

# A Evolução do Ataque Marshall

## Análise de desempenho dos jogadores ao longo do tempo

Daniel Vieira  
Eduardo Darrazão

### Abstract

Neste trabalho analisamos partidas de xadrez para capturar a difusão de novas ideias dentro do ataque Marshall, suas propagações no tempo e espaço, capturando fatores que facilitam e/ou dificultam a implementação das mesmas.

### 1 Introdução

O Ataque Marshall tem uma boa reputação e por isso muitos jogadores que jogam a abertura espanhola (Ruy Lopez) com as peças brancas temem a aplicação do Ataque Marshall por parte das pretas. É uma decisão arriscada que produz dificuldades para os dois jogadores. Neste trabalho queremos entender como essas dificuldades podem ser capturadas e como podemos compreender a evolução das jogadas.

Na descrição dos dados, abordamos dois aspectos principais: popularidade e domínio (resultados das partidas e desequilíbrio durante a partida) de cada nova ideia. No fim, construímos doze modelos de regressão linear múltipla para projetar a dificuldade que os jogadores terão ao enfrentar adversários nas variantes classificadas neste trabalho como relevantes.

### 2 Materiais e Métodos

Os registros de jogos são armazenados no formato (.pgn) contendo os dados das jogadas e metadados da partida. Usamos diferentes fontes de dados para cruzar informações e tratar as falhas dos dados, assim obtendo um conjunto completo e uniforme: *pgnmentor.com*, *365chess.com*, *chessbase.com*, *fide.com*.

**Table 1.** Jogos por variante

Variante	Jogos
9. d3 dxe4 10. dxe4 Qxd1	18
9. d4 Nxe4 10. dxe5 Be6	11
9. d4 exd4 10. e5 Ne4	31
9. exd5 Nxd5 10. Nxe5 Nxe5	957
9. exd5 Nxd5 10. d4 exd4	26
9. exd5 e4 10. Ng5 4	11
9. exd5 e4 10. dxc6 exf3	41

In: IV Concurso de Teses e Dissertações (CTD 2022), Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

© 2024 SBC – Sociedade Brasileira de Computação.

ISSN 2596-1683



**Figure 1.** Posição inicial do ataque Marshall. É uma formação de peças para as pretas contra a abertura Ruy Lopez. O ataque Marshall é uma variação da abertura de xadrez "Ruy Lopez" (ou abertura espanhola) e o nome do ataque é em homenagem ao seu criador, o Grão-Mestre Frank James Marshall.

As bases de dados do *ChessBase* possuem jogos catalogados desde o ano de 1842, onde para cada partida há informações referentes a: quem estava jogando com as peças brancas e quem estava jogando com as peças pretas, o ELO (*rating*) de cada jogador, o resultado, a quantidade de jogadas, o ECO (Código da abertura, 'Encyclopaedia of Chess Openings'), o torneio e em que *round* foi jogada a partida, a data, as jogadas em notação algébrica<sup>1</sup>, material restante no final, média do ELO e ID da partida.

Selecionamos as principais variantes (tabela 1) que são jogadas depois da sequência que leva ao ataque Marshall (1. e4 e5 2. Nf3 Nc6 3. Bb5 a6 4. Ba4 Nf6 5. O-O Be7 6. Re1 b5 7. Bb3

<sup>1</sup>Notação do tipo: <Lance>. <Movimento de uma peça branca> <Movimento de uma peça preta>, exemplo: 1. Cf3 d5 representa que Cavalo branco foi para a coordenada f3 e o peão da coluna 'd' preto foi para a coordenada d5. Com peças indo para a coordenada e2, a notação das peças em inglês é (King = K (e.g. Ke2), Queen = Q (e.g. Qe2), Rook = R (e.g. Re2), Bishop = B (e.g. Be2), Knight = N (e.g. Ne2) e a notação do peão é convencionalmente omitida (e.g. e2))

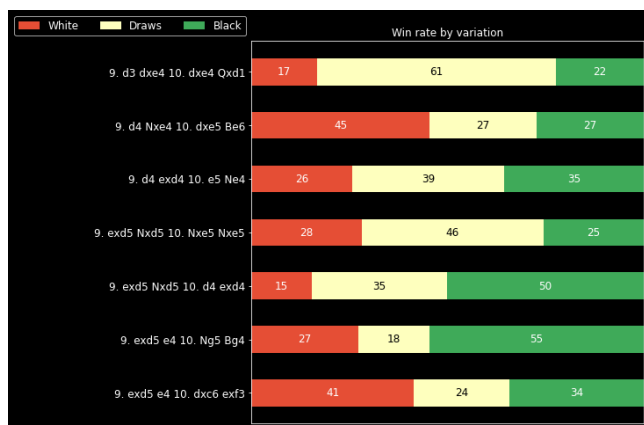


Figure 2. Resultados por variante

O-O 8. c3 d5) filtrando pela jogadas que, desde sua primeira aparição, foram jogadas pelo menos 10 vezes por jogadores com 2500 ou mais de ELO. De acordo com a popularidade, a principal variante é a variante original (9. **exd5 Nxd5 10. Nxe5 Nxe5**, criada pelo próprio Marshall), seguida muito de longe por outras variantes que classificamos como esporádicas, pois percebemos que são ideias criadas ou resgatas para surpreender outros jogadores.

