

Métodos Estatísticos para Análises de Dados do Esporte - Aula 3

Daniel Takata Gomes

IX Workshop on Probabilistic and Statistical Methods

4 de fevereiro de 2022

Aula 3: Métodos para análise de dados de esportes individuais

Sumário

1 Introdução

Sumário

- ① Introdução
- ② Teoria de Valores Extremos

Sumário

- ① Introdução
- ② Teoria de Valores Extremos
- ③ Análise de eventos extremos no futebol

Sumário

- ① Introdução
- ② Teoria de Valores Extremos
- ③ Análise de eventos extremos no futebol
- ④ Análise de eventos extremos no atletismo/natação

Sumário

- ① Introdução
- ② Teoria de Valores Extremos
- ③ Análise de eventos extremos no futebol
- ④ Análise de eventos extremos no atletismo/natação
- ⑤ Considerações finais



Original research

Swimming performance index based on extreme value theory

Daniel T Gomes¹ and Ligia Henriques-Rodrigues²

International Journal of
Sports Science
& Coaching

International Journal of Sports Science
& Coaching
2019, Vol. 14(1) 51–63
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journalsPermissions
DOI: [10.1177/17479541808068](https://doi.org/10.1177/17479541808068)
journals.sagepub.com/home/sps



Abstract

The International Swimming Federation has developed a points system that allows comparisons of results between different events. Such system is important for several reasons, since it is used as a criterion to rank swimmers in awards and selection procedures of national teams. The points system is based entirely on the world record of the correspondent event. Since it is based on only one observation, this work aims to suggest a new system, based on the probability distribution of the best performances in each event. Using extreme value theory, such distribution, under certain conditions, converges to a generalized Pareto distribution. The new performance index, based on the peaks over threshold methodology, is obtained based on the exceedance probabilities correspondent to the swimmers' times that exceed a given threshold. We work with 17 officially recognized events in 50 m pool, for each women and men, and considered all-time rankings for all events until 31 December 2016. A study on the adequacy of the proposed generalized Pareto distribution index and a comparison between the performances of Usain Bolt and Michael Phelps are also conducted.

Keywords

Aquatic sport, International Swimming Federation, performance analysis

Introduction

Competitive swimming has a non-subjective ranking system, based on times. When one analyzes the results of a given event, such criterion is free from arbitrariness and is the universal way to choose the "best (fastest) swimmers." However, for some reasons, it might be reasonable to compare performances in different events. For example, the winning time of the 100 m freestyle in a given competition is better than the winning time of the 200 m butterfly?

Obviously, the concept of "better" needs to be defined. Intuitively, a great performance is the one that, in its respective event, corresponds to a disreputant result in comparison to the others.

Such comparison is made by the International Swimming Federation (FINA) through a points system. It is important, since it is used to determine the winners of the World Cup.¹ It is also used by several

(Great Britain) were chosen the best performers of the year, by their performances in women's 800 m freestyle and men's 100 m breaststroke, respectively, in Rio de Janeiro Olympic Games.³

The FINA points system is based on what is called the *base times*.² The base times are defined every year, based on the latest world record that was approved by FINA. The base times are defined with the cut date of 31 December. So, in a given event, let $T_i(t)$ be the time obtained by the swimmer i , in seconds, at the year t , and $B(t-1)$ the base time, which is the world record, in

Reviewer: John Ennalah (Tilburg University, Netherlands),
Ian G. McHale (University of Liverpool, UK),
Jonathan Tawn (Lancaster University, UK).

¹National School of Statistics, Brazilian Geography and Statistics Institute,
Rio de Janeiro, Brazil

²Department of Statistics, Institute of Mathematics and Statistics,



STATISTICAL ANALYSIS: HOW FAST IS ADAM PEATY'S WORLD RECORD?



On Sunday, Adam Peaty became the first man to break 57 seconds in the 100 breaststroke. Could a swimmer be capable of a similar performance in another event? Courtesy photo via "Rafael Domeyko Photography"

BY SWIMSWAM CONTRIBUTORS

July 28th, 2019

51

Industry, International, Records

Courtesy: Daniel Takata

A couple of days ago, **Adam Peaty** became the first man to break 57 seconds in the 100 breaststroke as he delivered an incredible 56.88.

He is also the only swimmer to ever break 58 and he has the 17 fastest performances of all time.



