LI Simpósio Brasileiro de Pesquisa Operacional
Limeira–SP, 02 a 06 de setembro de 2019

Otimização multiobjetivo aplicada à montagem de equipes de beisebol

Gabriela Rodrigues Niquini[1]
[1]Departamento de Engenharia Química – Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Pampulha - Belo Horizonte - MG - CEP 31270-901
gabrielaniquini@gmail.com

Andréa Oliveira Souza da Costa[1]
andreaosc@yahoo.com.br

Esly Ferreira da Costa Junior[1]
esly@deq.ufmg.br

RESUMO

Sob o pretexto de equilibrar os times participantes, algumas ligas esportivas contam com uma restrição salarial para a montagem dos elencos, fazendo com que os times devam decidir como distribuir seu orçamento. Neste trabalho, propõe-se a montagem de equipes de beisebol, com diferentes ênfases em ataque e defesa, por meio de otimização multiobjetivo. Utilizaram-se dados reais da *Major League Baseball* – principal liga estadunidense – relativos ao ano de 2017. A estatística defensiva utilizada foi a média de corridas merecidas cedidas por jogo (ERA – *Earned runs average*), e a estatística ofensiva foi a razão entre as rebatidas obtidas pelo número de idas ao bastão (AVG – *Average*). O time com maior ênfase no ataque apresentou um AVG de 0,3188 e um ERA de 3,172. O time mais defensivo apresentou um AVG de 0,3027 e um ERA de 2,496. Foi possível observar o *trade-off* entre essas duas variáveis, por meio da Frente de Pareto.

PALAVRAS CHAVE. Otimização multiobjetivo, Ligas esportivas, Montagem de times.

Eixos temáticos
- PM – Programação matemática
- OA – Outras aplicações em PO

ABSTRACT

In order to balance teams' strength, some sports leagues impose a salary restriction to roster construction, forcing each team to decide the most suitable manner to distribute their budget. In this paper, it is proposed the assembly of baseball teams, with different emphases in offense and defense, through multi-objective optimization. It was used real data from Major League Baseball - main American league - related to the 2017 year. The defensive statistic used was ERA (Earned Runs Average) - average of earned runs conceded per game, and the offensive statistic used was AVG (Average) - number of hits divided by at bats. The team with more emphasis on offense had an AVG of 0.3188 and an ERA of 3.172. The team with more emphasis on defense had an AVG of 0.3027 and an ERA of 2.496. It was possible to observe the trade-off between these two variables through Pareto frontier.

KEYWORDS. Multi-objective optimization, Sports leagues, Teams' assembly.

Paper topics
- PM – Mathematical programming
- OA – Other applications in OR

https://proceedings.science/p/106934?lang=pt-br

1. Introdução

1.1. Problemas de otimização no meio esportivo

No meio esportivo, é comum a tomada de decisões por meio de sorteios ou conhecimentos subjetivos. Escalas de arbitragem ou a tabela de jogos de um campeonato, por exemplo, são usualmente definidas por sorteio, não havendo preocupação em otimizar as distâncias percorridas pelos times ou árbitros. Entretanto, observa-se que, caso algoritmos de otimização sejam implementados, é possível a economia significativa de recursos. Em Añon et al. (2017), os autores propõem escalas de arbitragem para o Campeonato Brasileiro cumprindo variadas restrições, de modo a reduzir a distância total percorrida pelos árbitros. Em um dos cenários, os autores obtiveram uma redução de 38% na distância total percorrida pelos árbitros em relação aos dados reais empregados [Añon et al. 2017]. Hoshino e Kawarabayashi (2011) propuseram um modelo de otimização para os enfrentamentos na liga japonesa de beisebol, logrando uma redução de 25% na distância percorrida pelos times em relação aos dados reais [Hoshino e Kawarabayashi 2011].

