

Vilson Raimundo Ferreira

ANÁLISE DE DESEMPENHO E PREVISÃO DE FINALISTAS EM EVENTOS ESPORTIVOS DE TIRO COM ARCO

Trabalho de Conclusão de Curso apresentado na
Escola de Administração de Empresas de São Paulo
da Fundação Getúlio Vargas, para obtenção de
certificado de conclusão de Curso de Especialização.

São Paulo

2017

Vilson Raimundo Ferreira

**ANÁLISE DE DESEMPENHO E PREVISÃO DE FINALISTAS EM
EVENTOS ESPORTIVOS DE TIRO COM ARCO**

Coordenador do Curso
José Luiz C. Kugler

Avaliação
Satisfatório Não Satisfatório

São Paulo

2017

ANÁLISE DE DESEMPENHO E PREVISÃO DE FINALISTAS EM EVENTOS ESPORTIVOS DE TIRO COM ARCO

A gestão de desempenho nos esportes sempre foi uma grande aliada na preparação de atletas de alto rendimento, objetivando seu autoconhecimento, acompanhamento contínuo, consistência, evolução, diagnóstico de mudanças de treinamento e comportamento. Mais recentemente, tem sido de extrema importância na análise de competitividade para detectar técnicas, táticas e condutas dos concorrentes para direcionar estratégias de um treinamento, com foco no curto prazo para obter resultados em uma competição ou para um objetivo específico. A estatística é ferramenta vital para tornar isso possível, e, consequentemente, contribui também para a busca de previsibilidade baseada na coleta, organização e processamento dos dados com aplicação de algoritmos de agrupamento e algoritmos preditivos, que podem assim permitir visualizar consistência dos competidores, grupos de competitividade, tendências baseadas em resultados obtidos, e, porque não, prever possíveis finalistas em uma competição. Este projeto visa a testar alguns destes algoritmos no esporte tiro com arco com o intuito de aplicar métodos que possam descobrir grupos de competidores de desempenho semelhante, que, portanto, competem entre si, e chegar em um cenário de possíveis finalistas com tendência a conquistar uma medalha olímpica.

Palavras-chave

Previsão, desempenho, esporte, tiro com arco,

SUMÁRIO

1. INTRODUÇÃO	5
1.1 O ESPORTE TIRO COM ARCO	5
1.2 DINÂMICA DAS COMPETIÇÕES NOS JOGOS OLÍMPICOS	8
2. CONTEXTO	10
2.1 É POSSÍVEL PREVER FINALISTAS DE UMA COMPETIÇÃO?.....	10
2.2 COMO UM ATLETA OBTÉM O FAVORITISMO	16
2.3 MÉTRICAS DA FEDERAÇÃO INTERNACIONAL.....	17
3. ABORDAGEM METODOLÓGICA	20
3.1 PROCESSO APLICADO E ETAPAS	20
3.2 ORGANIZAÇÃO DE DADOS E MÉTRICAS	23
3.3 ANÁLISE E SELEÇÃO DE VARIÁVEIS.....	25
4. ANÁLISES E RESULTADOS	31
4.1 AGRUPAMENTOS BASEADO NO RANKING	31
4.2 AGRUPAMENTOS BASEADO NA PONTUAÇÃO MÁXIMA	34
4.3 AGRUPAMENTOS BASEADO NA ZONA AMARELA.....	35
4.4 AGRUPAMENTOS BASEADO EM ANÁLISE DE CLUSTER	36
4.5 PREVISÃO USANDO ALGORITMO DE CLASSIFICAÇÃO.....	42
4.6 PREVISÃO USANDO ALGORITMO DE REGRESSÃO	45
CONCLUSÃO	49
REFERÊNCIAS BIBLIOGRÁFICAS	51

1. INTRODUÇÃO

Atletas de alto desempenho têm como rotina de vida: planejamento, treinamento, objetivos, metas, resultados e adaptação constante. Esta rotina é válida para qualquer esporte, seja ele individual ou coletivo, e não é diferente para o tiro com arco, que é, atualmente, um esporte olímpico. Esta modalidade permaneceu nos jogos olímpicos de 1900 até 1920 e retornou em 1972, sendo um dos esportes mais prestigiados destes jogos pela sua dificuldade e precisão dos atletas.

Neste esporte, a consistência é fator predominante devido ao volume necessário de tiros para garantir uma boa classificação e pela necessidade de manter os resultados em um raio o mais curto possível, que representa maior precisão e desempenho competitivo para as etapas eliminatórias. A seguir, vamos entender um pouco do funcionamento desta modalidade esportiva e como acontece a competição individual, bem como o sistema de pontuação e métricas, que são de entendimento necessário para a coleta e análise dos dados.

1.1 O esporte tiro com arco

O esporte tiro com arco pode ser competido com dois tipos de arcos: o arco recurvo, também chamado de arco olímpico, e o arco composto. O arco recurvo é uma versão renovada do arco tradicional de uma corda, com algumas evoluções em seus materiais, que objetivaram reduzir interferências externas ao arqueiro, como, por exemplo vibrações e instabilidades mecânicas, sem interferir na mecânica corporal do arqueiro e seus reflexos. O arco composto é um arco bem mais evoluído mecanicamente, tendo como principal diferença suas três cordas, duas roldanas e uma mira com lente, que o torna mais preciso por exigir menos esforço do arqueiro, atingindo maior potência, precisão, velocidade de voo da flecha e maior visibilidade do alvo. Nos jogos olímpicos, somente o arco recurvo é utilizado, o que permite que o arqueiro se mantenha no total controle do tiro.



Figura 1 - Atleta Kim Woojin, recordista olímpico com arco recurvo



Figura 2 - Atleta Reo Wilde, recordista mundial com arco composto

Dadas as diferenças citadas, as competições são separadas primeiramente pelos tipos de arcos, ou seja, atletas de arco recurvo não competem com atletas de arco composto e vice-versa. Conforme descrito no regulamento¹ da World Archery,² no arco recurvo, os atletas atiram a uma distância de 70 metros contra um alvo de 122 centímetros de diâmetro contendo 10 zonas de pontuações (de 1 a 10) e um centro utilizado como zona de desempate chamado de xis. As classificações são separadas entre homens e

¹ World Archery. **Rolebook Target Archery, Book 3**. Disponível em: <<https://rulebook.worldarchery.org/PDF/Official/2016-10-01/EN-Book3.pdf>>. Acesso em: 21 nov. 2016.

² Federação Internacional de Tiro com Arco, organização sem fins lucrativos atuante como principal regulador do esporte no mundo, denominado a partir de 2015 como World Archery (www.worldarchery.org).

mulheres, e, sendo assim, há duas categorias principais: recurvo masculino (RM) e recurvo feminino (RF). No arco composto, os atletas atiram a uma distância de 50 metros contra um alvo de 80 centímetros de diâmetro contendo 6 zonas de pontuação (de 6 a 10) e um centro utilizado como zona de desempate, também chamado de xis. Da mesma forma, as classificações são divididas entre homens e mulheres, e há duas categorias principais: composto masculino (CM) e composto feminino (CF).

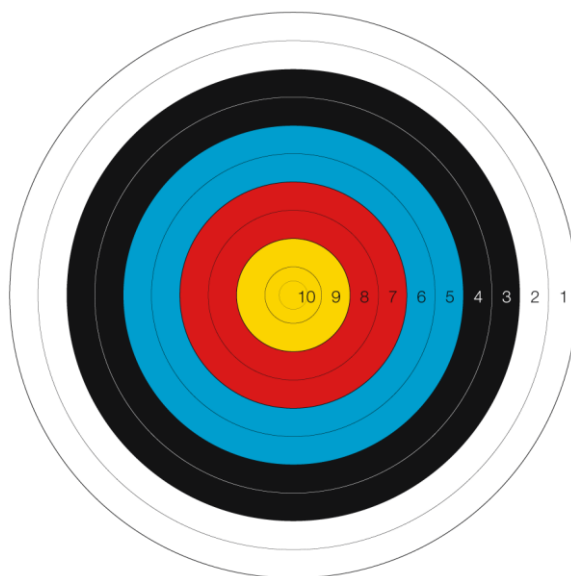


Figura 3 - Alvo utilizado por arco recurvo

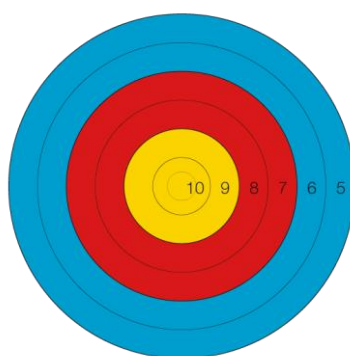


Figura 4 - Alvo utilizado por arco composto

As competições normalmente têm sua importância baseada em uma hierarquia regionalizada. No caso do Brasil acontecem os Campeonatos Estaduais, com etapas

mensais e premiação anual, o Campeonato Brasileiro, como evento único anual, e o Campeonato Pan-americano, como evento único a cada 4 anos. Seus resultados contribuem para pontuação em uma classificação nacional para permitir a seletividade nas duas principais competições: o Campeonato Mundial, com premiação anual, baseado em 4 etapas no ano, e os Jogos Olímpicos, evento único a cada 4 anos. Ambas as competições normalmente são divididas em duas etapas – classificação e eliminatórias –, com o objetivo de concluir com três finalistas e medalhistas de ouro, prata e bronze.

1.2 Dinâmica das competições nos jogos olímpicos

Nos jogos olímpicos os atletas competem individualmente em duas etapas chamadas de Round Classificatório e Round Eliminatório.

No Round Classificatório, todos os atletas atiram juntos e disparam 72 flechas no alvo divididos em dois rounds: o primeiro round de 6 séries de 6 tiros, totalizando 36 tiros, uma pausa normalmente de 15 a 30 minutos e mais um round de 6 séries de 6 tiros, totalizado assim os 72 tiros e 720 pontos máximos possíveis. Após os 72 tiros, as pontuações são somadas e assim é gerada a classificação final, na qual somente os 64 primeiros atletas estão aptos a competir na etapa seguinte.



Figura 5 - Atletas atirando no round classificatório

No Round Eliminatório, também muito conhecido como combates, os 64 atletas melhores classificados no round anterior são separados em duplas: o 1º colocado disputa contra o 64º, o 2º contra o 63º e assim por diante. Cria-se um chaveamento de 1/32, onde cada dupla atira junta e o vencedor continua na competição para disputar os próximos chaveamentos de 1/16, 1/8, quartas de finais, semifinais e finais (disputas de ouro e bronze). Em cada etapa das chaves os atletas atiram no máximo 5 *sets*, que são pontuados da seguinte forma:

- Em cada *set* os atletas atiram 3 flechas para obter a pontuação máxima de 30 pontos;
- Quem fizer mais pontos em um *set* ganha 2 pontos. Se o *set* tiver empate cada atleta ganha 1 ponto;
- A eliminatória tem no máximo 5 *sets*. Vence quem fizer primeiro 6 pontos ou mais;
- Caso haja um empate ao final de 5 *sets*, os atletas terão que realizar o *shoot-off*, também conhecido como tiro de morte. Cada atleta atira apenas uma flecha e vence aquele que pontuar mais próximo ao centro.

Esta dinâmica de eliminatória acontece até a disputa de ouro, na qual competem os dois atletas que não perderam nenhum combate, e a disputa de bronze, com atletas que perderam apenas nas semifinais.



Figura 6 - Atletas disputando em uma chave no round eliminatório

2. CONTEXTO

2.1 É possível prever finalistas de uma competição?

No esporte tiro com arco, a grande busca dos atletas, antes de focar um alto desempenho, é conseguir manter uma consistência em seus resultados. Entende-se por consistência o atleta conseguir, em seus treinos e em suas participações em competições, manter rigidamente resultados com a menor oscilação possível entre as séries. Como citado anteriormente, durante a etapa de classificação, os atletas atiram 72 flechas divididas em 12 séries de 6 tiros cada. Isso significa que a cada série o atleta tem condições de atingir o máximo de 60 pontos, e desta forma, a consistência refere-se a manter uma pontuação semelhante em cada série. Uma pontuação consistente normalmente é apresentada com uma diferença máxima de 3 pontos entre as séries, seja para mais ou para menos. Como exploração dos dados, visualizamos a seguir os resultados dos 4 melhores atletas nos Jogos Olímpicos Rio 2016, na fase classificatória, e como a consistência acontece:

Rio 2016 Olympic Games Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016											
Archer: Kim Woojin											
Country: KOR - Korea											
						Total	16B				
						0	R M				
70m-1	1	2	3	Sum	Tot.	Total	10+X	X			
1	X	X	X	30			3	3			
	10	9	9	28	58	58	1				
2	X	X	X	30			3	3			
	9	9	9	27	57	115					
3	X	X	10	30			3	2			
	10	9	9	28	58	173	1				
4	10	10	10	30			3				
	10	10	9	29	59	232	2				
5	X	X	X	30			3	3			
	X	10	10	30	60	292	3	1			
6	10	10	9	29			2				
	9	9	9	27	56	348					
Total						348	348	24	12		

Rio 2016 Olympic Games Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016											
Archer: Kim Woojin											
Country: KOR - Korea											
						Total	16B				
						348	R M				
70m-2	1	2	3	Sum	Tot.	Total	10+X	X			
1	X	X	X	30			3	3			
	10	10	9	29	59	59	2				
2	X	X	X	30			3	3			
	10	10	10	30	60	119	3				
3	10	10	10	30			3				
	9	9	9	27	57	176					
4	X	X	10	30			3	2			
	10	9	9	28	58	234	1				
5	X	X	10	30			3	2			
	10	10	9	29	59	293	2				
6	X	X	X	30			3	3			
	X	10	9	29	59	352	2	1			
Total						352	700	28	14		

