

MATRIZ DE CORRELAÇÃO ENTRE AS QUATRO PRINCIPAIS LIGAS DE ESPORTES DOS ESTADOS UNIDOS

GUILHERME DE MORAES MASUKO

2022

RESUMO

A presente análise buscou estudar o comportamento entre a performance dos times na quatro ligas mais famosas dos Estados Unidos e Canadá. Conhecidas como "Big Four", as quatro ligas de esportes coletivos são a Major League Baseball (MLB), a National Basketball Association (NBA), a National Football League (NFL) e a National Hockey League (NHL). Foi utilizado o coeficiente de correlação de Pearson, como medidor da relação entre a performance desses times e as regiões metropolitanas como indexador. O resultado final do trabalho conta com uma matriz de correlação medindo a relação entre a performance média dos times das regiões metropolitanas em seus pares de esportes. Um resultado de destaque foi a correlação entre as ligas NBA e NHL, onde constatou uma correlação de 0,76.

Palavras-chave: Correlação. Esporte.

SUMÁRIO

1	INTRODUÇÃO	3
2	METODOLOGIA	4
3	DADOS	6
4	CONCLUSÃO	9
	REFERÊNCIAS	10

1 INTRODUÇÃO

Os Estados Unidos é um país conhecido pelo seu ótimo desempenho no esporte. As quatro principais ligas esportivas profissionais nos Estados Unidos e Canadá, comumente referenciada como "Big Four", são as competições profissionais mais competitivas de esportes coletivos nesses países. As quatro ligas inclusas nessa definição são a Major League Baseball (MLB), a National Basketball Association (NBA), a National Football League (NFL) e a National Hockey League (NHL). É comum verificar que vários times entre as ligas Big Four estão situados em mesma região metropolitana.

O presente estudo objetivou-se em questionar as relações que se têm entre estas regiões metropolitanas e a performance média obtida por seus times. Utilizando o coeficiente de correlação de Pearson, foi medida a correlação entre a população destas regiões com a performance média de seus times. O coeficiente de Pearson, como mostra em Wikipédia (2021b), pode ter valores entre -1 e 1. Um resultado muito próximo de um nos diria que uma região metropolitana com grande população está ligada à boas performances médias, enquanto regiões com baixa população estão ligadas à performances médias menos regulares. Já um resultado próximo de menos um nos entregaria uma inferência inversa à acima.

Para o restante do trabalho, foi estudado o relacionamento que se têm entre a performance média das regiões metropolitanas entre pares de esporte. Isto é, queremos saber qual o comportamento dos times de mesma região metropolitana, mas em esportes diferentes. Por exemplo, queremos descobrir se os times da NHL da região metropolitana de New York City, New Jersey Devils, New York Islanders e New York Rangers, tem performance média parecida com os times New York Knicks e Brooklyn Nets da liga de basquete.

2 METODOLOGIA

Para mensurar a relação entre a performance média das regiões metropolitanas nas ligas norte-americanas NBA, MLB e NFL, será proposto como medidor de performance, visualizado na equação 2.1, a razão entre vitórias e jogos de cada time, ou seja

$$P_i = \frac{W_i}{G_i} \quad (2.1)$$

onde P_i é a performance do time i , W_i é a quantidade de vitórias na temporada regular do time i e G_i é a quantidade de jogos da temporada regular do time i .

Para a liga NHL, como há um esquema de pontuação diferente, onde o time ganha dois pontos caso vença, um ponto caso seja derrotado em um overtime, e zero caso seja derrotado no tempo normal, o medidor de performance é proposto a partir da razão entre pontuação do time e pontuação possível, este último travado em 164 (pontuação caso um time ganhar as 82 partidas que contém uma temporada regular). A equação 2.2 ilustra isso.

$$P_i = \frac{Pt_i}{maxPt} = \frac{Pt_i}{164} \quad (2.2)$$

onde P_i é a performance do time i , Pt_i é a pontuação na temporada regular do time i e $maxPt$ é a pontuação máxima, de 164 pontos, da temporada regular que um time pode atingir.