Para a montagem de elencos, a análise torna-se mais abstrata, considerando que os times devem possuir habilidades em aspectos distintos, e a influência do elenco no resultado não é trivial. No entanto, frente a restrições salariais, a análise de estatísticas relativas aos jogadores torna-se um diferencial na montagem de elencos. Mulholland e Jensen (2019) buscaram um modelo ótimo para a distribuição do teto salarial entre as posições na montagem de um time na liga estadunidense de futebol americano, a *National Football League* (NFL). No caso desta liga, o teto salarial existe visando prover um maior equilíbrio entre as equipes e não pode ser violado. O modelo criado pelos autores visou maximizar o número de vitórias esperadas em uma temporada. Concluiu-se que, em um mercado eficiente, as principais posições são a de quarterback, guard, linha defensiva e linebacker [Mulholland e Jensen 2019].

Em alguns esportes, o problema da montagem de um elenco torna-se mais complicado porque os mesmos jogadores devem possuir habilidades em aspectos distintos. Nesse caso, uma alternativa de otimização que se apresenta trata-se da otimização multiobjetivo, em que se varia o peso de uma habilidade frente à outra na função objetivo. Ahmed et al. (2013) realizaram a otimização de um time de críquete do principal torneio indiano (*Indian Premier League*), impondo variados limites salariais e sopesando a importância do desempenho nas rebatidas e nos arremessos. O resultado obtido foi uma frente de Pareto em que observa o *trade-off* entre o desempenho nos arremessos e nas rebatidas [Ahmed et al. 2013].

1.2. O jogo de beisebol

Na partida de beisebol, dois times com nove jogadores cada se enfrentam alternando-se entre ataque e defesa. São realizados nove turnos de ataque e defesa, chamados de entradas. Os jogadores do time que está no ataque possuem por objetivo atravessar as quatro bases do campo, anotando corridas. O avanço entre as bases é realizado quando ocorre uma rebatida.

A defesa, por sua vez, deve evitar que o time adversário anote essas corridas. A principal figura defensiva é o arremessador, cuja função consiste em lançar a bola para que o jogador adversário tente rebater, de maneira que dificulte o trabalho do rebatedor. Os demais jogadores do time, no turno de defesa, atuam fazendo lançamentos entre as bases ou apanhando a bola, para eliminar os adversários.

Com exceção do arremessador em algumas ligas, todos os jogadores de defesa atuam também no ataque. Na situação em que o arremessador não participa da rotação ofensiva, ele é substituído por um jogador conhecido como rebatedor designado, sendo que este não atua na defesa.

2. Objetivo

O objetivo do trabalho consiste em criar equipes de beisebol com diferentes ênfases em ataque e defesa por meio de otimização multiobjetivo. Para a montagem dessas equipes, serão utilizados dados reais da *Major League Baseball*, a principal liga estadunidense do esporte.

3. Metodologia

3.1. Variáveis e função objetivo

O elenco de um time da *Major League Baseball* conta com 25 jogadores, dentre os quais estão presentes os arremessadores iniciais, os arremessadores secundários – incluindo um fechador –, primeira base, segunda base, terceira base, interbases, campista esquerdo, campista central, campista direito e receptor. Nos times da *American League*, há também um rebatedor designado, que tem por função substituir o arremessador nos turnos de ataque.

Neste trabalho, foram construídos elencos contando com cinco arremessadores iniciais, um receptor, um primeira base, um segunda base, um terceira base, um interbases, um rebatedor designado e três defensores de campo externo – não havendo distinção entre campista esquerdo, central ou direito, por serem posições em que é comum o remanejamento de atletas. Para formar um elenco titular na *Major League Baseball*, são de fato necessários múltiplos arremessadores iniciais, uma vez que trata-se de uma posição em que o desgaste físico é maior, e o tempo de recuperação de um atleta dura cerca de cinco dias. Nas demais posições, os jogadores atuam em uma frequência quase diária.