Analisando o domínio (Figura 3)) e os resultados (Figura 2) para compreender o que faz uma variante ser popular, percebemos que para uma variante do ataque Marshall é fundamental que ela ofereça para ambos jogadores perspectivas relativamente interessantes, onde cada jogador faz concessões para conseguir alguma vantagem exclusiva e dessa forma tentar produzir desequilíbrio nas estratégias, o que demonstra que é o ataque Marshall possui intrinsecamente uma ambição de minimizar a probabilidade de empate sob a condição de correr riscos moderados.

No jogo de xadrez, uma variante é definida pelos dois jogadores, através de um acordo em que cada um faz uma concessão conforme sua estratégia para a partida. É comum um jogador diminuir suas chances de vitórias para diminuir suas chances de derrota (fazendo um jogo mais seguro e equilibrado levando possivelmente a um empate), por outro lado também é plausível que um jogador que precisa da vitória se colocar em um risco maior de derrota para poder minimizar o equilíbrio e, consequentemente, minimizar as chances de empate.

Por causa do acordo dinâmico, podemos perceber que algumas variantes que apresentam frequentemente resultados muito favoráveis para apenas um lado perdem popularidade em pouco tempo (figura 4), justamente porque com o tempo os jogadores aprenderam que aquele caminho era provavelmente desfavorável e então não aceitavam os acordos dessas variantes (executando lances diferentes dos necessários para caracterizar a variante desfavorável).

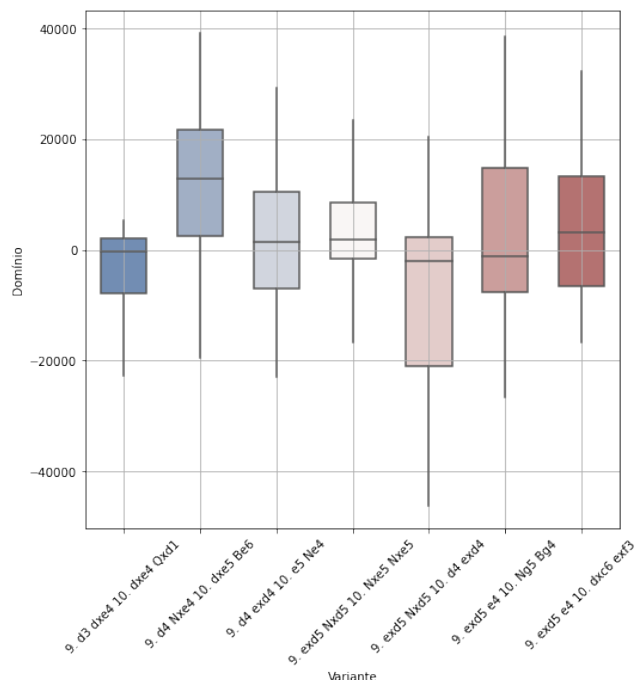


Figure 3. Domínio das partidas por variante

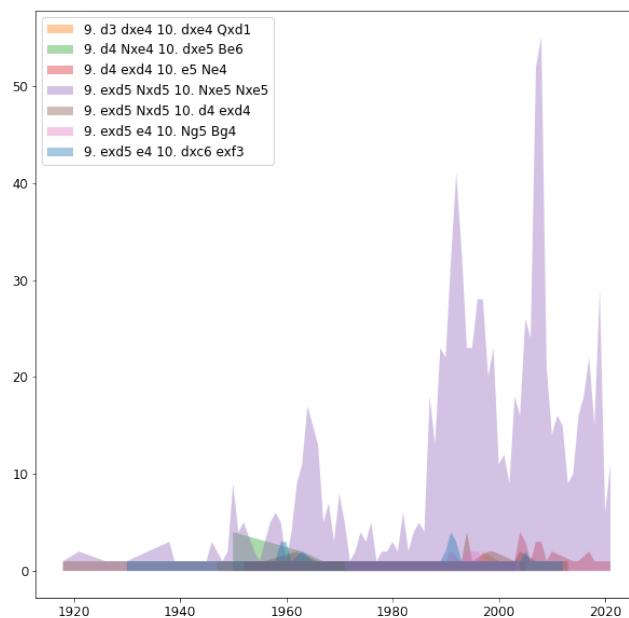


Figure 4. Partidas de cada variante por ano.