Figura 7 - Planilha de resultados do 1º colocado na classificação Rio 2016

Rio 2016 Olympic Games
Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016

Archer: Ellison Brady
Country: USA - USA

7C

70m-1	1	2	3	Sum	Tot.	Total	10+X	X
1	X	X	10	30			3	2
	10	9	9	28	58	58	1	
2	X	X	10	30			3	2
	9	9	7	25	55	113		
3	X	10	10	30			3	1
	10	9	9	28	58	171	1	
4	X	10	10	30			3	1
	9	9	9	27	57	228		
5	X	X	10	30			3	2
	9	9	9	27	57	285		
6	X	X	X	30			3	3
	9	9	9	27	57	342		
Total						342	342	20 11

Rio 2016 Olympic Games
Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016

Archer: Ellison Brady
Country: USA - USA

7C

70m-2	1	2	3	Sum	Tot.	Total	10+X	X
1	X	X	10	30			3	2
	10	10	9	29	59	59	2	
2	10	9	9	28			1	
	9	9	9	27	55	114		
3	X	X	10	30			3	2
	9	9	9	27	57	171		
4	X	X	X	30			3	3
	10	10	9	29	59	230	2	
5	X	10	10	30			3	1
	10	9	9	28	58	288	1	
6	10	10	10	30			3	
	10	10	10	30	60	348	3	
Total						348	690	24 8

Figura 8 - Planilha de resultados do 2º colocado na classificação Rio 2016

Rio 2016 Olympic Games
Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016

Archer: Pasqualucci David
Country: ITA - Italy

12C

70m-1	1	2	3	Sum	Tot.	Total	10+X	X
1	10	9	9	28			1	
	9	9	8	26	54	54		
2	X	10	9	29			2	1
	9	9	9	27	56	110		
3	X	X	10	30			3	2
	9	9	9	27	57	167		
4	10	10	10	30			3	
	9	9	8	26	56	223		
5	X	10	10	30			3	1
	10	10	9	29	59	282	2	
6	X	10	9	29			2	1
	9	9	8	26	55	337		
Total						337	337	16 5


Rio 2016 Olympic Games
Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016

Archer: Pasqualucci David
Country: ITA - Italy

12C


70m-2	1	2	3	Sum	Tot.	Total	10+X	X
1	10	10	10	30			3	
	10	9	9	28	58	58	1	
2	10	10	10	30			3	
	9	9	9	27	57	115		
3	X	X	10	30			3	2
	10	10	9	29	59	174	2	
4	X	X	10	30			3	2
	10	10	9	29	59	233	2	
5	10	10	10	30			3	
	9	9	9	27	57	290		
6	X	X	10	30			3	2
	10	9	9	28	58	348	1	
Total						348	685	24 6

Figura 9 - Planilha de resultados do 3º colocado na classificação Rio 2016



Rio 2016 Olympic Games

Rio de Janeiro (BRA), From 05-08-2016 to 12-08-2016



Archer:

Van Den Berg Sjeff

Country:

NED - Netherlands

8B

Total

0

R M

70m-1	1	2	3	Sum	Tot.		Total	10+X	X
1	X	10	10	30				3	1
	9	9	8	26	56	56	56		
2	10	10	10	30				3	
	9	9	9	27	57	113	113		
3	X	10	10	30				3	1
	10	10	10	30	60	173	173	3	
4	X	X	10	30				3	2
	10	10	10	30	60	233	233	3	
5	10	10	10	30				3	
	10	9	9	28	58	291	291	1	
6	10	10	10	30				3	
	9	9	8	26	56	347	347		
				Total	347	347	347	25	4

Figura 10 - Planilha de resultados do 4º colocado na classificação Rio 2016

Tiro	Kim	Brady	David	Sjef
1	10	10	10	10
2	10	10	9	10
3	10	10	9	10
4	10	10	9	9
5	9	9	9	9
6	9	9	8	8
7	10	10	10	10
8	10	10	10	10
9	10	10	9	10
10	9	9	9	9
11	9	9	9	9
12	9	7	9	9
13	10	10	10	10
14	10	10	10	10
15	10	10	10	10
16	10	10	9	10
17	9	9	9	10
18	9	9	9	10

19	10	10	10	10
20	10	10	10	10
21	10	10	10	10
22	10	9	9	10
23	10	9	9	10
24	9	9	8	10
25	10	10	10	10
26	10	10	10	10
27	10	10	10	10
28	10	9	10	10
29	10	9	10	9
30	10	9	9	9
31	10	10	10	10
32	10	10	10	10
33	9	10	9	10
34	9	9	9	9
35	9	9	9	9
36	9	9	8	8
37	10	10	10	10
38	10	10	10	10
39	10	10	10	9
40	10	10	10	9
41	10	10	9	9
42	9	9	9	8
43	10	10	10	10
44	10	9	10	10
45	10	9	10	9
46	10	9	9	9
47	10	9	9	9
48	10	9	9	8
49	10	10	10	10
50	10	10	10	10
51	10	10	10	9
52	9	9	10	9
53	9	9	10	9
54	9	9	9	8
55	10	10	10	10
56	10	10	10	10
57	10	10	10	10
58	10	10	10	10
59	9	10	10	10
60	9	9	9	9
61	10	10	10	10

62	10	10	10	10
63	10	10	10	10
64	10	10	9	9
65	10	9	9	9
66	9	9	9	8
67	10	10	10	10
68	10	10	10	10
69	10	10	10	10
70	10	10	10	9
71	10	10	9	9
72	9	10	9	9

Uma segunda tabela foi criada para avaliar os resultados das séries, que representa a pontuação a cada 6 tiros:

Serie	Kim	Brady	David	Sjef
1	58	58	54	56
2	57	55	56	57
3	58	58	57	60
4	59	57	56	60
5	60	57	59	58
6	56	57	55	56
7	59	59	58	55
8	60	55	57	55
9	57	57	59	55
10	58	59	59	59
11	59	58	57	56
12	59	60	58	57

Com os dados tabulados, uma primeira análise dos quartis, média, mediana e intervalo interquartil nos possibilita ver a pouca variação nos resultados tiro-a-tiro, demonstrando o quanto é comum a meta de acerto de amarelos (região de pontuação 9, 10 e xis).

Kim	Brady	David	Sjef
Min. : 9.000	Min. : 7.000	Min. : 8.000	Min. : 8.0
1st Qu.: 9.000	1st Qu.: 9.000	1st Qu.: 9.000	1st Qu.: 9.0
Median :10.000	Median :10.000	Median :10.000	Median :10.0
Mean : 9.722	Mean : 9.583	Mean : 9.514	Mean : 9.5
3rd Qu.:10.000	3rd Qu.:10.000	3rd Qu.:10.000	3rd Qu.:10.0
Max. :10.000	Max. :10.000	Max. :10.000	Max. :10.0

A variância foi avaliada conforme matriz a seguir, o que apresentou como mais uma força de que a consistência e pouca variação dos tiros contribui para melhores resultados, aliados com os resultados da análise de quartis, afinal, pode-se encontrar variância baixa também em profissionais com resultados fora da área amarela:

	Kim	Brady	David	Sjef
Kim	0.2034429	0.1502347	0.1447574	0.1408451
Brady	0.1502347	0.3309859	0.1748826	0.1971831
David	0.1447574	0.1748826	0.3378326	0.2042254
Sjef	0.1408451	0.1971831	0.2042254	0.4225352

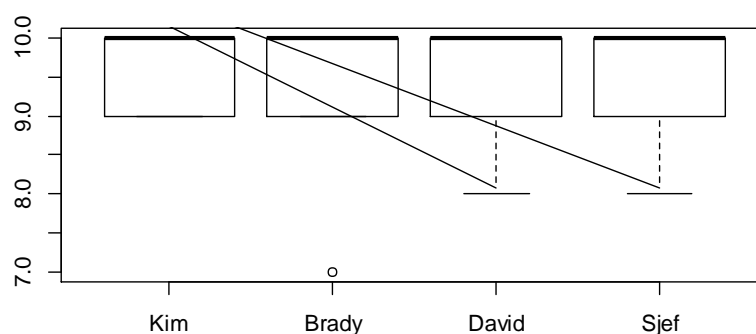


Figura 11 - Box-plot das séries (conjunto de 6 tiros) dos 4 melhores atletas

A mesma análise foi realizada utilizando os resultados série a série conforme a seguir:

Kim	Brady	David	Sjef
Min. :56.00	Min. :55.00	Min. :54.00	Min. :55.00
1st Qu.:57.75	1st Qu.:57.00	1st Qu.:56.00	1st Qu.:55.75
Median :58.50	Median :57.50	Median :57.00	Median :56.50
Mean :58.33	Mean :57.50	Mean :57.08	Mean :57.00
3rd Qu.:59.00	3rd Qu.:58.25	3rd Qu.:58.25	3rd Qu.:58.25
Max. :60.00	Max. :60.00	Max. :59.00	Max. :60.00

	Kim	Brady	David	Sjef
Kim	1.5151515	0.1818182	0.6969697	0.1818182
Brady	0.1818182	2.2727273	0.6818182	0.4545455
David	0.6969697	0.6818182	2.6287879	0.1818182
Sjef	0.1818182	0.4545455	0.1818182	3.4545455

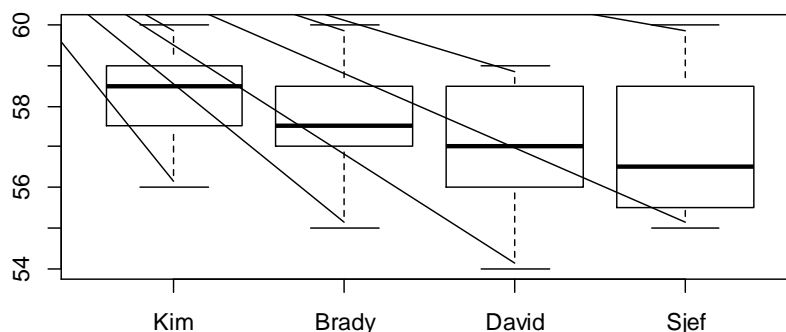


Figura 12 - Box-plot dos tiros dos 4 melhores atletas

Mediante estes resultados é possível ter como hipótese a previsão dos finalistas devido ao fato de que consistência indica os atletas em melhores condições no dia da competição, juntamente com alguns outros fatores, sendo eles:

1. A pontuação total em cada round;
2. O número de flechas atingidas no xis, pois esta é a maior pontuação, e indica o acerto no centro do alvo e é útil para critérios de desempate;
3. O número de flechas atingidas no 10, pois esta é a maior pontuação do alvo;
4. O número de flechas atingidas na zona amarela do alvo, ou seja, o total de flechas com pontuações 9, 10 e xis.

Além destes fatores que indicam o desempenho dos atletas na competição atual, algumas outras métricas podem ser obtidas baseadas nos resultados do ano, as quais serão tratadas mais adiante, na seção 2.3 (Métricas da Federação Internacional).

2.2 Como um atleta obtém o favoritismo

Para um atleta ser identificado como um destaque em uma competição, alguns dados precisam ser organizados com base na sua planilha de resultados. O primeiro dado é a pontuação total, que, afinal, indica a classificação e colocação de cada atleta. O segundo fator é o número de tiros atingidos na zona amarela do alvo, pois, devido ao fato

de a menor pontuação desta zona equivaler a 9 pontos, um atleta que atinge constantemente esta área naturalmente terá 90% de desempenho nos seus tiros. Um terceiro fator a ser considerado é a diferença entre os dois rounds na classificação, o que indica o quanto um atleta ganha ou perde em desempenho durante a classificação. E um quarto e último fator é avaliar seu desempenho anterior à competição, baseado em resultados nas participações de competições prévias e mais recentes, conforme as métricas já existentes pela World Archery, e desta forma, pode-se ter uma melhor noção de quais atletas têm sido mais competitivos recentemente e participam da competição atual com maior visibilidade e condições de vencer. Estas são métrica básicas sem tantos fundamentos estatísticos, mas que contribuem para uma “indicação” de um favorito, porém, conforme verificamos anteriormente, é essencial avaliar os resultados do round classificatório no momento da competição, e minimamente explorar seus dados com análise da variância e quartis para se ter uma visão mais clara do distanciamento dos atletas e quanto seus tiros e resultados estão mais concentrados ou mais dispersos. Essa concentração, neste esporte, é popularmente denominada pelos atletas como agrupamento, ou seja, quanto melhor o agrupamento dos tiros do atleta (flechas mais próximas uma das outras) maior e mais consistente tende a ser seu resultado.

2.3 Métricas da Federação Internacional

A World Archery é o principal órgão do esporte responsável pelo registro dos atletas, pela organização de todas as competições, federações e confederações e por regulamentar e governar cada resultado, ranking e premiações existentes. Além do regulamento e sistema de pontuação, conforme citado na seção 1.2 (Dinâmica da competição nos jogos olímpicos), é responsável também por manter um ranking mundial e algumas estatísticas do desempenho dos atletas nas competições.