A otimização a ser feita consiste em uma otimização multiobjetivo, em que haverá a variação da importância dada para o ataque ou defesa na constituição do elenco. A potência ofensiva de um time foi definida por meio da estatística AVG (*average* – média) de cada jogador, dada por:

$$\text{AVG} = \frac{\text{número de rebatidas válidas}}{\text{número de idas ao bastão}}$$

Quanto maior o AVG de um jogador, melhor ele é no ataque. Neste trabalho, foram considerados os AVG de todos os jogadores, exceto dos arremessadores.

O poderio defensivo dos times, por sua vez, será descrito pelo ERA (*earned runs average* – média de corridas merecidas) cedidas por seus arremessadores:

$$\text{ERA} = 9 * \frac{\text{número de corridas merecidas cedidas}}{\text{entradas arremessadas}}$$

Dado que o objetivo do arremessador consiste em evitar que o time adversário anote corridas, este jogador é melhor quanto menor seu ERA.

Como o AVG é melhor quanto maior, e o ERA é melhor quanto menor, optou-se por inverter o ERA, para que o problema de otimização se torne inteiramente de maximização. Outra medida tomada foi a normalização das variáveis, para que as duas – AVG e o inverso do ERA – variassem entre 0 e 1, 0 sendo o pior valor e 1 o melhor valor. Para o AVG, consideraram-se o maior e o menor valores global. Sendo assim, as variáveis utilizadas no problema de otimização foram:

$$AVG_{norm} = \frac{AVG - AVG_{min}}{AVG_{max} - AVG_{min}}$$

$$ERA^{-1}{}_{norm} = \frac{ERA^{-1} - ERA^{-1}{}_{min}}{ERA^{-1}{}_{max} - ERA^{-1}{}_{min}}$$

A função objetivo foi escrita como:

$$\text{Maximizar } f(\mathbf{X}) = \alpha \sum_{i=1}^{m(1)} x_{i,1} ERA^{-1}{}_{norm,i,1} + (1-\alpha) \sum_{j=2}^{8} \sum_{i=1}^{m(j)} x_{i,j} AVG_{norm,i,j}$$

O termo α diz respeito à otimização multiobjetivo. Seu valor pode variar entre 0 e 1, sendo mais próximo de 0 quanto maior a ênfase dada ao AVG e mais próximo de 1 quanto maior a ênfase dada ao ERA.

Na função objetivo, **X** é uma matriz de números binários, cada um deles atrelado a um jogador. Cada coluna da matriz refere-se a uma posição e, dentro dessas colunas, cada elemento corresponde a um jogador. O número $x_{i,j}$ (jogador i da posição j) assume o valor de 1 caso o jogador esteja presente no time formado, e 0 caso esteja ausente. A primeira coluna trata dos arremessadores, os quais são avaliados em relação ao ERA. As demais colunas correspondem às posições de rebatedores, avaliados, por sua vez, em relação à média de rebatidas. A correspondência entre as colunas da matriz e posições é dada na Tabela 1. Esta tabela também traz o número de jogadores disponíveis para cada posição. Na função objetivo, a primeira coluna da matriz aparece somente no primeiro termo do somatório, pois este trata dos arremessadores e de seus respectivos ERAs. Este somatório parte de $i = 1$ e termina em $i = m(1)$, em que $m(1)$ é o número de arremessadores disponíveis para escolha – primeira coluna da matriz de jogadores. No termo correspondente ao AVG, o somatório interno varia de $i = 1$ até $i = m(j)$, com $m(j)$ variando em função da coluna. Esta variação se deve ao fato de que a disponibilidade de jogadores é diferente para cada posição (Tabela 1). Como o arremessador (primeira coluna) não é considerado em relação ao AVG, o somatório externo ocorre da segunda até a oitava coluna.