Combinamos os dados das partidas com os dados dos jogadores para tentar projetar a dificuldade que cada perfil de jogador tende a encontrar dada uma variante dentre as avaliadas. Para isso consideramos dois modelos para cada cor sendo que cada um tem como variável dependente o número de movimentos bons e ruins respectivamente. Acreditamos

que variantes apresentam mais dificuldade para determinados contextos e pretendemos capturar isso avaliando o número de jogadas consideradas ruins quando comparadas com as jogadas sugeridas pelo *Stockfish*. As hipóteses iniciais que avaliamos são:

- O ELO do jogador de pretas, O ELO do jogador de brancas, o ano da partida e a variante jogada influenciam o número de **lances ruins** (variável dependente) que o jogador de **brancas** cometerá.
- O ELO do jogador de pretas, O ELO do jogador de brancas, o ano da partida e a variante jogada influenciam o número de **lances ruins** (variável dependente) que o jogador de **pretas** cometerá.
- O ELO do jogador de pretas, O ELO do jogador de brancas, o ano da partida e a variante jogada influenciam o número de **lances bons** (variável dependente) que o jogador de **brancas** cometerá.
- O ELO do jogador de pretas, O ELO do jogador de brancas, o ano da partida e a variante jogada influenciam o número de **lances bons** (variável dependente) que o jogador de **pretas** cometerá.

Após a análise inicial dos lances bons e ruins, seguimos a orientação do nosso professor Thiago Silva e tentamos criar métricas para capturar elementos culturais que podem influenciar na partida. Historicamente, é parte do senso comum dos entusiastas de xadrez que alguns países possuem uma tradição cultural forte, que influencia significativamente o estilo de jogo e desempenho geral dos jogadores daquele país. Por isso analisamos as hipóteses iniciais adicionando a região que o jogador representa (não necessariamente é o local de nascimento, representa a federação que ele defende) como variável independente, para capturar a influência da região na projeção dos lances bons e ruins. Dessa forma foi produzido mais um modelo para cada hipótese, totalizando oito modelos.

Dado o contexto de aplicação desse trabalho decidimos distribuir os jogadores de acordo com suas federações para as seguintes regiões (como grupos): Américas, Ásia, Europa, Oceania, África, URSS. Sendo que URSS inclui os países que durante século XX fizeram parte do União Soviética, as outras regiões são mapeamentos diretos para seus respectivos continentes.

Outro conjunto de hipóteses foi feito de forma similar as regiões mas dessa vez com as federações diretamente. Novamente, foram adicionados mais um modelo para cada hipótese inicial, para testar a influência das federações, totalizando finalmente doze modelos.

### 3 Resultados

Os quatro modelos que possuíam os jogadores distribuídos por região foram descartados em razão de diluir as informações contidas em cada país, dado que as regiões não obtiveram nenhuma relevância estatística nos modelos.

A tabela 2 representa o impacto da variável que contém a informação da federação do jogador com as peças brancas nos modelos que projetam a quantidade de jogadas boas e ruins das brancas. O valor corresponde ao coeficiente da variável no modelo e os asteriscos expressam a relevância estatística (*P-Value*). Podemos inferir que os países citados influenciam a quantidade de lances bons ou ruins do jogador de brancas, principalmente para os jogadores da Bielorrússia, Geórgia, Nova Zelândia e do Uzbequistão.

**Table 2.** Federação (brancas) por lance (brancas)

White Federation	White Good	White Bad
Belarus	26.841**	
Bulgaria	16.015*	
Georgia	38.739***	
India	19.512*	
Netherlands	16.641*	
New Zealand	28.926**	
United States	18.311*	
Uzbekistan	33.964***	

Na tabela 3 temos o impacto da variável que contém a informação da federação do jogador com as peças brancas nos modelos que projetam a quantidade de jogadas boas e ruins das pretas, respectivamente. Podemos compreender estas informações de modo que quando o jogador com as brancas é de algum destes países, o jogador com as peças pretas realiza mais jogadas boas ou ruins, dependendo do país.

Acreditamos que os casos dos países em que o número de lances bons das pretas aumentam em função da nacionalidade do jogador de brancas implicam em um jogo mais fácil para o jogador de pretas devido a algum elemento (possivelmente cultural) no estilo de jogo do jogador de brancas. A mesma crença se aplica de forma relativa no sentido oposto: quando há um aumento em jogadas ruins das pretas, pode ser devido a um fator cultural no estilo de jogo das brancas que atrapalha a tomada de decisão do jogador de pretas.