O ranking mundial é um outro sistema de pontuação que leva em consideração os resultados dos atletas nos rounds classificatórios e nos rounds eliminatórios, e, de forma acumulada, mantém o desempenho dos atletas. Baseado nesta sistemática, este é um ótimo fator para se utilizar como consistência de um atleta, por ser um acumulado de suas participações e considerar naturalmente os cenários de desempenho em classificações e

combates nas competições, sendo acrescido e decrescido de forma temporal conforme competições realizadas. A seguir, temos duas visões do ranking mundial publicado no site da federação:

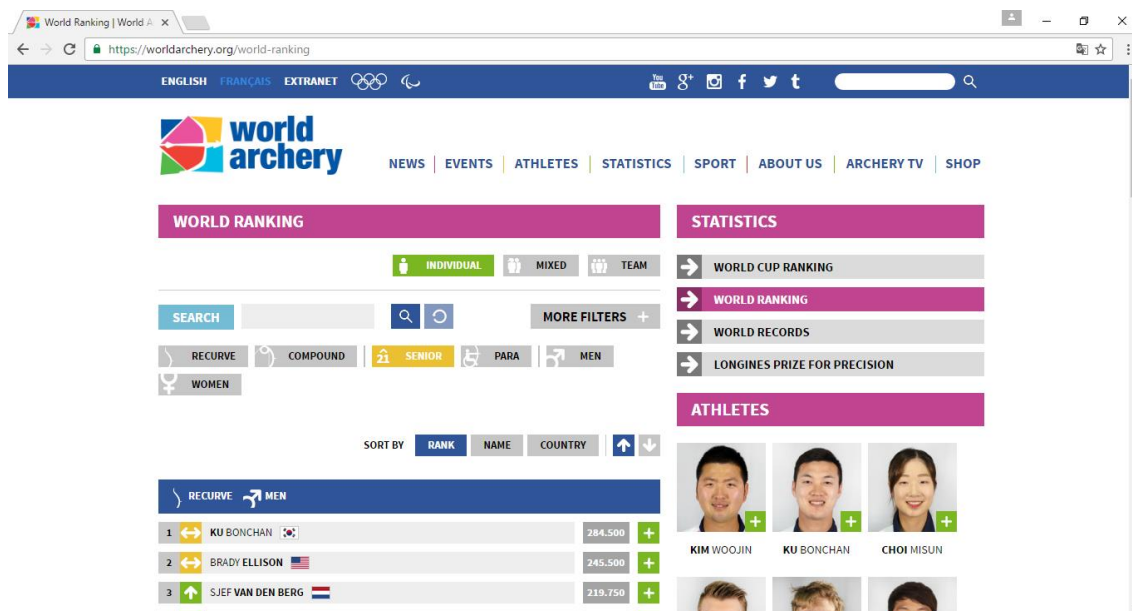


Figura 13 - Consulta ao ranking mundial no site da Federação Internacional

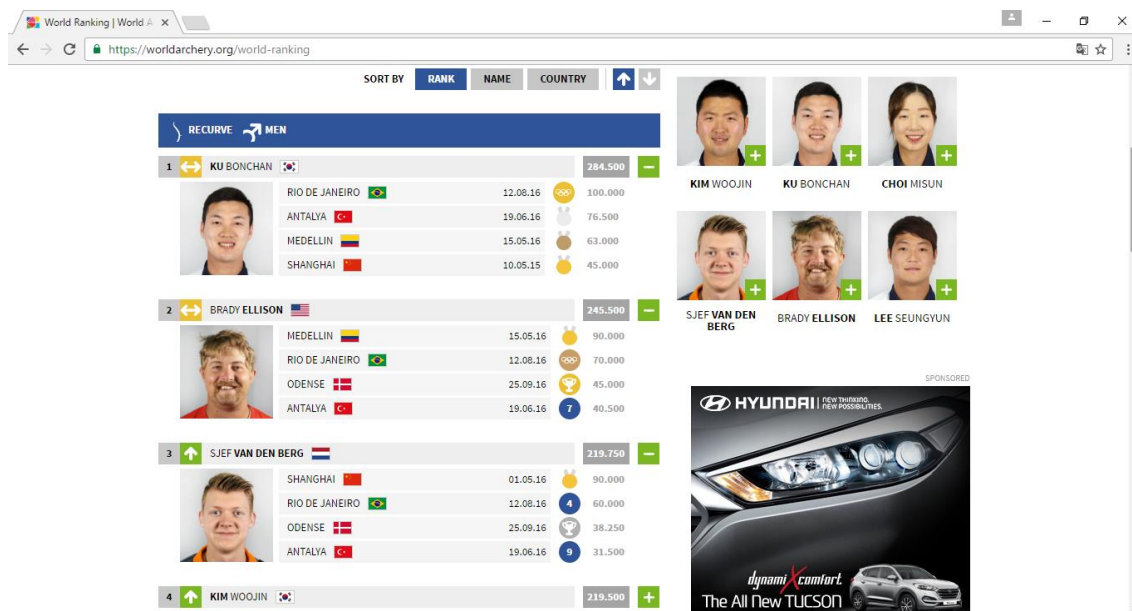


Figura 14 - Consulta aos 3 primeiros colocados do ranking mundial

Ainda dentro das métricas publicadas pela Federação Internacional, encontramos também alguns dados individuais, conforme consulta a seguir:

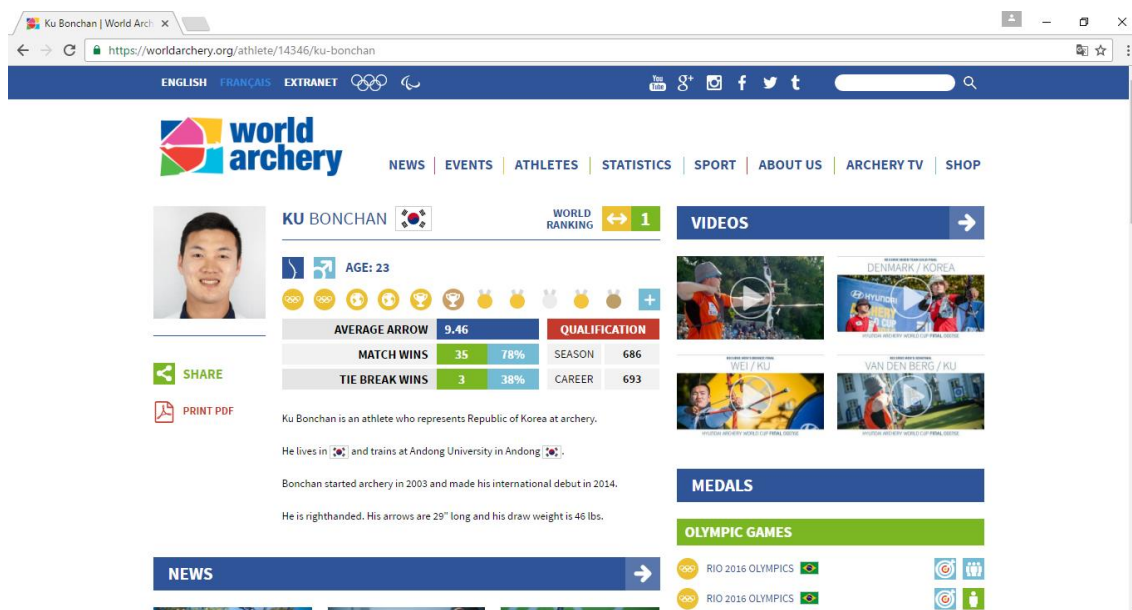


Figura 15 - Métricas públicas de um atleta

Podemos, nesta figura, identificar três métricas importantes:

1. **Average Arrow:** este atleta, por exemplo, possui o valor de 9,46, o que indica sua pontuação média nos tiros;
2. **Match Wins:** este atleta possui um valor de 35, demonstrando a quantidade de combates vencidos, equivalente, neste caso, a 78% dos combates totais que disputou;
3. **Tie Break Wins:** este atleta possui um valor de 3, demonstrando que nas disputas em que empatou e teve que atirar o *shoot-off* (flecha de desempate), ele venceu 38% das vezes.

3. ABORDAGEM METODOLÓGICA

3.1 Processo aplicado e etapas

De acordo com RIES (2012), os produtos que uma *startup* (empresa recém-criada ainda em fase de desenvolvimento) desenvolve são experimentos. À medida que os clientes interagem com os produtos, *feedbacks* e dados são gerados, e tais informações são de extrema importância para o desenvolvimento de novos experimentos ou evolução contínua de um experimento anterior, buscando constantemente atender às necessidades dos clientes e entregar assim produtos mínimos viáveis (MVP, na sigla em inglês) tanto para o consumo como para o crescimento e desenvolvimento sustentável da empresa. O processo sugerido é apresentado conforme a seguir:

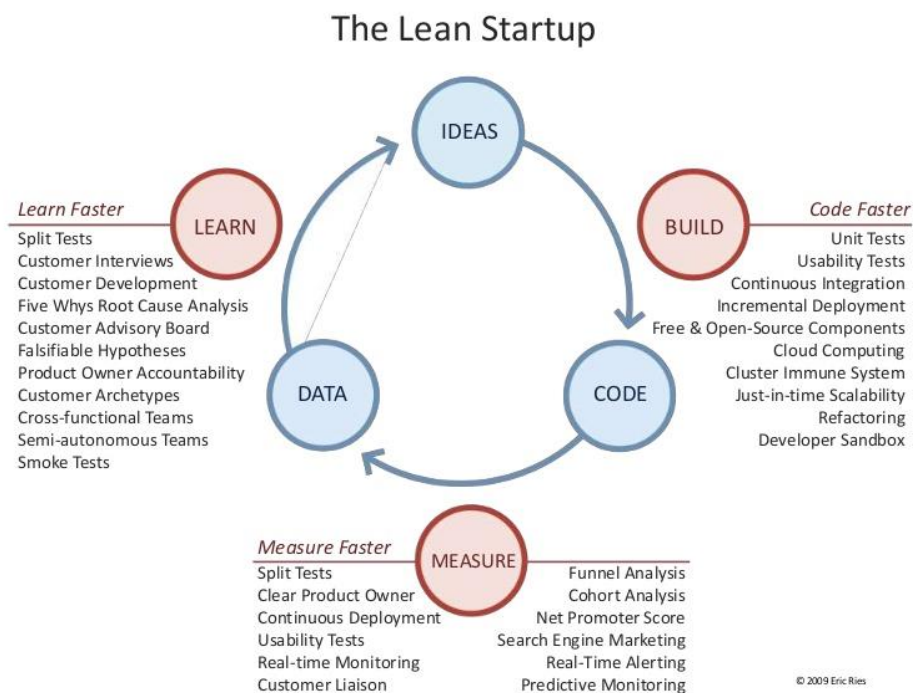


Figura 16 - Ciclo de Feedback Construir-Medir-Aprender

Um produto mínimo viável é algo construído que passou por completo pelo ciclo Construir-Medir-Aprender, indicado por RIES (2012). Tudo parte de uma ou mais ideias que devem ser testadas a partir da construção de algo que possa ser entregue ao seu público-alvo (etapa Construir/*Build*), que tenha alguma métrica a fim de validar seu

mínimo resultado (etapa Medir/*Measure*) com objetivo de gerar algum tipo de dado ou conjunto de dados que a partir do qual possa ser possível ter algum feedback (etapa Aprender/*Learn*), e os resultados deste aprendizado possam apresentar novas ideias ou, minimamente, evoluções das ideias aplicadas no ciclo anterior, e, consequentemente, permitir que o próximo ciclo mantenha um foco naquilo que for realmente necessário para quem a empresa está entregando o seu produto.

Como processo produtivo, RIES (2012) também indica um processo baseado em gerenciamento ágil de projetos, com objetivo de garantir uma mesma cultura de idealização, construção e testes de produtos em seu processo de produção ou de desenvolvimento. Dentre os apresentados, o Scrum³ idealizado por SCHWABER e SUTHERLAND (1995) fez total sentido a este projeto pela sua natureza de executar ciclos pequenos de construção ou prototipação, exigindo um planejamento focado em uma meta que deseja-se atingir, com organização macro, micro, estimativas de esforço, acompanhamento e controle de curto prazo, revisão e feedback, sem exigências de documentações de projetos burocráticas. E mais: dado seu modelo iterativo e incremental, possibilita a continuidade da cultura do ciclo Contruir-Medir-Aprender de RIES (2012), garantindo integração entre ciclos para assegurar o foco e objetivo a serem atingidos, sempre em evolução contínua e refinamento constante a cada etapa executada.

Com base nestes dois modelos, a decisão para este projeto foi combiná-los considerando como produto mínimo viável: um modelo estatístico que possa prever possíveis finalistas de uma competição de tiro com arco. Para a produção deste produto, um planejamento macro fez-se necessário objetivando: definição do tamanho de cada ciclo; metas; e quantidades de ciclos possível para construção de um produto mínimo viável. Deste planejamento macro obteve-se as seguintes metas:

1. Entender e documentar a dinâmica de resultados de competições;
2. Organizar dados e métricas úteis;
3. Explorar e definir variáveis para uso em um modelo;
4. Aplicar o método de cluster para segmentar atletas;

³ Scrum (subs): Um *framework* dentro do qual pessoas podem tratar e resolver problemas complexos e adaptativos, enquanto produtiva e criativamente entregam produtos com o mais alto valor possível. SCHWABER e SUTHERLAND (1995).

5. Testar métodos de previsões baseado em algoritmos de classificação e modelo de regressão logística como modelo de *scoring* (pontuação) para previsão de finalistas.

Cada uma destas metas foi transformada em ciclos de uma semana, conhecidos dentro do Scrum como *Sprints*, a serem gerenciados e executados individualmente, na ordem especificada, conforme processo indicado por SCHWABER e SUTHERLAND (1995), que exige:

SCRUM FRAMEWORK

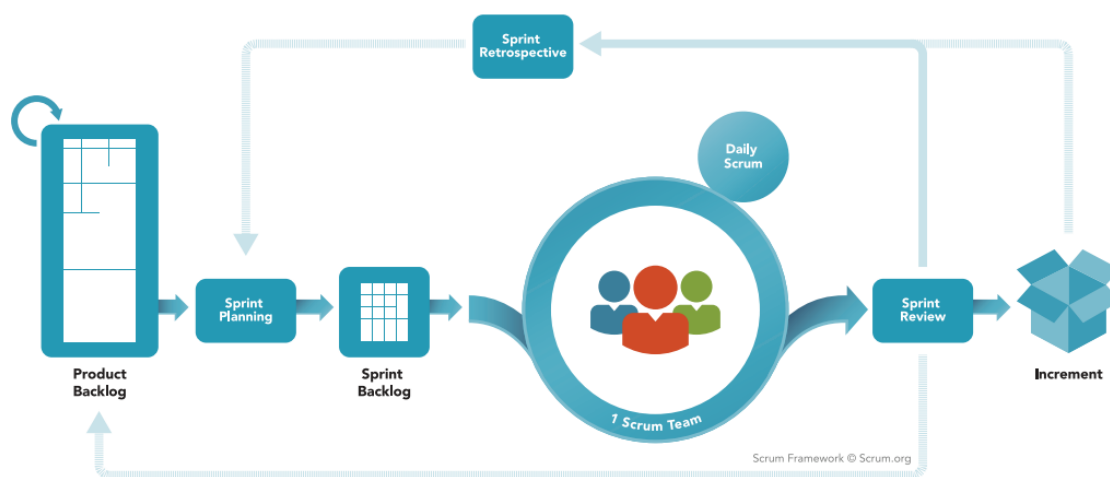


Figura 17 - Scrum Framework: Fonte scrum.org.

