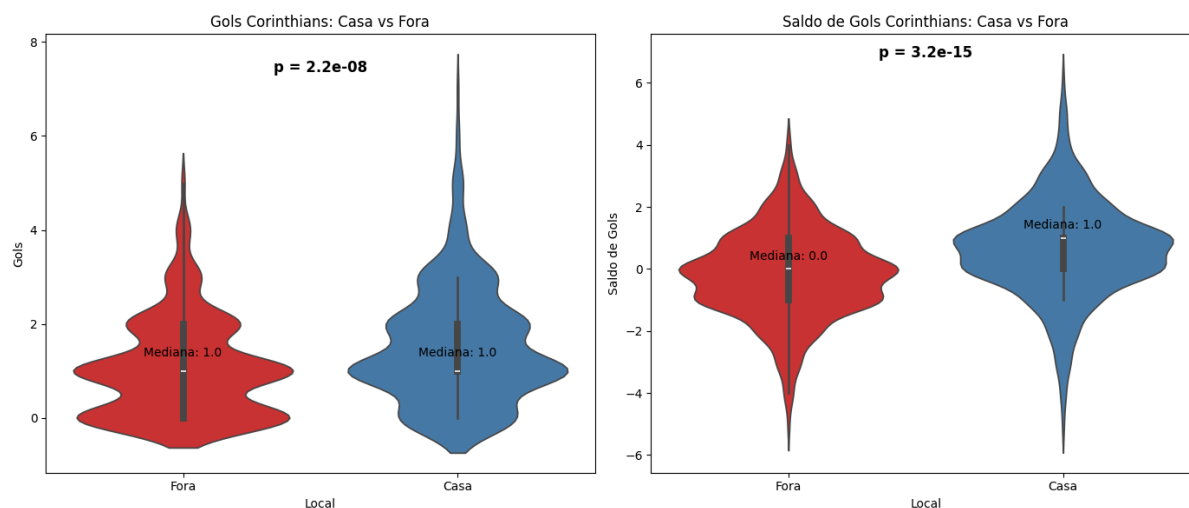
Os Gráficos 2 e 3 correspondem, respectivamente, às representações em boxplot e em violino, obtidas com base na análise dos dados.

Gráfico 2. (a) Comparação entre o desempenho ofensivo do Corinthians em casa e fora. **(b)** Comparação entre o saldo de gols do Corinthians em casa e fora



Fonte: Próprio autor.

Gráfico 3. (a) Distribuição dos gols do Corinthians no Brasileirão (2003-2024). **(b)** Distribuição dos saldo de gols do Corinthians no Brasileirão (2003-2024).



Fonte: Próprio autor.

Com base nos dados apresentados, nas tabelas e nos gráficos, é possível constatar que, para análise dos GM do Corinthians, apesar da mediana ser a mesma tanto para Casa quanto para Fora, para jogos como mandante há uma distribuição mais simétrica e concentrada em 1 e 2 gols, enquanto para jogos como visitante há uma maior dispersão, com mais jogos com 0 gols. Referente aos valores discrepantes, percebemos que os jogos em casa obtiveram o maior valor, 7 gols, enquanto fora o maior valor foi de 5 gols. Através do teste de Mann-Whitney foi obtido um valor de $p = 2.1573e-08$, muito menor que 0.05, diferença estatisticamente significativa, mostrando que, apesar da mediana igual, jogar em casa tende a influenciar a forma como o time marca gols.

Já para análise do SG, é possível visualizar uma mediana e média maior para jogos em casa, tendo saldo positivo na maioria dos jogos como mandante. Como visitante, observou-se uma alta variabilidade, com muitos jogos negativos. Referente aos valores discrepantes tivemos um valor máximo maior para jogos em casa do que para jogos fora (saldo de 6 gols em casa e 4 gols fora). O teste de Mann-Whitney mostrou um valor de $p = 3.2165e-15$, o que, assim como para a relação entre gols e o local, mostra que jogar em casa tende a influenciar no saldo de gols do Corinthians.

5. Considerações Finais

A partir desta análise, baseando-se nos valores de p menores que 0,05 (tanto para GM quanto para SG), foi possível concluir que o desempenho do Corinthians apresenta diferenças

estatisticamente significativas entre partidas realizadas em casa e fora de casa, indicando que o mando de campo pode influenciar na sua quantidade de GM e no seu SG.

Referências

- [1] RIGON, Jéssica. *Fator casa no futebol: vantagem decisiva para apostas assertivas*. Disponível em: <https://blog.mrjack.bet.br/esportes/futebol/fator-casa-no-futebol/>. Acesso em: 4 abr. 2025.
- [2] CASSUCI, Bruno; RUDÁ, Yago. Corinthians bate recorde de público na Neo Química Arena em final contra o Palmeiras. *GE.globo.com*, São Paulo, 27 mar. 2025. Disponível em: <https://ge.globo.com/futebol/times/corinthians/noticia/2025/03/27/corinthians-bate-recorde-d-e-publico-na-neo-quimica-arena-em-final-contra-o-palmeiras.ghtml>. Acesso em: 13 maio 2025.
- [3] BASE DOS DADOS. *Campeonatos de Futebol*. Disponível em: <https://basedosdados.org/dataset/c861330e-bca2-474d-9073-bc70744a1b23?table=18835b0d-233e-4857-b454-1fa34a81b4fa>. Acesso em: 3 abr. 2025.
- [4] COURNEYA, K. S.; CARRON, A. V. The home advantage in sports competitions: a literature review. *Journal of Sport and Exercise Psychology*, v. 14, n. 1, p. 13–27, 1992.
- [5] SILVA, C. D. Vantagem de jogar em casa: uma avaliação no futebol brasileiro na temporada de 2003. *EFDeportes: Revista Digital*, Buenos Aires, ano 10, n. 71, 2004. Disponível em: <https://www.efdeportes.com/efd71/casa.htm>. Acesso em: 3 abr. 2025.
- [6] FOOTURE. *A influência da torcida no futebol brasileiro*. Disponível em: <https://footure.com.br/a-influencia-da-torcida-no-futebol-brasileiro/>. Acesso em: 6 mar. 2025.
- [7] MACHADO, A. A. (org.). *Psicologia do esporte: temas emergentes I*. Jundiaí: Ápice, 1997.
- [8] UNIVERSIDADE ESTADUAL DE LONDRINA (UEL). Teste de Shapiro-Wilk. Disponível em: <https://www.uel.br/projetos/experimental/pages/arquivos/Shapiro.html>. Acesso em: 13 maio 2025.

[9] SEABORN DEVELOPMENT TEAM. seaborn.kdeplot — seaborn 0.12.2 documentation. Disponível em: <https://seaborn.pydata.org/generated/seaborn.kdeplot.html>. Acesso em: 14 maio 2025.

[10] SUNDJAJA, Joshua Henrina; SHRESTHA, Rijen; KRISHAN, Kewal. McNemar and Mann-Whitney U Tests. StatPearls Publishing, 2025. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK560699/>. Acesso em: 14 maio 2025.

[11] CARMOS, Vera. Teste de Mann-Whitney. Universidade Federal de Santa Catarina (UFSC). Disponível em: https://www.inf.ufsc.br/~vera.carmo/Testes_de_Hipoteses/Testes_nao_parametricos_Mann-Whitney.pdf. Acesso em: 15 maio 2025.

[12] MELLO, Mariana. *Corinthians Performance Analysis*. GitHub, 2023. Disponível em: <https://github.com/maricmello/corinthians-performance-analysis>. Acesso em: 18 maio 2025.