Tabela 1: Construção da matriz de jogadores

Coluna	**Posição no jogo**	**Número de jogadores disponíveis**
1	Arremessador	56
2	Receptor	14
3	Primeira base	26
4	Segunda base	19
5	Terceira base	20
6	Interbases	20
7	Rebatedor designado	10
8	Campistas externos	49

As restrições da função objetivo são as seguintes:
$x_{i,j} = \text{binário}$
$\sum_{i=1}^{m(1)} x_{i,1} = 5$ (número de arremessadores)
$\sum_{i=1}^{m(2)} x_{i,2} = 1$ (número de receptores)
$\sum_{i=1}^{m(3)} x_{i,3} = 1$ (número de primeira base)
$\sum_{i=1}^{m(4)} x_{i,4} = 1$ (número de segunda base)
$\sum_{i=1}^{m(5)} x_{i,5} = 1$ (número de terceira base)

$\sum_{i=1}^{m(6)} x_{i,6} = 1$ (número de interbases)
$\sum_{i=1}^{m(7)} x_{i,7} = 1$ (número de rebatedores designados)
$\sum_{i=1}^{m(8)} x_{i,8} = 3$ (número de campistas externos)
$\sum_{j=1}^{8} \sum_{i=1}^{m(j)} x_{i,j} \text{Salário}_{i,j} \leq$ Teto salarial (restrição salarial)

Na restrição salarial, a matriz Salário contém os valores anuais recebidos por cada jogador, estando esses valores organizados da mesma forma que a matriz de jogadores, havendo correspondência entre os elementos de cada posição nas duas matrizes.

3.2. Coleta de dados e implementação do problema

Para a criação dos times, foram coletados dados estatísticos relativos à temporada de 2017 da *Major League Baseball* [Fangraphs 2017]. Na posição de arremessador inicial, foram considerados apenas os atletas qualificados para as premiações da temporada, ou seja, com 162 entradas arremessadas no ano. Para as demais posições, consideram-se qualificados os atletas com pelo menos 502 passagens no bastão. Este critério foi adotado para todas as posições de rebatedores deste trabalho, exceto para os receptores, em que foram listados jogadores com a partir de 400 passagens no bastão.

Foram coletados também os salários anuais de todos os jogadores listados [Spotrac 2017], a fim de se impor a restrição salarial. A restrição para o ano de 2017 na *Major League Baseball* foi de US$ 195 milhões para o elenco de 25 jogadores [Major League Baseball 2017]. Neste trabalho, adotou-se uma restrição salarial de US$ 100 milhões, uma vez que não se criou um elenco completo, e a adoção da restrição real faria com que o melhor time pudesse ser montado para qualquer situação.

Para a resolução do problema de otimização, a função objetivo foi maximizada por meio do método Simplex LP, implementado no *solver* do Microsoft Excel 2016. O programa foi executado em um Windows 10 com processador Intel i5-5200U dual-core de 2,2 GHz e memória RAM de 8GB. Foram utilizados valores de α de 0,01 até 0,99, com passos de 0,01.

4. Resultados

A Tabela 2 mostra os elencos obtidos para as diferentes faixas de valores de α, juntamente com os ERA e AVG médios obtidos e o valor da folha salarial anual.

Para valores menores de α (mais próximos de 0), foram obtidos elencos com maior média de rebatidas e maior número de corridas merecidas cedidas – times com maior potencial ofensivo. Para valores de α mais próximos de 1, os times possuem uma média inferior de rebatidas, porém cedem menos corridas merecidas – melhor desempenho defensivo. Este é o resultado esperado, de acordo com a definição de α.

A diferença entre as equipes para $\alpha = 0{,}56$ e α entre 0,57 e 0,66 está na troca do arremessador Max Scherzer e do primeira base Eric Hosmer pelo arremessador Clayton Kershaw e primeira base Trey Mancini. Kershaw apresenta um ERA melhor que o de Scherzer (2,31 contra 2,51), entretanto seu salário é cerca de US$ 13 milhões mais elevado. Por esse motivo, o time com menor α conta com o arremessador mais barato, e o dinheiro restante pôde ser empregado na aquisição de Eric Hosmer (AVG 0,318), um primeira base melhor que Trey Mancini (AVG 0,293). A partir de $\alpha = 0{,}67$, foram escolhidos os cinco arremessadores com menor ERA – o melhor cenário defensivo possível –, o que justifica a não alteração no elenco a partir deste valor.