- a. Definição de uma lista de itens entregáveis e priorizados, conhecidos como *Product Backlog*, com uma visão macro do que se pretende ter ao final, sem detalhes de atividades ou tarefas de execução;
- b. Realização da *Sprint Planning*, que é um planejamento do ciclo, com objetivo de definir as tarefas a serem realizadas com suas estimativas de esforços para realização, a partir dos itens do *product backlog*, mas agora planejando de uma visão mais micro de execução. Desta reunião conclui-

- se o *Sprint Backlog*, que é a lista de tarefas a serem executadas, devidamente analisadas, criticadas, identificadas riscos e estimadas;
- c. Reunião diária (*Daily Scrum*) para acompanhamento das tarefas, permitindo a inspeção diária e adaptação da execução das tarefas planejadas conforme seus riscos e necessidades de atuação;
 - d. Reunião de revisão (*Sprint Review*), realizada ao final do ciclo como um processo de qualidade para garantir que as entregas previstas de cada tarefa foram realizadas conforme o planejado;
 - e. Reunião de retrospectiva (*Sprint Retrospective*), realizada ao final do ciclo para obter *feedback* da execução do ciclo, avaliando o que deu certo, o que deu de errado e, principalmente, discutir o que pode e deve ser diferente para o próximo ciclo.

Os resultados ou entregáveis de cada ciclo são apresentados conforme demais conteúdos adiante.

3.2 Organização de dados e métricas

A principal dificuldade no processo de organização dos dados e métricas para este caso foi a tabulação. Apesar dos órgãos competentes terem alguns dados publicados em seus sites, como os casos demonstrados da Federação Internacional no que diz respeito às métricas dos atletas e das competições realizadas, não há dados tabulados em formatos padrões, como arquivos de textos ou algum modelo de banco de dados ou planilha eletrônica. O primeiro passo para ter dados organizados foi transformar as planilhas oficiais de resultados em tabelas que minimamente possibilitassem seu uso em softwares para uma análise e exploração dos dados.

Como primeira tabulação, foram organizados, em arquivo Microsoft Excel®, os resultados dos quatro melhores atletas conforme resultados nas Olimpíadas Rio 2016, e, a partir destes, foi realizada uma análise exploratória conforme descrito no item 2.1 (É possível prever finalistas de uma competição?). Esta etapa foi necessária não apenas para conhecer os dados e a dinâmica dos resultados, mas também diagnosticar como pode ser avaliada a consistência dos atletas de mais alto desempenho. Desta tabulação, foi criada

uma planilha de resultados tiro-a-tiro dos quatro atletas, sendo 72 linhas para cada um, e uma segunda planilha com os resultados série-a-série, sendo 12 linhas para cada.

Como segunda tabulação, foram organizados os dados do ranking do Round Classificatório Rio 2016, realizado em 05 de agosto de 2016, e inclusas as variáveis identificadas como possivelmente úteis para análises de desempenho dos atletas, sendo elas:

- **Rank:** posição do atleta no round classificatório;
- **Atleta:** nome do atleta;
- **País:** país ao qual o atleta representa;
- **Round1:** pontuação obtida no primeiro round (1° ao 36° tiro);
- **Round2:** pontuação obtida no segundo round (37° ao 72° tiro);
- **RoundTotal:** pontuação total da classificação na competição (somatório dos rounds 1 e 2);
- **Dez:** número de tiros atingidos na pontuação máxima;
- **X:** número de tiros atingidos na pontuação máxima e critério de desempate;
- **Nove:** número de tiros atingidos na pontuação nove;
- **WorldRanking:** pontuação atual do atleta no ranking mundial no dia da competição olímpica;
- **Prop.Dez:** proporção de tiros atingidos no dez em relação ao total de disparos;
- **Prop.X:** proporção de tiros atingidos no xis em relação ao total de disparos;
- **Prop.Nove:** proporção de tiros atingidos no nove em relação ao total de disparos;
- **Amarelos:** número de tiros atingidos na região central do alvo, que indica a soma dos tiros atingidos na pontuação dez e nove;
- **Finalista:** variável dicotômica indicando os atletas que chegaram pelo menos uma vez a uma final no ano;
- **PontosCombates:** número de pontos obtidos em combates dentro do ano.

Estes dados foram computados manualmente em Microsoft Excel® e transformados em arquivos CSV separados por ponto e vírgula como separador de colunas, para uso nos softwares R Studio 1.0 e Tableau 9.1®.

3.3 Análise e seleção de variáveis

Conforme tabulação dos dados indicados no item 3.2 (Organização de dados e métricas), cada variável foi analisada para visualização de seu comportamento, e consequentemente foram verificados também os resultados e desempenho dos atletas na competição no Rio 2016. As seguintes análises foram as identificadas como mais relevantes para garantir uma melhor seleção de variáveis:

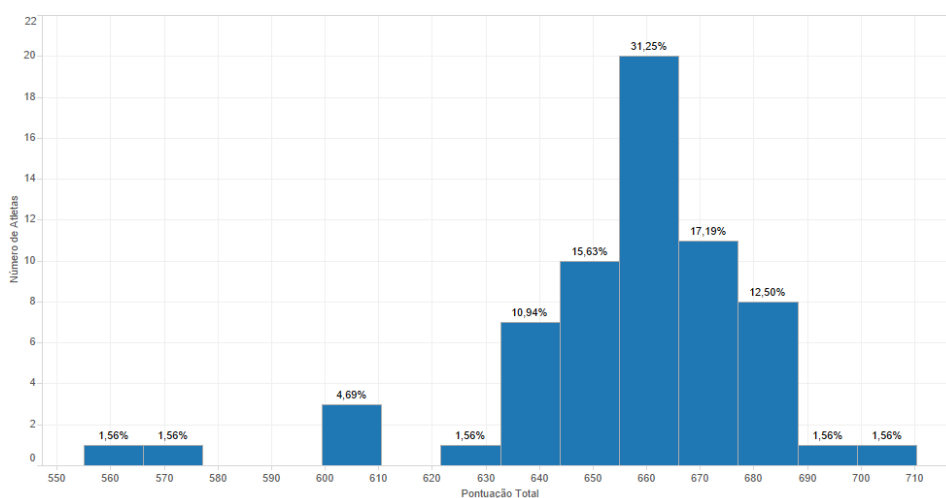


Figura 18 - Histograma da Pontuação Total dos atletas

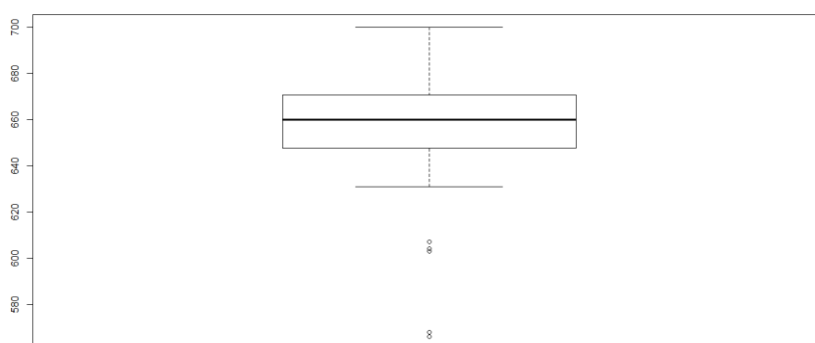


Figura 19 - Box-plot da Pontuação Total dos atletas

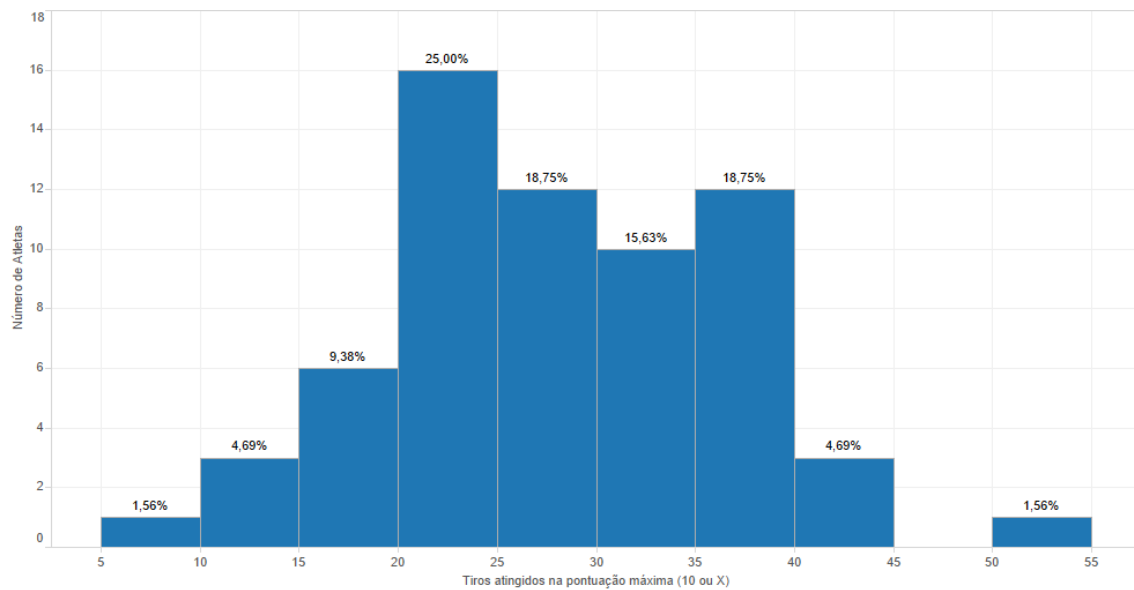


Figura 20 - Histograma dos tiros atingidos na pontuação 10

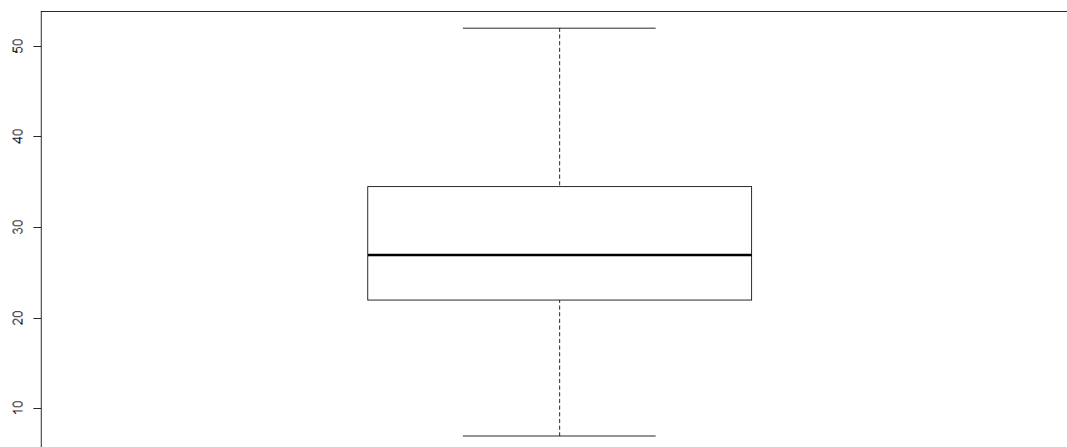


Figura 21 - Box-plot dos tiros atingidos na pontuação 10

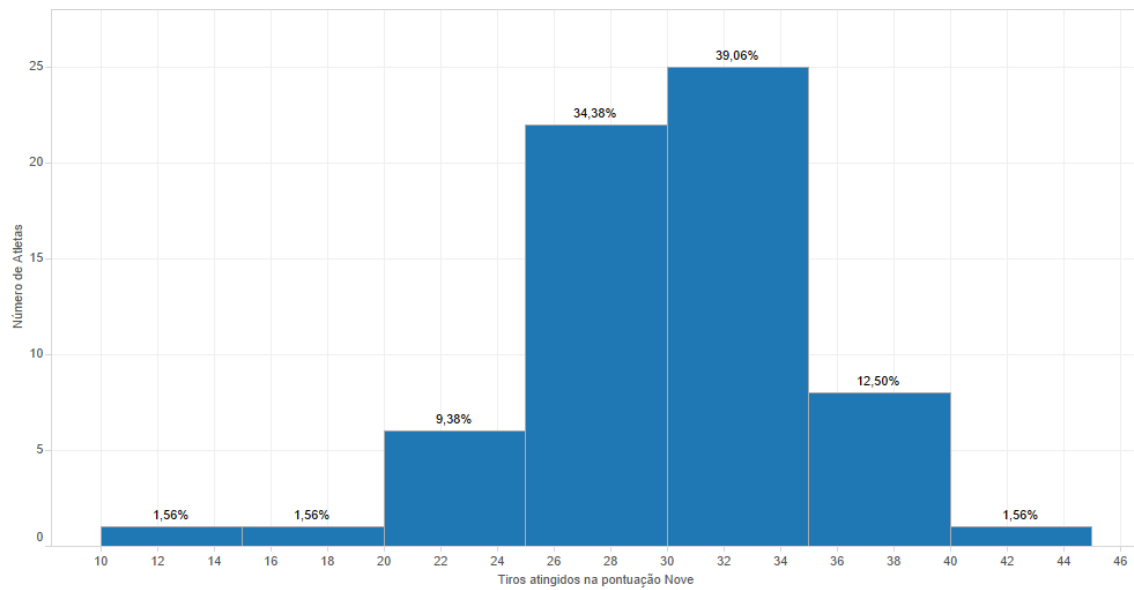


Figura 22 - Histograma dos tiros atingidos na pontuação 9

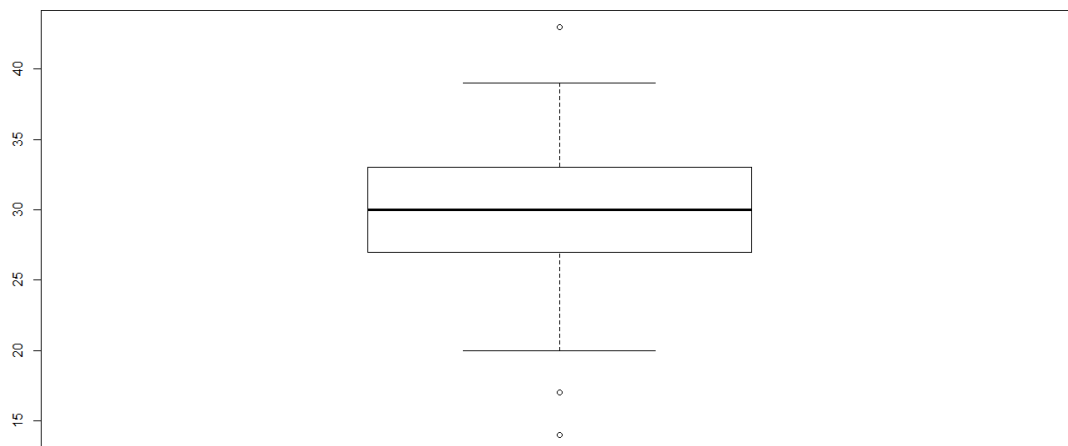


Figura 23 - Box-plot dos tiros atingidos na pontuação 9

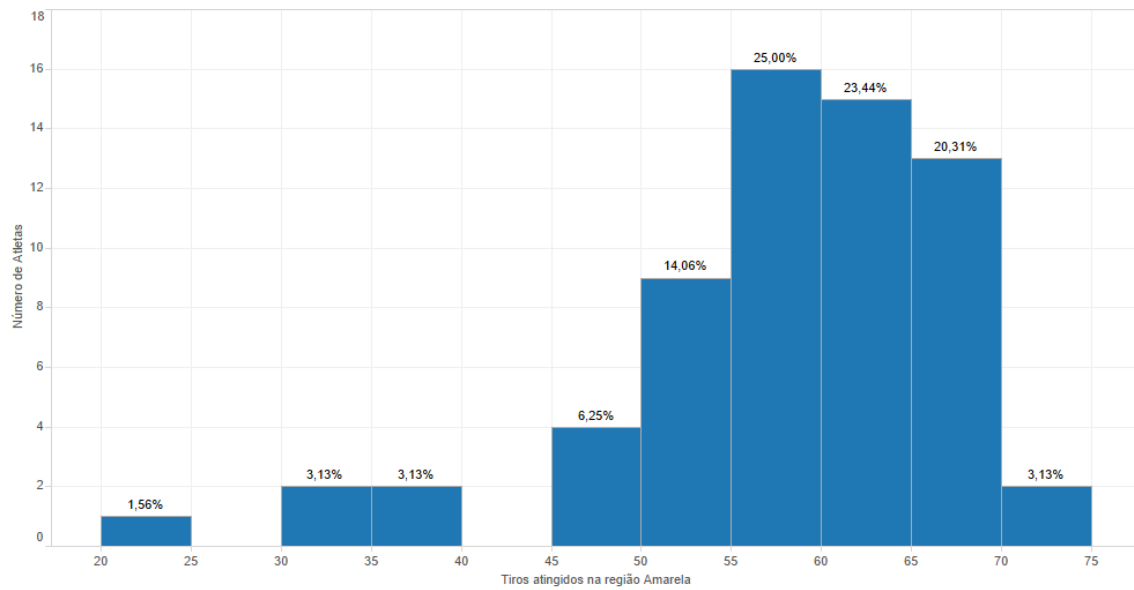


Figura 24 - Histograma dos tiros atingidos na região amarela do alvo

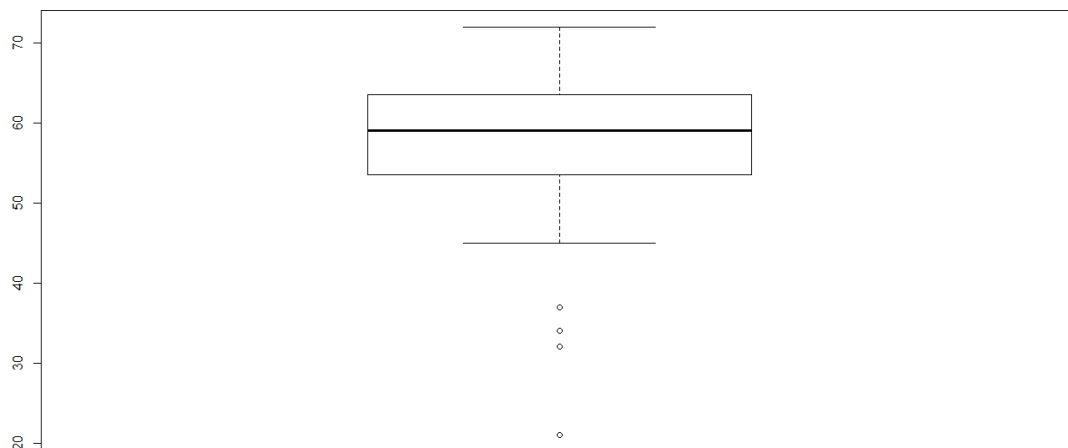


Figura 25 - Box-plot dos tiros atingidos na região amarela do alvo

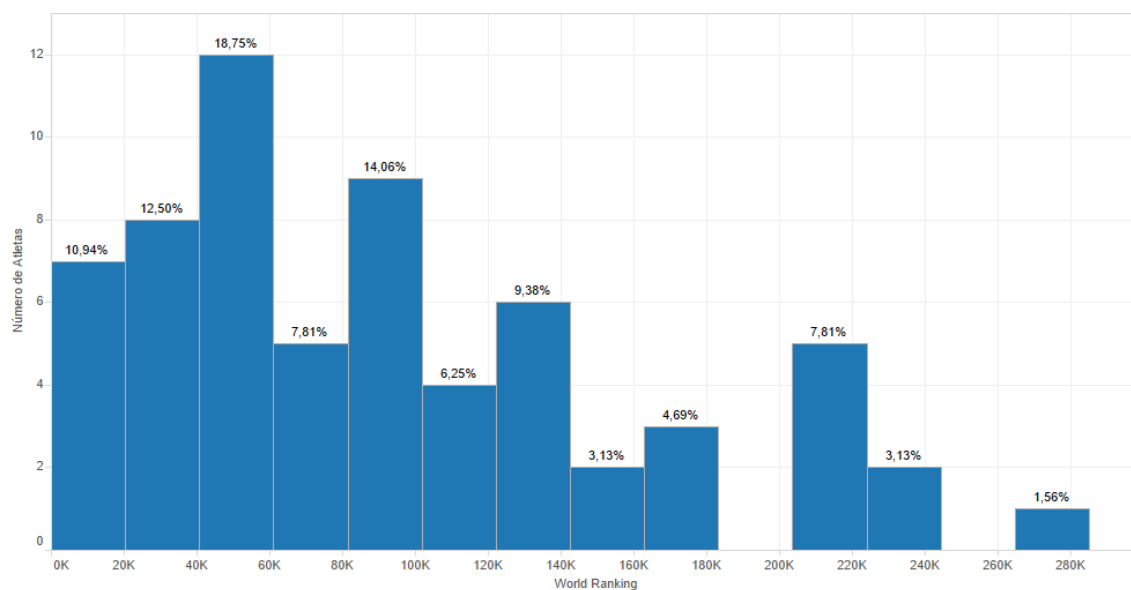


Figura 26 - Histograma do posicionamento no Ranking Mundial

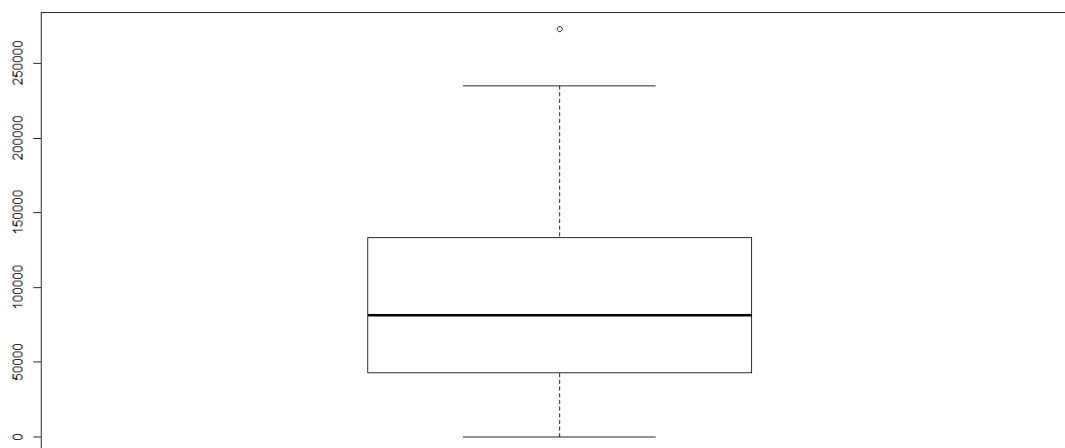


Figura 27 - Box-plot do posicionamento no Ranking Mundial

Destas análises podemos observar as seguintes informações:

1. Na pontuação total detecta-se um grupo de atletas de maior competitividade na região dos 650 até 700 pontos, que é onde no ranking a diferença entre os atletas começa a ser menor;
2. Avaliando o número de tiros na pontuação 10, detectamos que temos cerca de 6% de atletas que atingem acima de 40 vezes esta região, e, desta forma,

se destacam dos demais na competição. Esta variável tende a distanciar os atletas de maior desempenho;

3. Avaliando o número de tiros na pontuação 9, detectamos que temos cerca de 14% de atletas que atingem acima de 35 vezes esta região e 70% entre 25 e 35 vezes, o que demonstra o quanto a região de pontuação 9 é bem explorada pela grande parte dos participantes;
4. Avaliando o número de tiros na região amarela do alvo, ou seja, considerando as pontuações 9 e 10, temos 3,13% de atletas que se destacam em relação aos demais, mas vemos que esta variável aproxima boa parte dos atletas por refletir no gráfico pouco mais de 80% de atletas que atingem esta região de 50 à 70 vezes. Esta variável é bastante interessante, pois a zona amarela é de forte competitividade nos rounds eliminatórios e pode agrupar de forma mais refinada os competidores de desempenho semelhante em momentos de eliminação;
5. Já a variável que considera o Ranking Mundial distingue fortemente alguns atletas, cerca de 12% deles, pois é uma métrica acumulativa do ano e os atletas que participam mais de uma vez das semifinais ou finais, têm sua pontuação mais distanciada da maioria, o que pode ser um problema em algoritmos de predição por ganhar muito peso em relação às demais.

Após estas análises, identifica-se a necessidade de cruzar estas variáveis a ponto de observar o posicionamento dos atletas e seu desempenho de forma a não considerar simplesmente o ranking da competição atual ou o ranking mundial, e, para uso em algoritmos de previsão, aparenta-se necessária uma padronização para evitar que algumas variáveis tenham peso muito superior às demais nos modelos.



The scatter plot displays the relationship between an athlete's Total Score (X-axis) and their World Ranking (Y-axis). The X-axis ranges from 560 to 700, and the Y-axis ranges from 0K to 250K. A horizontal dotted line is drawn at the 0K ranking level. Most athletes are clustered between 560 and 700 on the X-axis and 0K to 250K on the Y-axis. The data points are labeled with the names of the athletes, and some are highlighted in blue.

Athlete	Pontuação Total (X)	World Ranking (Y)
SUTHERLAND Gavin	568	30K
EL GHIRANI Ali	570	0K
DAVID Amos	590	0K
JENSEN Arne	605	0K
BALAZ Boris	630	25K
ELDER Robert	635	35K
KOUASSI Rene Philippe	645	5K
WUJAYA Muhammad	650	20K
PURNAMA Hendra	650	45K
DUZELBAYEV Sultan	655	15K
GALLAZZO Marco	655	0K
PIIPPO Samuli	635	55K
DIELEMANS Mitch	640	70K
MALAVE Elias	645	85K
PLIHON Pierre	650	80K
PRILOPOV Anton	640	125K
FERNANDEZ Antonio	645	115K
EGA AGATHA Riau	650	105K
D'ALMEIDA Marcus	655	135K
DANIEL Lucas	660	110K
RUBAN Viktor	665	95K
MOHAMAD Khairul Anuar	670	85K
FLOTO Florian	675	55K
SOTO Ricardo	680	45K
RODRIGUEZ Juan I.	685	95K
PASQUALUCCI David	690	100K
FURUKAWA Takaharu	685	125K
WEI Chun-Heng	675	165K
ERNESTO	680	155K
BOARDMAN	665	145K
XING Yu	655	150K
ALVARINO GARCIA Miguel	650	165K
GARRETT Zach	660	210K
LEE Seungyun	675	205K
ELLISON Brady	685	215K
VAN DEN BERG Sjef	690	225K
KU Bonchan	685	235K
KIM Woojin	700	265K

Figura 30 - Segundo grupo de alto desempenho

Um terceiro grupo foi identificado por não ter pontuação expressiva na competição e também por não estar bem qualificado no ranking mundial, apesar de minimamente apresentar alguma pontuação neste ranking. Consideramos estes atletas como pertencentes a um terceiro grupo de desempenho.