Algumas ligas possuem divisões de localidades para que seja possível manter a quantidade de jogos da temporada regular. O presente estudo não discriminará em qual divisão cada time pertence, fazendo apenas uma agregação simples da performance de cada time, considerando que essas divisões mantêm o mesmo nível de dificuldade.

Possivelmente vários times podem estar situados em mesma região metropolitana e para chegar na performance da região será utilizado a média simples da performance de cada time.

E, finalmente, para mensurar a relação entre a performance média das regiões metropolitanas usaremos a matriz de correlação para as quatro principais ligas norte-americanas. Como pode ser visto em Wikipédia (2021a), correlação é uma medida estatística de relação (causal ou não-causal) entre duas variáveis abordando o comportamento da relação entre elas. O coeficiente de correlação utilizado pela presente análise será o de Pearson, o coeficiente de correlação mais comumente usado.

O coeficiente de correlação de Pearson, segundo Wikipédia (2021b), define o quanto e em qual direção duas variáveis estão relacionadas, sendo obtida a partir da seguinte fórmula demonstrada na equação 2.3 abaixo.. Usado normalmente como ρ , o coeficiente pode ter valores em um range de -1 à 1 , isto é, $-1 \leq \rho \leq 1$, onde $\rho = -1$ significa uma relação perfeitamente negativa entre as variáveis, $\rho = 1$ uma relação perfeitamente positiva, e $\rho = 0$

uma relação de não dependência linear (isso não diz nada sobre não haver dependência de modo geral entre as variáveis).

$$\rho_{X,Y} = \frac{\text{COV}(X,Y)}{\sqrt{\text{VAR}(X)}\sqrt{\text{VAR}(Y)}} \quad (2.3)$$

onde $\rho_{X,Y}$ é a correlação entre X e Y .

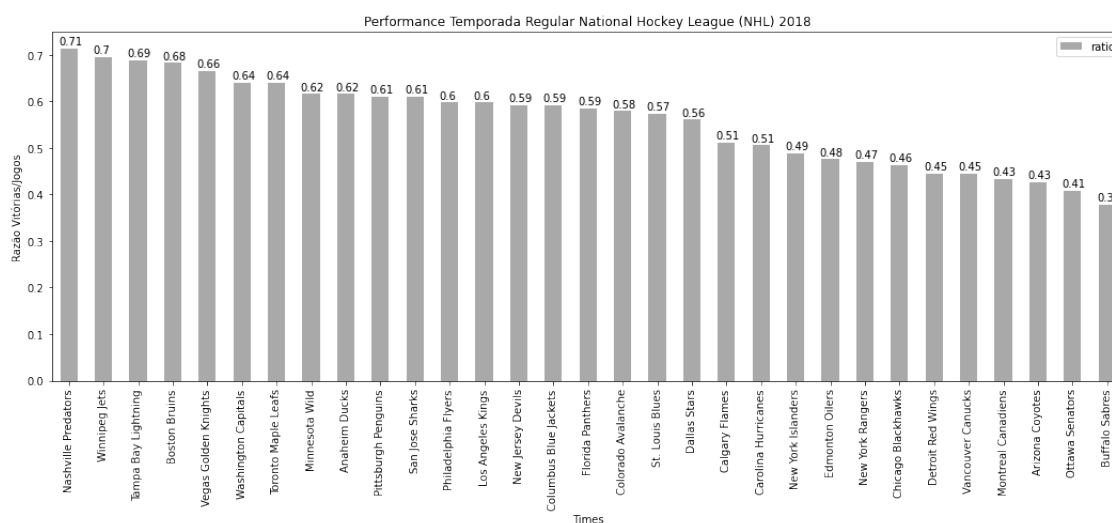
A matriz de correlação é uma ferramenta estatística capaz de medir a correlação de uma coleção de variáveis em seus pares, uma ótima maneira de obter de forma reduzida, o comportamento de várias séries em uma só tabela. Uma matriz de correlação de n variáveis X_1, \dots, X_n , é uma matriz $n \times n$, no qual o i, j -ésimo elemento da matriz é a correlação entre X_i e X_j , isto é, ρ_{X_i, X_j} .