Os jogadores com seus nomes em itálicos são os que estão presentes em todos os elencos montados. O arremessador Robbie Ray possui um salário de US$ 570.400,00 e um ERA de 2,89, o que o torna um jogador barato e vantajoso. O segunda-base Jose Altuve recebe US$ 4.687.500,00 – um

Tabela 2: Elencos obtidos para os diferentes valores de α

Posição	$\alpha = 0{,}01$	$0{,}02 < \alpha < 0{,}14$	$0{,}15 < \alpha < 0{,}55$	$\alpha = 0{,}56$	$0{,}57 < \alpha < 0{,}66$	$0{,}67 < \alpha < 0{,}99$
Arremessadores	*Robbie Ray* Gio Gonzalez Luis Severino Jimmy Nelson Aaron Nola	Corey Kluber *Robbie Ray* Gio Gonzalez Luis Severino Marcus Stroman	Corey Kluber Stephen Strasburg *Robbie Ray* Gio Gonzalez Luis Severino	Corey Kluber Max Scherzer Stephen Strasburg *Robbie Ray* Gio Gonzalez	Corey Kluber Clayton Kershaw Stephen Strasburg *Robbie Ray* Gio Gonzalez	Corey Kluber Clayton Kershaw Max Scherzer Stephen Strasburg *Robbie Ray*
Receptor	Buster Posey	Buster Posey	Buster Posey	Gary Sanchez	Gary Sanchez	Gary Sanchez
Primeira base	Joey Votto	Eric Hosmer	Eric Hosmer	Eric Hosmer	Trey Mancini	Trey Mancini
Segunda base	*Jose Altuve*	*Jose Altuve*	*Jose Altuve*	*Jose Altuve*	*Jose Altuve*	*Jose Altuve*
Terceira base	Justin Turner	Justin Turner	Jose Ramirez	Jose Ramirez	Jose Ramirez	Jose Ramirez
Interbases	Jean Segura	Jean Segura	Corey Seager	Corey Seager	Corey Seager	Corey Seager
Rebatedor designado	Nelson Cruz	Nelson Cruz	Nelson Cruz	Nelson Cruz	Nelson Cruz	Robbie Grossman
Campistas externos	Charlie Blackmon *Avisail Garcia* Marcell Ozuna	Charlie Blackmon *Avisail Garcia* Marcell Ozuna	Charlie Blackmon *Avisail Garcia* Marcell Ozuna	Charlie Blackmon *Avisail Garcia* Marcell Ozuna	Charlie Blackmon *Avisail Garcia* Marcell Ozuna	*Avisail Garcia* Marcell Ozuna Tommy Pham
ERA médio	3,172	2,834	2,720	2,626	2,586	2,496
AVG médio	0,3188	0,3186	0,3176	0,3129	0,3101	0,3027
Preço da equipe (US$)	99.462.652,00	99.971.652,00	96.651.385,00	96.623.390,00	98.342.561,00	99.483.318,00

*Os jogadores em itálico são aqueles que estão presentes em todos os elencos

valor intermediário –, mas possui uma média de rebatidas igual a 0,346, o maior valor entre todos os jogadores considerados. O campista externo Avisail Garcia possui o segundo maior AVG de sua categoria (0,330), e um salário de 3 milhões de dólares – valor considerado intermediário.

A Figura 1 mostra a Frente de Pareto obtida para a otimização multiobjetivo. Nela, é possível observar o *trade-off* entre o AVG e o inverso do ERA, as duas estatísticas utilizadas na função objetivo. O gráfico ilustra o crescimento de uma variável atrelado ao decrescimento da outra, resultado esperado nesse tipo de problema. A existência do *trade-off* evidencia que a presença de um limite salarial influencia diretamente na montagem de um elenco e que o teto salarial empregado neste trabalho foi eficaz ao ilustrar essa realidade.