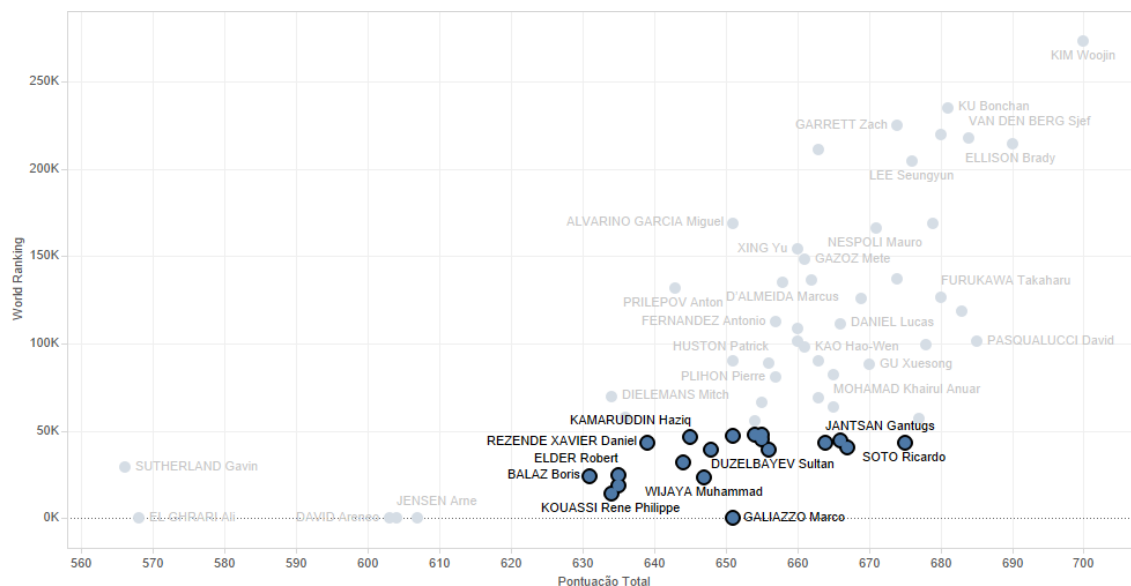


Figura 31 - Terceiro grupo de desempenho

E por último identificamos um quarto grupo, que teve pontuação bem abaixo do rendimento comum, tanto na competição quanto no ranking mundial.

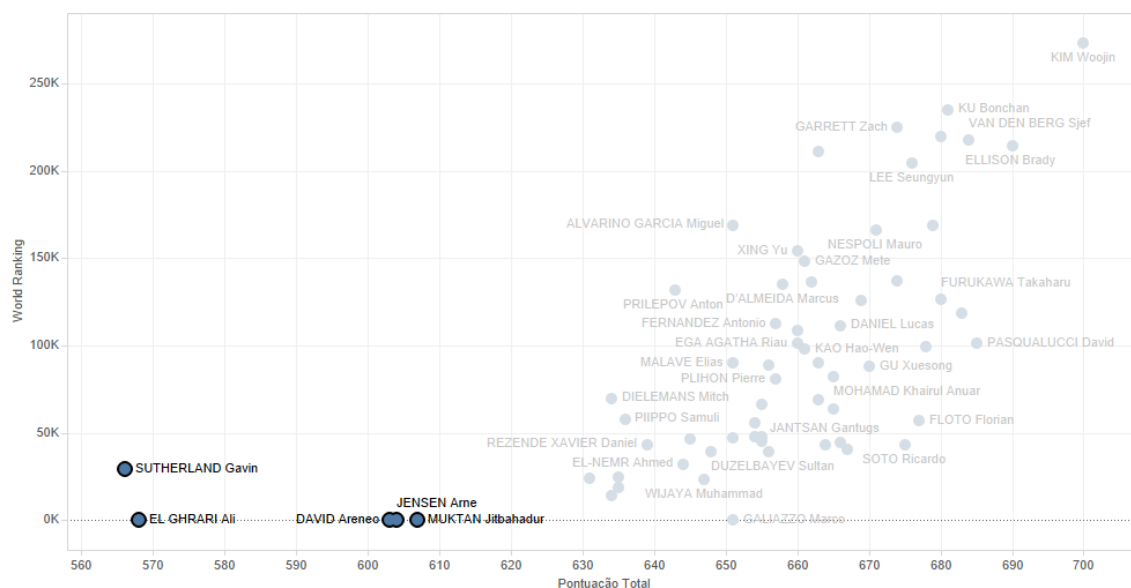


Figura 32 - Quarto grupo de desempenho

4.2 Agrupamentos baseado na pontuação máxima

Uma outra forma de agrupar os atletas foi cruzando a quantidade de tiros na pontuação máxima possível contra a pontuação total da competição e pontuação no ranking mundial. Como a zona máxima de pontuação (pontuações 10 e xis) privilegia os atletas de melhor desempenho, a análise consequentemente distancia mais os atletas e assim consegue demonstrar um pouco melhor a disputa dos melhores colocados na competição.

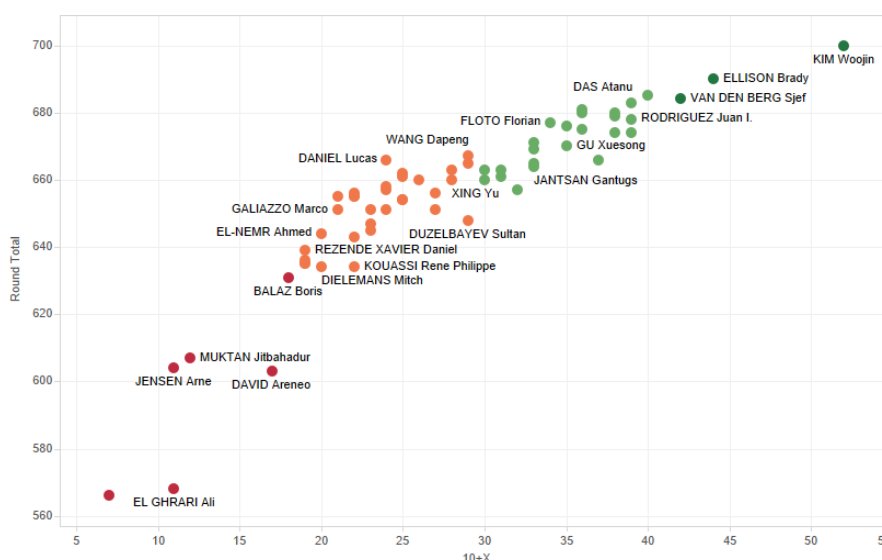


Figura 33 - Agrupamento baseado em tiros de Pontuação Máxima vs. Pontuação Total

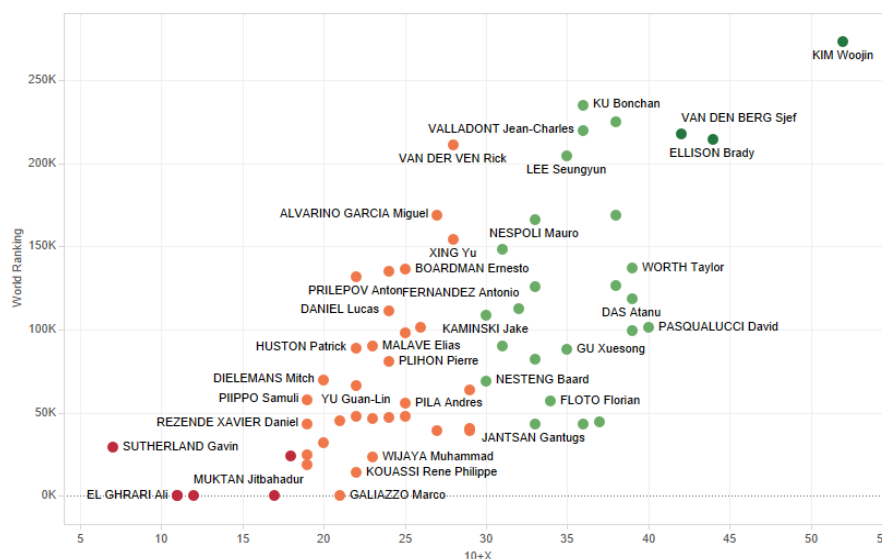


Figura 34 - Agrupamento baseado em tiros de Pontuação Máxima vs. Ranking Mundial

4.3 Agrupamentos baseado na zona amarela

O agrupamento baseado na zona amarela foi realizado pois, diferentemente do agrupamento baseado na pontuação máxima que distancia pelo maior desempenho, este aproxima mais os atletas. Isso se dá pelo fato de que o desempenho dos atletas normalmente se mostra diferente no round de classificação e nos rounds eliminatórios. Por mais que nos rounds classificatórios os atletas se distanciem pela pontuação, eles continuam tendo um alto número de tiros na região amarela, porém, perdem pontos por atingirem menos a região de pontuações dez e xis.

Nos rounds eliminatórios, a zona amarela do alvo, independentemente de a pontuação ser 9 ou 10, é uma zona de competitividade alta pois os atletas não atiram 6 tiros como na etapa de classificação, mas apenas 3, e com isso tendem a realizar tiros melhores. Por este motivo, este agrupamento e aproximação dos atletas é essencial para visualizar grupos de resultados semelhantes e de grande competitividade.

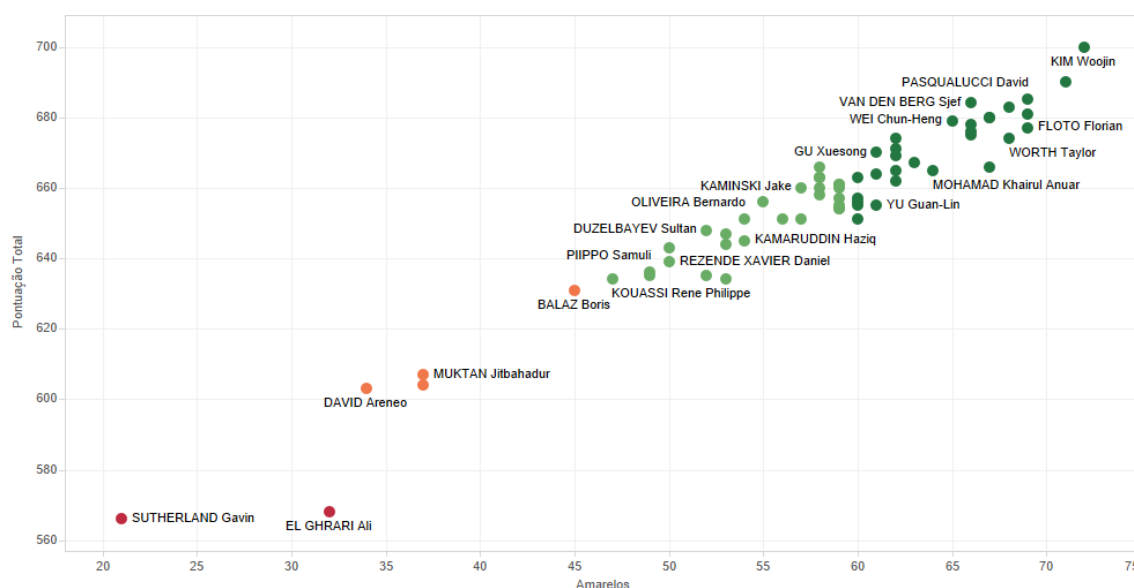


Figura 35 - Agrupamento baseado em tiros na Zona Amarela vs. Pontuação Total

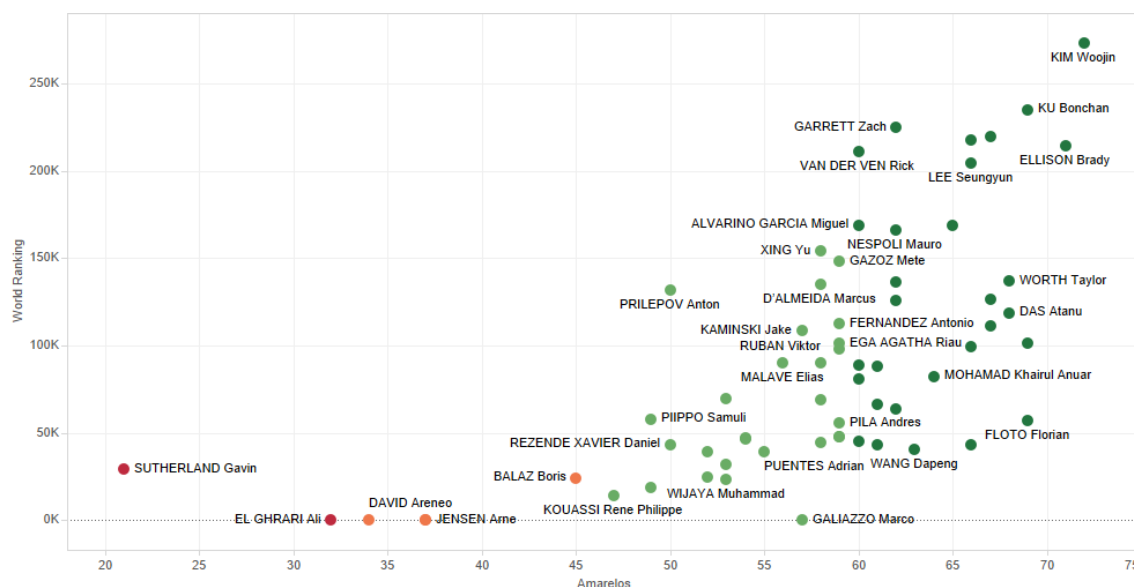


Figura 36 - Agrupamento baseado em tiros na Zona Amarela vs. Ranking Mundial

4.4 Agrupamentos baseado em análise de cluster

Apesar dos agrupamentos apresentados anteriormente terem boas interpretações, eles acabam sendo direcionados conforme cada variável, de uma forma individual e direcionada baseada no olhar de quem o interpretou sem uso de um algoritmo ou técnica de agrupamento científica e mais eficaz.