3 DADOS

Os dados das regiões metropolitanas dos Estados Unidos e Canadá utilizados nesse estudo foram obtidos através de Wikipédia (2016), onde uma tabela pode ser encontrada relacionando cada região metropolitana com os times da big four. Nesta tabela também há informações sobre a quantidade populacional de cada região, com dados atualizados para o ano de 2016.

A primeira liga abordada foi a National Hockey League. Para esta e todas as demais ligas presentes neste trabalho, foi escolhido os resultados da temporada regular de 2018. No gráfico 1, podemos verificar a performance de cada time na temporada regular utilizando o medidor proposto em particular à esta liga, como a razão entre a pontuação do time sobre a máxima pontuação possível. A liga é dividida entre quatro divisões, são elas a do Atlântico, Metropolitana, Central e do Pacífico.

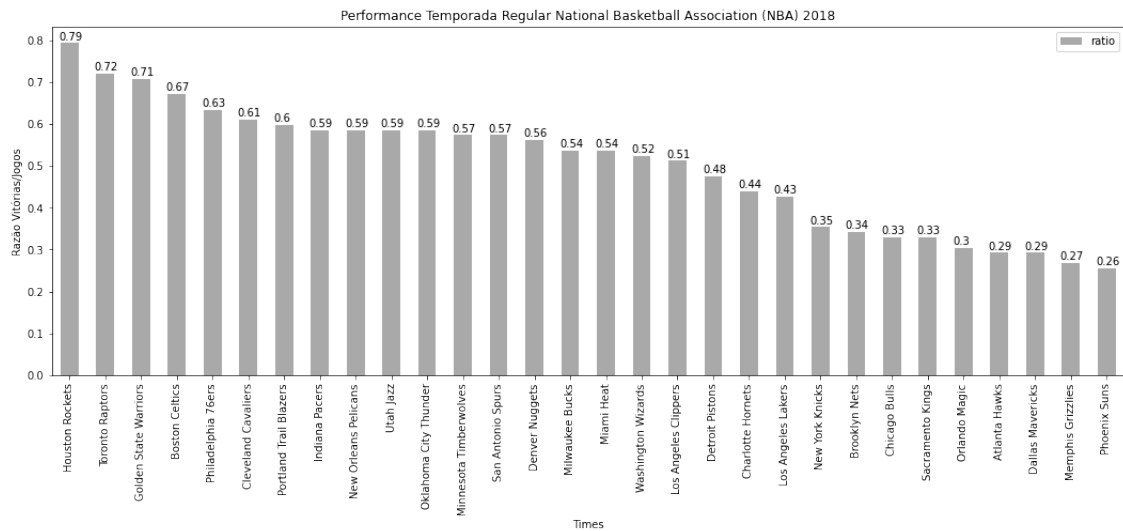
Figura 1 – Performance Temporada Regular National Hockey League (NHL) 2018



Fonte: Elaboração própria a partir de Reference (2018d)

O destaque para a liga de hóquei ficou para o Nashville Predators. O time de Nashville se mostrou muito forte com uma razão de 71% de performance na temporada regular, mas mesmo assim acabou caindo na segunda rodada dos playoffs para o Winnipeg Jets, este com a segunda melhor campanha na temporada regular, com 70% de performance. Por fim, o time que acabou sendo campeão foi o Washington Capitals.