**Table 3.** Federação (brancas) por lance (pretas)

White Federation	Black Good	Black Bad
Belarus	27.871*	
Georgia	37.917**	
Lithuania		12.295*
New Zealand	25.808*	
Switzerland		14.729***
United States		7.144*
Uzbekistan	36.904***	

Na tabela 4 temos o impacto da variável que contém a informação da federação do jogador com as peças pretas nos

modelos que projetam a quantidade de jogadas boas e ruins das brancas. A interpretação é similar a 3, com a inversão das cores. Percebemos um fenômeno em Kosovo, onde o efeito nas jogadas ruins é negativo, ou seja, diminuí a quantidade de erros, e de certo modo facilita a tomada de decisão do jogador de brancas.

**Table 4.** Federação (pretas) por lance (brancas)

Black Federation	White Good	White Bad
Argentina	27.951*	
Belarus	24.719*	
Brazil	29.521**	
Denmark	15.872*	
Spain	29.347*	
Iceland	41.872***	
Kosovo		-12.794*
Philippines	33.964***	
Poland	19.679**	
Sweden	22.649**	

A tabela 5 apresenta o efeito da informação da federação do jogador com as peças pretas nos modelos que projetam a quantidade de jogadas boas e ruins das pretas. A análise é semelhante a 2, entretanto conseguimos constatar (similamente à 4) que onde o coeficiente é negativo o efeito é o oposto. De acordo com nossa análise, no caso de Kosovo podemos concluir que existe uma tendência a jogos mais simples e equilibrados, em que ambos os lados encontram os lances bons mais facilmente.

**Table 5.** Federação (pretas) por lance (pretas)

Black Federation	Black Good	Black Bad
Bosnia		7.319*
Greece		8.052*
Iceland	39.320***	
Kosovo		-13.699**
Philippines	36.904***	
Sweden		6.741*
Wales		17.047***

Os próximos modelos foram construídos desconsiderando a informação da federação dos jogadores, de forma que ampliou o efeito e relevância das demais variáveis independentes, para testar as hipóteses iniciais isoladamente.

Na tabela 6 temos o impacto das variantes, do ELO e do ano na qualidade das jogadas das brancas, e em 7 na qualidade das jogadas das pretas.

Conjecturamos que quanto mais uma variante penaliza a quantidade de bons movimentos e aumenta a quantidade de jogadas ruins, mais difícil ela é, pois torna-se difícil encontrar os melhores lances. Podemos inferir também que

variantes que penalizam similarmente ambas as cores são mais balanceadas (casos em que as condições são equivalentes mesmo que não sejam iguais) para ambos os lados. Em relação ao ano, percebemos que com o passar do tempo aumenta-se jogadas boas e diminuem os lances ruins, o que indica que os jogadores aprendem com os erros do passado. E por fim na análise do ELO vemos que um ELO maior de cada cor impacta no aumento de boas jogadas e diminui jogadas ruins na cor respectiva, e ainda captura um fenômeno conhecido no xadrez que é a vantagem intrínseca das brancas, onde o ELO do jogador de brancas impacta (apesar de muito fracamente) no aumento de jogadas ruins das pretas, porque o jogador de brancas decide a direção inicial do acordo que definirá a variante.

**Table 6.** Features básicas por lance (brancas)

Variable	White Good	White Bad
9. d4 Nxe4 10. dxe5 Be6	-5.380*	
9. d4 exd4 10. e5 Ne4	-7.461***	
9. exd5 Nxd5 10. d4 exd4	-5.283**	
9. exd5 e4 10. Ng5 Bg4		4.880**
9. exd5 e4 10. dxc6 exf3	-9.789***	
WhiteElo	0.008***	-0.006***
Year	0.129***	-0.021**

**Table 7.** Features básicas por lance (pretas)

Variable	Black Good	Black Bad
9. d4 exd4 10. e5 Ne4	-7.747***	
9. exd5 Nxd5 10. d4 exd4	-5.776*	
9. exd5 e4 10. Ng5 Bg4	-10.519*	5.425***
9. exd5 e4 10. dxc6 exf3	-11.905***	
BlackElo	0.005***	-0.007***
WhiteElo		0.001*
Year	0.127***	

## 4 Conclusão

A análise estatística e a regressão pareceram coerentes com as hipóteses que desenvolvemos e com as interpretações que propusemos no texto. O modelo de regressão linear múltipla apresentou resultados interessantes, com coeficiente de determinação relativamente baixo porém com correlações expressivas e estatisticamente significativas.

Nas regressões obtivemos um  $R^2$  de até  $0.4^2$ , o que é um resultado bastante satisfatório levando em consideração que utilizamos poucas *features* para caracterizar o perfil dos jogadores.

<sup>2</sup>Outros valores de coeficiente de determinação que obtivemos: [0.405; 0.335; 0.388; 0.3640.066; 0.147; 0.045; 0.135].