Visualizamos o caso da pontuação total da competição *versus* ranking mundial, que privilegia os atletas de maior desempenho na competição atual e participação contínua nas competições anteriores, e distancia demais atletas e grupos, não considerando sua real competitividade do momento.

Visualizamos o caso de número de tiros de pontuação máxima combinados com a pontuação total e ranking mundial, que aproxima mais os atletas mais ainda assim os distancia, visto que ainda há casos de desempenho muito superior aos demais.

E visualizamos também o cruzamento das variáveis de número de tiros na região amarela do alvo conjuntamente com a pontuação total e ranking mundial, que aproxima mais os atletas em relação às demais comparações, trazendo um cenário um pouco mais

coerente com a realidade, tendo em vista que nas etapas de eliminatórias todo tiro na região amarela ser um tiro de alta competitividade.

Dados estes fatores, uma análise de cluster torna-se necessária para poder agrupar os atletas com todas as variáveis existentes e que consideradas relevantes. Para isso, as seguintes variáveis foram consideradas:

1. Pontuação no Round Total;
2. Número de tiros atingidos na pontuação máxima (pontuação 10);
3. Pontuação no Ranking Mundial;
4. Pontuações em Combates no Ranking Mundial;
5. Número de tiros atingidos na região amarela (pontuação 9 e 10).

Para garantir que todas as variáveis sejam consideradas com um mesmo peso, uma padronização foi realizada utilizando *Z-Score*.

Carregando estes dados no *software* R Studio, conforme detalhado no item 3.2 (Organização de dados e métricas), foi aplicada a análise de cluster buscando obter quatro e cinco grupos usando o algoritmo *k-means* conforme a seguir:

```
#####  
# Bibliotecas utilizadas  
#####  
library(gmodels);  
  
#####  
# Carga dos dados  
#####  
Dados = read.csv("RoundOlimpico_Rio2016_Individual.csv", sep=";", dec=",")  
view(Dados);  
  
# Número de Amarelos  
Dados$Amarelos = Dados$Dez + Dados$Nove;  
  
#####  
# Padronização de Variáveis  
#####  
Dados$Round1_S = scale(Dados$Round1);  
Dados$Round2_S = scale(Dados$Round2);  
Dados$RoundTotal_S = scale(Dados$RoundTotal);  
Dados$Dez_S = scale(Dados$Dez);  
Dados$X_S = scale(Dados$X);  
Dados$Nove_S = scale(Dados$Nove);
```

```
Dados$worldRanking_S = scale(Dados$worldRanking);
Dados$PontosCombates_S = scale(Dados$PontosCombates);
Dados$Amarelos_S = scale(Dados$Amarelos);
Dados$Classe = "Não Finalista";
Dados$Classe[which(Dados$Finalista==1)] = "Finalista";
Dados$Classe = factor(Dados$Classe);

#####
# Análise de Cluster
#####
Cluster.Dados = Dados[, c(19,20,23,24,25)];

# Análise com 4 Grupos
Cluster.Modelo = kmeans(Cluster.Dados,4);
Dados$Cluster.4 = Cluster.Modelo$cluster;
CrosstTable(Dados$Cluster.4, Dados$Classe, chisq=FALSE);

aggregate(cbind(Dados$RoundTotal,          Dados$Dez,          Dados$worldRanking,
Dados$PontosCombates, Dados$Amarelos), list(Dados$Cluster.4), mean);

# Análise com 5 Grupos
Cluster.Modelo = kmeans(Cluster.Dados,5);
Dados$Cluster.5 = Cluster.Modelo$cluster;
CrosstTable(Dados$Cluster.5, Dados$Classe, chisq=FALSE);

aggregate(cbind(Dados$RoundTotal,          Dados$Dez,          Dados$worldRanking,
Dados$PontosCombates, Dados$Amarelos), list(Dados$Cluster.5), mean);
```

Após execução do algoritmo, tivemos como resultado os grupos desejados e executamos uma tabela cruzada para indicar em cada grupo quantos atletas foram finalistas em alguma etapa do ano anterior à competição atual. Como análise, agrupando os atletas em quatro grupos de desempenho, tivemos:

Group	V1	V2	V3	V4	V5
1	589.6000	11.60000	5800.00	0.000000	32.20000
2	645.8636	22.04545	47880.68	7.636364	53.95455
3	664.5238	30.19048	107991.67	0.000000	61.00000
4	677.5000	37.43750	159385.81	157.937500	66.06250

Cell Contents

```
|-----|
|                                     N |
| Chi-square contribution |
|      N / Row Total |
|      N / Col Total |
|      N / Table Total |
|-----|
```

Total Observations in Table: 64

Dados\$Cluster.4	Dados\$Classe		
	Finalista	Não Finalista	Row Total
1	0	5	5
	0.859	0.178	
	0.000	1.000	0.078
	0.000	0.094	
	0.000	0.078	
2	0	22	22
	3.781	0.785	
	0.000	1.000	0.344
	0.000	0.415	
	0.000	0.344	
3	2	19	21
	0.718	0.149	
	0.095	0.905	0.328
	0.182	0.358	
	0.031	0.297	
4	9	7	16
	14.205	2.948	
	0.562	0.438	0.250
	0.818	0.132	
	0.141	0.109	
Column Total	11	53	64
	0.172	0.828	

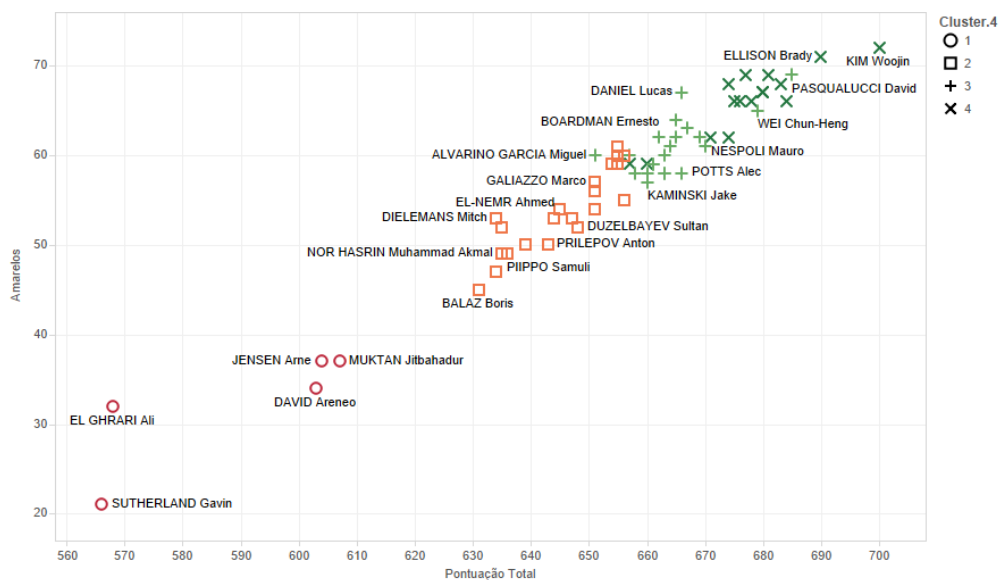


Figura 37 - Distribuição da Análise de 4 Clusters de Desempenho

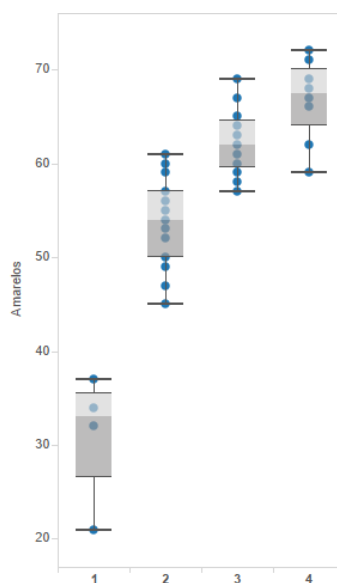


Figura 38 - Desempenho dos 4 Clusters em Zona Amarela

E, como análise agrupando os atletas em cinco grupos de desempenho, tivemos:

Group	V1	V2	V3	V4	V5
1	645.381	22.04762	45922.62	8.0000	53.66667
2	662.350	28.90000	104341.25	0.0000	60.35000
3	589.600	11.60000	5800.00	0.0000	32.20000
4	688.000	43.33333	181083.33	0.0000	68.66667
5	676.000	36.46667	151794.87	168.4667	65.66667

Cell Contents

N
Chi-square contribution
N / Row Total
N / Col Total
N / Table Total

Total Observations in Table: 64

	Dados\$Classe		
Dados\$Cluster.5	Finalista	Não Finalista	Row Total
1	0	21	21
	3.609	0.749	
	0.000	1.000	0.328
	0.000	0.396	
	0.000	0.328	

2	1	19	20
	1.728	0.359	
	0.050	0.950	0.312
	0.091	0.358	
	0.016	0.297	
3	0	5	5
	0.859	0.178	
	0.000	1.000	0.078
	0.000	0.094	
	0.000	0.078	
4	2	1	3
	4.273	0.887	
	0.667	0.333	0.047
	0.182	0.019	
	0.031	0.016	
5	8	7	15
	11.402	2.367	
	0.533	0.467	0.234
	0.727	0.132	
	0.125	0.109	
Column Total	11	53	64
	0.172	0.828	

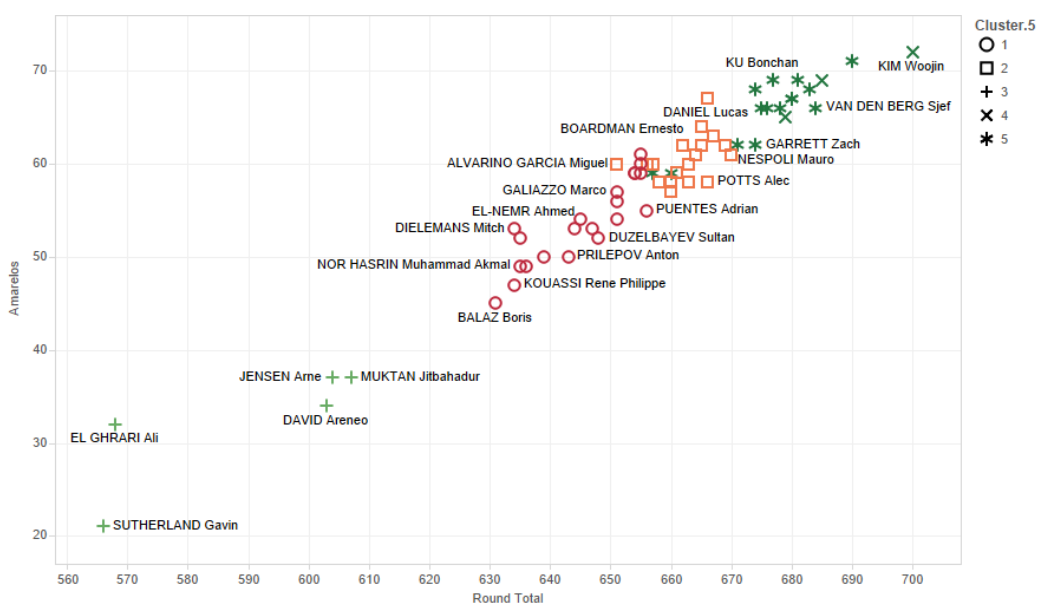


Figura 39 - Distribuição da Análise de 5 Clusters de Desempenho

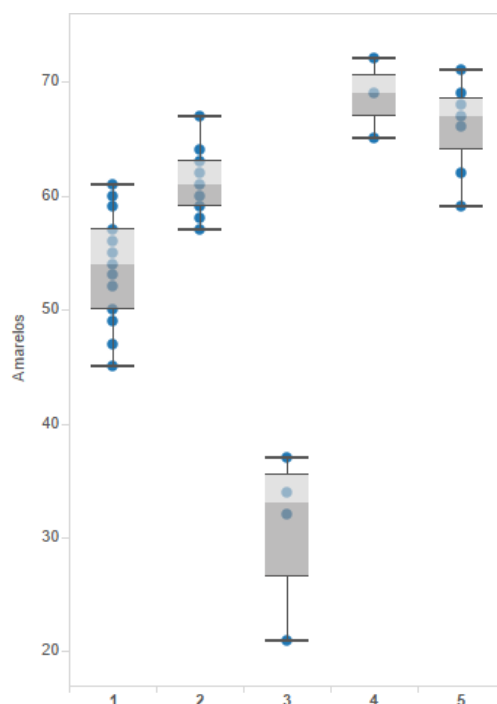


Figura 40 - Desempenho dos 5 Clusters em Zona Amarela

4.5 Previsão usando algoritmo de classificação

Com o objetivo de tentar prever quais atletas têm maiores condições de ser finalistas da competição, uma primeira opção foi a utilização do algoritmo K-Nearest Neighbors (KNN). Devido à sua simplicidade na classificação de características por não produzir um modelo, e sua previsão ser baseada em um treinamento anterior dos dados, consideramos ser uma tentativa interessante, usando as variáveis deste cenário de competição, que, naturalmente, refletem características de desempenho dos atletas e fazer com que a previsão seja realizada a partir da comparação da distância entre os competidores.