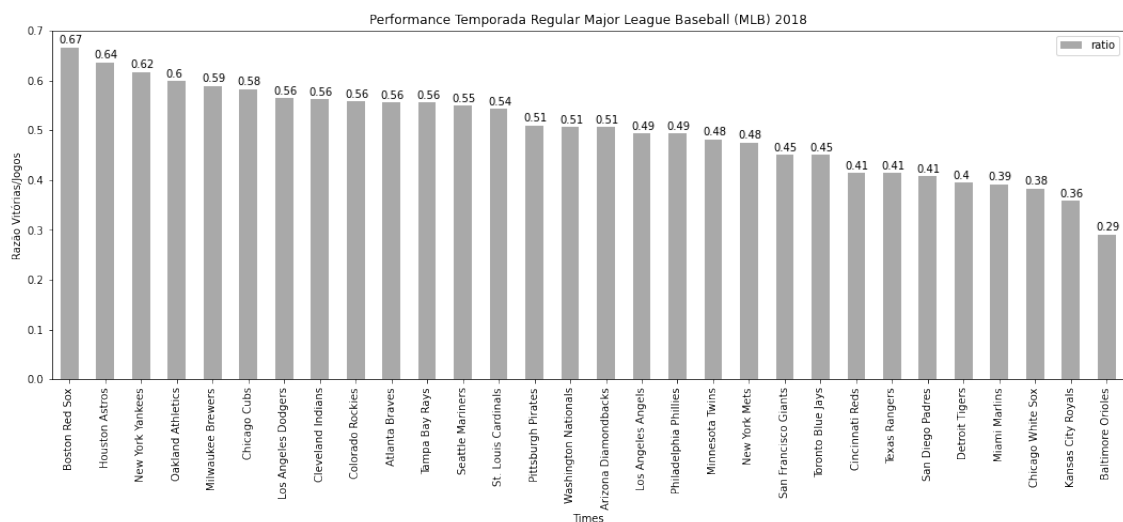
Para as ligas restantes o estimador de performance é obtido a partir da equação 2.1. O segundo esporte analisado é o basquete. Utilizando dados da temporada regular da National Basketball Association, pode se ver a lista de times e suas respectivas performance através do gráfico 2. A NBA é dividida em duas divisões, conhecidas como Conferências Leste e Oeste.

Figura 2 – Performance Temporada Regular National Basketball Association (NBA) 2018

Fonte: Elaboração própria a partir de Reference (2018b)

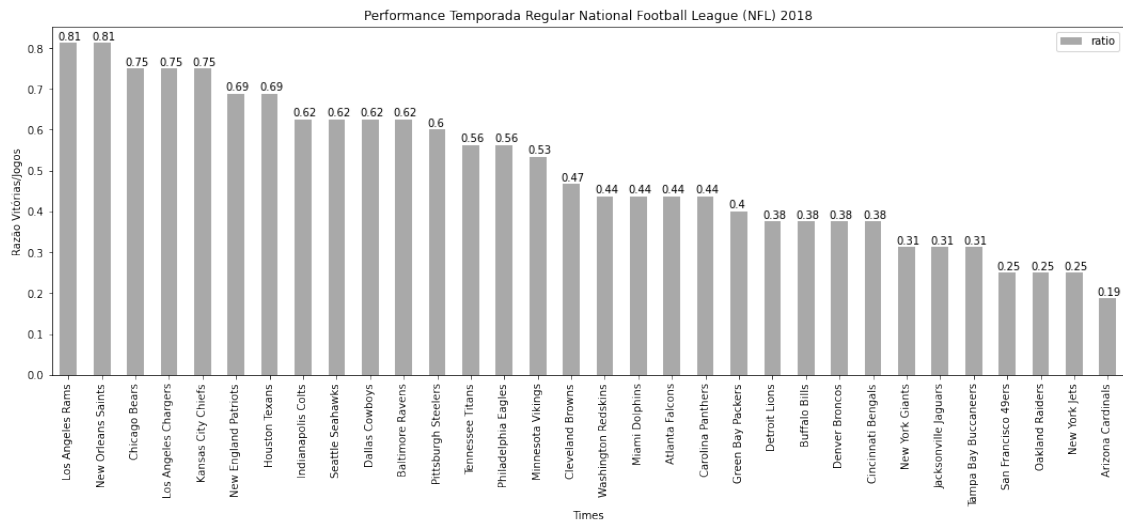
Com destaque para o time Houston Hockets, o time obteve 65 vitórias. Mesmo com a tamanha regularidade na temporada, com 79% em performance, o time caiu diante do Golden State Warriors, nas finais da Conferência Oeste. O time da região metropolitana de San Francisco Bay Area se sagrou campeão vencendo também a final da NBA sobre o Cleveland Cavaliers.

A terceira liga é a Major League Baseball. O gráfico 3 mostra a superioridade do time Boston Red Sox durante a temporada regular. O time também foi o campeão dos playoffs neste ano.