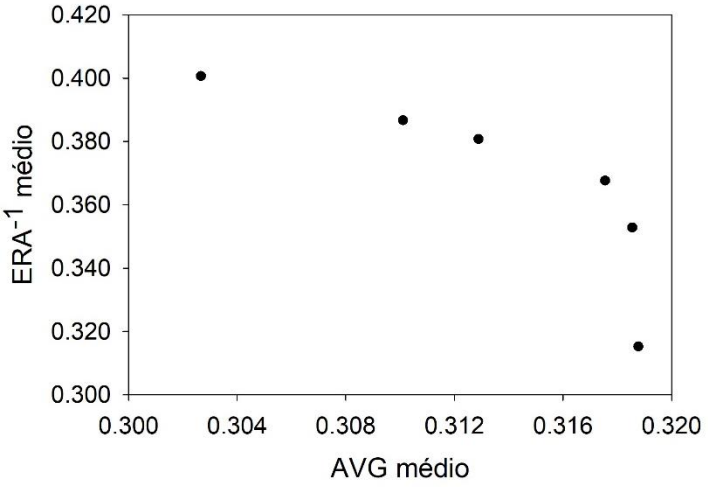


Figura 1: Frente de Pareto para o AVG e o inverso do ERA

5. Conclusão

A proposta deste trabalho consistiu na montagem de equipes de beisebol com diferentes ênfases em ataque e defesa. A estatística ofensiva utilizada foi a razão entre as rebatidas e o número de idas ao bastão (AVG – *average*), e a estatística defensiva foi a média de corridas cedidas a cada nove entradas (ERA – *Earned runs average*). A existência de um limite salarial impõe um *trade-off* entre esses dois aspectos: para que um deles melhore, há a necessidade de reduzir o outro.

O problema de otimização foi resolvido por meio da metodologia de otimização multiobjetivo, com a variável α ditando a relevância do ataque ou da defesa no elenco. Com o valor de α variando de 0,01 até 0,99, em incrementos de 0,01, foram obtidas seis diferentes composições, atendendo à restrição salarial imposta no algoritmo. Esses resultados, dispostos na Frente de Pareto, ilustram graficamente o esperado *trade-off* entre o ERA e a AVG.

A resolução deste problema envolveu baixo esforço computacional, evidenciando a potencialidade do método para a avaliação de um elenco completo de um time da *Major League Baseball* e o uso de múltiplas estatísticas representativas do desempenho dos jogadores. Esta metodologia pode ser aplicada também a outros esportes com funcionamento similar ao do beisebol, em particular frente a um amplo banco de dados e estatísticas relativos aos atletas.

6. Referências bibliográficas

Ahmed, F., Deb, K. e Jindal, A. (2013). Multi-objective optimization and decision making approaches to cricket team selection. *Applied Soft Computing*, 13:402-414.

Añon, I. C., Scaglia, A. J. e Torezzan, C. (2017). Um modelo de programação linear inteira para otimização de recursos na designação de árbitros para o campeonato brasileiro de futebol. *Pesquisa Operacional para o Desenvolvimento*, 9:10-17.

Fangraphs (2017). Major League Team Stats. https://www.fangraphs.com/.Acessado: 2018-10-21.

Hoshino, R. e Kawarabayashi, K. (2011). A multi-round generalization of the traveling tournament problem and its application to Japanese baseball. *European Journal of Operational Research*, 215:481-497.

Major League Baseball (2017). Competitive Balance Tax. http://m.mlb.com/glossary/transactions/competitive-balance-tax/. Acessado: 2018-10-20.

Mulholland, J. e Jensen, S. T. (2019). Optimizing the allocation of funds of an NFL team under the salary cap. *International Journal of Forecasting*, 35:767-775.

Spotrac (2017). MLB Salary Rankings. https://www.spotrac.com/mlb/rankings/2017/. Acessado: 2018-10-20.