Como fator de classificação, uma variável categórica foi criada apenas para indicar quem já foi finalista em alguma competição do ano. Para esta, indicamos os termos **Finalista** e **Não Finalista**, e assim passamos a utilizar esta indicação como fator de classificação para o algoritmo e para análise de seus resultados. A aplicação do algoritmo foi realizada conforme a seguir:

```
#####  
# Previsão de Finalistas com KNN  
#####  
library(class);  
  
knn.Base = Dados[, c(19,20,23,24,25)];  
knn.SorteioTeste = order(runif(64))[1:16];  
knn.BaseTeste = knn.Base[knn.SorteioTeste,];  
knn.ClassesTeste = Dados$Classe[knn.SorteioTeste];  
knn.BaseTreino = knn.Base[-knn.SorteioTeste,];  
knn.ClassesTreino = Dados$Classe[-knn.SorteioTeste];  
  
knn.Pred = knn(train=knn.BaseTreino, test=knn.BaseTeste, cl=knn.ClassesTreino,  
k=8);  
Crosstable(knn.ClassesTeste, knn.Pred, chisq=TRUE);
```

A mesma base de dados e variáveis padronizadas utilizadas na análise de cluster foram utilizadas, e os seguintes trabalhos foram realizados para conseguir realizar a previsão:

1. Foi utilizada técnica de particionamento de dados por aleatoriedade, também conhecida como *holdout sampling*. A base total é de 64 observações e para o teste foram utilizadas 16 observações, o equivalente a 25%, de forma aleatória, para quebrar a ordenação natural do ranking da competição;
2. Os demais 75% das observações foram utilizados para treinamento do algoritmo;
3. Uma variável de classificação foi criada apenas para uso neste algoritmo, indicando atletas que foram Finalistas ou Não Finalistas em competições do ano, e para ser utilizada como resultado da previsão;
4. Como número de vizinhos mais próximos, utilizamos como parâmetro o valor 8 por dois motivos:
 - a. Considerar até 8 atletas mais próximos, o que é um número interessante dado um total de 64 participantes;
 - b. Uma métrica que é considerada comum para este algoritmo, que tem como sugestão trabalhar com a raiz quadrada do total de observações, que neste caso também dá 8.

Desta estratégia foram obtidos os seguintes resultados:

Cell Contents

	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

Total observations in Table: 16

Knn.ClassesTeste	Knn.Pred		Row Total
	Finalista	Não Finalista	
Finalista	2	0	2
	12.250	1.750	
	1.000	0.000	0.125
	1.000	0.000	
Não Finalista	0	14	14
	1.750	0.250	
	0.000	1.000	0.875
	0.000	1.000	
Column Total	2	14	16
	0.125	0.875	

Statistics for All Table Factors

Pearson's Chi-squared test

Chi² = 16 d.f. = 1 p = 6.334248e-05

Pearson's Chi-squared test with Yates' continuity correction

Chi² = 8.163265 d.f. = 1 p = 0.004274734

Apesar do algoritmo ter se apresentado extremamente eficaz e gerado, consequentemente, um possível cenário de *overfitting*, diversas execuções foram realizadas mantendo a aleatoriedade da base de teste que demonstrou alguns verdadeiros-negativos e falsos-positivos de forma coerente, nada fora do comum que comprometesse a qualidade do algoritmo.

4.6 Previsão usando algoritmo de regressão

Algoritmos de classificação utilizados para previsões podem apresentar bons resultados conforme demonstrado no item anterior, porém, o que falta para análise de desempenho esportivo é saber o quanto. Tudo normalmente é numérico, uma meta, um objetivo, um volume de treino, uma necessidade de evolução, e, com isso, prever um possível finalista ou não fica uma impressão vaga sem um número minimamente atrelado à previsão, por mais que se tenha um algoritmo eficaz e claro tecnicamente.

Dada esta necessidade, a proposta agora é realizar previsão usando um modelo que possa apontar um escore para a previsão, e para isso foi construído e aplicado um modelo, com as mesmas variáveis utilizadas anteriormente objetivando conseguir sanar esta necessidade numérica.

Como técnica, foi decidido utilizar um modelo baseado em regressão logística pelos seguintes fatores:

1. Utilização como um modelo binário baseado na classificação desejada de Finalista ou Não da competição atual;
2. Possibilitar gerar escore para cada caso;
3. Permitir, a partir do escore, analisar os dados conjuntamente e decidir pontos de corte conforme cada cenário, como exemplo, outras competições que conseqüentemente têm outras características de desempenho.

A seguir a execução do modelo:

```
#####  
# Previsão De Finalistas com Regressão Logística  
#####  
RegLog.Modelo = glm(Finalista ~ WorldRanking_S + PontosCombates_S + RoundTotal_S  
+ Amarelos_S + Dez_S, family=binomial(logit), data=Dados);  
  
# Predição  
RegLog.Prob = predict.glm(RegLog.Modelo, type="response");  
Dados = cbind(Dados, RegLog.Prob);  
  
# Matriz de Confusão  
PontoDeCorte = 0.9;
```

```
Dados$Previsao = "Não Finalista";
Dados$Previsao[which(Dados$RegLog.Prob>=PontoDeCorte)] = "Finalista";
CrossTable(Dados$Classe, Dados$Previsao, chisq=TRUE);
```

Com o resultado, o trabalho seguinte foi testar alguns pontos de corte e avaliar por meio de matriz de confusão um bom ponto a ser considerado. Os seguintes pontos de corte foram analisados:

1. A partir de 0,5, foram obtidos 9 finalistas possíveis;
2. A partir de 0,6, foram obtidos 7 finalistas possíveis;
3. A partir de 0,7, foram obtidos 7 finalistas possíveis;
4. A partir de 0,8, foram obtidos 7 finalistas possíveis;
5. A partir de 0,9, foram obtidos 5 finalistas possíveis.

Abaixo, é apresentada a matriz de confusão para o ponto de corte considerado como 0,9:

Cell Contents	
	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

Total Observations in Table: 64

Dados\$Classe	Dados\$Previsao		Row Total
	Finalista	Não Finalista	
Finalista	4	7	11
	11.478	0.973	
	0.364	0.636	0.172
	0.800	0.119	
	0.062	0.109	
Não Finalista	1	52	53
	2.382	0.202	
	0.019	0.981	0.828
	0.200	0.881	
	0.016	0.812	
Column Total	5	59	64
	0.078	0.922	

Statistics for All Table Factors

Pearson's Chi-squared test

Chi² = 15.03424 d.f. = 1 p = 0.0001055781

Pearson's Chi-squared test with Yates' continuity correction

Chi² = 10.62828 d.f. = 1 p = 0.001113715

Para os casos de ponto de corte a partir de 0,8 e 0,9, que significam atletas com escore acima de 80% e 90%, consideramos ambos os cenários otimizados para previsão, mesmo considerando que neste caso estão indicados 5 atletas como possibilidades de finalistas para apenas 4 vagas possíveis. Como avaliação do modelo a fim de verificar sua força de distinção dos resultados, a seguinte Curva de Característica de Operação do Receptor (Curva ROC) foi analisada, apresentando positivamente uma alta diferenciação dos escores:

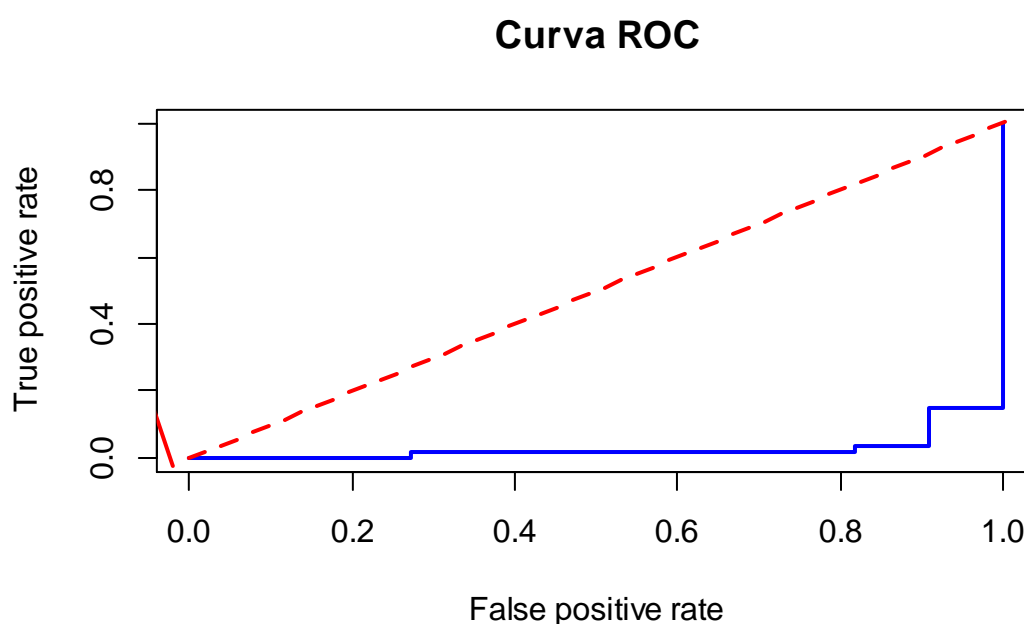


Figura 41 - Curva ROC do modelo de previsão aplicado

Também foi avaliado no modelo as métricas padrões de matrizes de confusão (acurácia, sensibilidade, especificidade, eficiência, previsão), conjuntamente com a estatística Kappa para aferir a eficiência do modelo, com os resultados apresentados conforme a seguir:

Confusion Matrix and Statistics

Prediction	Reference	
	Finalista	Não Finalista
Finalista	4	1
Não Finalista	7	52

Accuracy : 0.875
 95% CI : (0.7685, 0.9445)
 No Information Rate : 0.8281
 P-Value [Acc > NIR] : 0.2070

 Kappa : 0.4398
 McNemar's Test P-Value : 0.0771

 Sensitivity : 0.36364
 Specificity : 0.98113
 Pos Pred Value : 0.80000
 Neg Pred Value : 0.88136
 Prevalence : 0.17188
 Detection Rate : 0.06250
 Detection Prevalence : 0.07812
 Balanced Accuracy : 0.67238

 'Positive' Class : Finalista

Um fator de grande relevância neste modelo é que ele foi capaz de promover casos de atletas que não foram finalistas nenhuma vez no ano para possível finalista desta competição. Neste caso em especial, identificamos o atleta Jean Charles Valladont da França, que foi posicionado na competição como 8º colocado e saiu do cenário de não finalista para possível finalista com escore de 93%. Ao mesmo tempo, também ocorreram casos interessantes de atletas que tiveram ótimo desempenho na competição, mas que não foram classificados como possíveis finalistas, como os casos dos 3º, 5º e 7º colocados na classificação Rio 2016.

CONCLUSÃO

A grande hipótese levantada neste estudo de caso foi a possibilidade de prever finalistas em competições oficiais de tiro com arco. Para buscar as respostas e realizar as tentativas, o desenho de modelos estatísticos fez-se necessário a fim de não avaliar os atletas simplesmente pelo seu resultado final da competição, mas, principalmente, avaliar seu desempenho, a variação de seus tiros, resultados flecha-a-flecha e conflitar métricas de forma eficaz, como o resultado da competição atual, o posicionamento no ranking mundial, métricas específicas de desempenho baseadas no volume de tiros considerados perfeitos e o mesmo volume na região de competitividade do alvo (pontuações nove e dez). Com a aplicação de análise exploratória, cruzamento das métricas citadas e execução de algoritmos de previsão baseado em classificação e modelo de regressão binária, foi possível constatar que sim, é possível e eficaz.

As análises, modelos e resultados apresentados foram realizados e publicados em 05 de agosto de 2016, usando o site public.tableau.com, dia em que ocorreu a competição na fase classificatória, estimulado pela curiosidade não apenas da possibilidade de prever os finalistas que começariam a disputar no dia seguinte, como também comprovar o quanto os modelos seriam eficazes. Os resultados baseados no modelo de regressão logística binária, considerando um ponto de corte para atletas com escore acima de 90%, distinguiu 5 atletas como finalistas possíveis, sendo eles: 1º Kim Woojin, 2º Brady Elisson, 3º Sjef Van Den Berg, 4º Ku Bonchan e 5º Jean Charles Valladont. Para surpresa ou não, a competição continuou conforme o cronograma e no seu último dia foram classificados os seguintes atletas para as finais: Ku Bonchan e Jean Charles Valladont para disputar medalhas de ouro e prata, Brady Elisson e Sjef Van Den Berg para a disputa de medalha de bronze.

Como em qualquer esporte, há fatores que fogem completamente das estatísticas, e neste estudo de caso não foi diferente. O atleta Kim Woojin, primeiro colocado na fase classificatória com uma pontuação que bateu o recorde olímpico e mundial, foi eliminado na segunda etapa da fase de eliminatórias, e perdeu a chance de seguir adiante na competição.

O modelo com o ponto de corte escolhido conseguiu prever cinco atletas como possíveis finalistas, e acertou quatro que chegaram à disputa de medalhas, demonstrando em cenário real a possibilidade e eficiência de prever sim atletas em competições de tiro com arco e até mesmo em outros esportes onde a sorte tem baixa probabilidade de ocorrer.

REFERÊNCIAS BIBLIOGRÁFICAS

HAIR, Joseph et al. (Tradução: Adonai Schlup Sant'Anna). **Análise Multivariada de Dados**. Porto Alegre: Bookman, 2009.

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit: The definitive guide to dimensional modeling**. 3 ed. Indianapolis: Wiley, 2013. 600 p.

KNAFLIC, Cole. **Storytelling with Data: A Data Visualization Guide for Business Professionals**. New Jersey: Wiley, 2015. 288 p.

LANTZ, Brett. **Machine Learning with R: Learn how to use R to apply powerful machine learning methods and gain insight into real-world applications**. Birmingham: Packt Publishing, 2013. 375 p.

RIES, Eric (Tradução: Texto Editores) **A startup enxuta: como os empreendedores atuais utilizam a inovação contínua para criar empresas extremamente bem-sucedidas**. São Paulo: Lua de Papel, 2012. 288 p.

SCHWABER, Ken; SUTHERLAND, Jeff. **Guia do Scrum**. Disponível em: <<http://www.scrumguides.org/docs/scrumguide/v2016/2016-Scrum-Guide-Portuguese-Brazilian.pdf>>. Acesso em: 28 nov. 2016.

VERZANI, John. **Using R for introductory statistics**. 2 ed. Boca Raton: Chapman and Hall/CRC, 2014. 518 p.

World Archery. **Rolebook Target Archery, Book 3**. Disponível em: <<https://rulebook.worldarchery.org/PDF/Official/2016-10-01/EN-Book3.pdf>>. Acesso em: 21 nov. 2016.