Figura 3 – Performance Temporada Regular Major League Baseball (MLB) 2018

Fonte: Elaboração própria a partir de Reference (2018a)

A quarta e última liga abordada nesse trabalho é a National Football League.

Figura 4 – Performance Temporada Regular National Football League (NFL) 2018

Fonte: Elaboração própria a partir de Reference (2018c)

Mesmo que os times Los Angeles Rams e New Orleans Saints tivessem atingido a incrível marca de 81% de performance na temporada regular como aponta o gráfico 4, o time que acabou como campeão de um dos eventos esportivos mais famosos conhecido como Super Bowl, foi o New Englands Patriots.

4 CONCLUSÃO

Os primeiros resultados obtidos neste trabalho foram em relação a variável indexadora. Foi estimada a correlação entre a população de cada região metropolitana e a performance de seus respectivos times para cada liga.

Tabela 1 – Correlação entre População e Performance média da Região Metropolitana para cada Liga

	População (2016)
NFL	0.004922
NBA	-0.176572
NHL	0.006519
MLB	0.150277

Fonte: Elaboração própria.

Os resultados mostram que para as ligas NFL e NHL, os níveis de correlação foram muito próximos de zero, podendo dizer que não há uma relação linear entre a performance média das regiões metropolitanas nestes esportes e a população destas regiões. Para o basquete, a correlação entre a performance média das regiões metropolitanas na NBA e a população é menor que zero, isso demonstra que há uma relação negativa entre essas duas variáveis. Enquanto que para a MLB, a relação entre estas duas variáveis foi positiva, mostrando que uma população maior está positivamente relacionada com a performance média de uma determinada região metropolitana. Embora existam essas relações para NBA e MLB, os coeficientes de correlação não foram grandes o suficiente para encontrar casos distintos ao que foi inferido.

A matriz de correlação mostra a relação que se teve entre a performances média das regiões metropolitanas aos pares de ligas esportivas.

Tabela 2 – Matriz de Correlação da Performance média da Região Metropolitana entre as Ligas

	NFL	NBA	NHL	MLB
NFL		0.237000	0.304000	-0.050000
NBA	0.237000		0.760000	0.325000
NHL	0.304000	0.760000		0.432000
MLB	-0.050000	0.325000	0.432000	

Fonte: Elaboração própria.

O basquete e o hóquei foram dois esportes em destaque, onde a performance média das regiões metropolitanas dos times se mostraram positivamente correlacionados com um coeficiente relativamente alto. Isso mostra que times de mesma região metropolitana devem ter tido resultados mais parecidos. Esse não é o caso da correlação entre a NFL e MLB, onde o coeficiente de Pearson foi de -0,05.

REFERÊNCIAS

REFERENCE. *Major League Baseball (MLB)*. 2018. Disponível em: <<https://www.baseball-reference.com/leagues/majors/2018.shtml>>. Acesso em: 25 fev. 2022. 7

REFERENCE. *National Basketball Association (NBA)*. 2018. Disponível em: <https://www.basketball-reference.com/leagues/NBA_2018.html>. Acesso em: 25 fev. 2022. 7

REFERENCE. *National Football League (NFL)*. 2018. Disponível em: <<https://www.pro-football-reference.com/years/2018/>>. Acesso em: 25 fev. 2022. 8

REFERENCE. *National Hockey League (NHL)*. 2018. Disponível em: <https://www.hockey-reference.com/leagues/NHL_2018.html>. Acesso em: 25 fev. 2022. 6

WIKIPÉDIA. *Regiões Metropolitanas (EUA)*. 2016. Disponível em: <https://en.wikipedia.org/wiki/List_of_American_and_Canadian_cities_by_number_of_major_professional_sports_franchises>. Acesso em: 25 fev. 2022. 6

WIKIPÉDIA. *Correlação*. 2021. Disponível em: <<https://pt.wikipedia.org/wiki/Correla%C3%A7%C3%A3o>>. Acesso em: 25 fev. 2022. 4

WIKIPÉDIA. *Correlação de Pearson*. 2021. Disponível em: <https://pt.wikipedia.org/wiki/Coefficiente_de_correla%C3%A7%C3%A3o_de_Pearson>. Acesso em: 25 fev. 2022. 3, 4