Adam Peaty MBE
2.022 Tweets

Tweets Tweets e respostas Mídia Curtidas



Adam Peaty MBE  @adam_pe... · 43m

I've never thought about how fast it would be in another event, but this is a great analysis 

 **SwimSwam**  @swimswamn... · 8h

Statistical Analysis: How Fast is Adam Peaty's World Record?
swimswam.com/statistical-an...

3

7

76



More or Less

[More or Less Home](#) [Episodes](#) [Podcast](#) [More](#)



Peaty v Bolt: Which is the greatest world record?

Using statistics to compare world records in athletics and swimming

Professor Takata calculated the likelihood of Peaty's world record being broken by a swimmer as 0.291%.

Messi breaks Müller's record of goals scored in a calendar year

With his two goals against Betis, the Argentine broke the all-time record of goals scored in a calendar year set by Gerd Müller in 1972

Jordi Clos

09:00PM SAMEDI 08 DÉC.

ÉQUIPE PREMIÈRE

08 déc. 12

PARTAGER CET ARTICLE

 Twitter

 Facebook

 Lien copié

Messi's two goal against Betis took his year tally to 86 (with Barça and Argentina), equalling the record set by Gerd Müller in 1972 (with Bayern Munich and Germany).

Impeccable goalscorer

Messi, one of the three finalists for the FIFA Ballon d'Or 2012, is completing a dream-like year. He was La Liga's Pichichi last year with 50 goals; he was awarded the Golden Boot,

FORÇA BARÇA

21





FUTEBOL | Fla bate recordes de pontos com 20 clubes e de aproveitamento no geral

Fla bate recordes de pontos com 20 clubes e de aproveitamento no geral

Com a vitória sobre Ceará, rubro-negro chega a 73,68% em relação a todo o campeonato, índice maior até do que o dos 100 pontos do Cruzeiro em 2003

FUTEBOL | Eduardo Marini, do R7

28/11/2019 - 00H06 (ATUALIZADO EM 29/11/2019 - 00H33)

COMPARTILHE:



Ouvir: es e de aproveitamento no geral - Esportes

0:00 audílio

0 A- A+



[HOME > NBA](#)

Warriors Finish 73-9, Break Bulls' NBA Record for Regular-Season Wins

SCOTT POLACEK 

APRIL 14, 2016





Odds By **tipico**
SPORTSBOOK

[News](#)[\[Sports \]](#)[Entertainment](#)[Life](#)[Money](#)[Tech](#)[Travel](#)[Opinion](#)**RIO 2016**

Bob Beamon's Olympic long-jump record still stands nearly 50 years later

Paul Myerberg USA TODAY Sports

Published 11:26 p.m. ET Aug. 13, 2016 | Updated 11:38 p.m. ET Aug. 13, 2016



Ranking, and other properties, of elite swimmers using extreme value theory

Harry Spearing¹ | Jonathan Tawn¹ | David Irons² | Tim Paulden² | Grace Bennett²

¹Lancaster University, Lancashire, UK

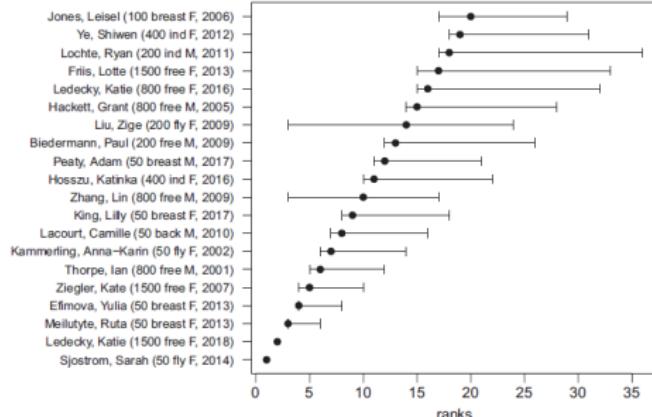
²ATASS sports, Exeter, UK

Correspondence

Harry Spearing, Lancaster University,
Lancashire, UK.
Email: h.spearing@lancaster.ac.uk

Abstract

The International Swimming Federation (FINA) uses a very simple points system with the aim to rank swimmers across all swimming events. The points acquired is a function of the ratio of the recorded time and the current world



Tratam-se de desempenhos extraordinários, extremos.

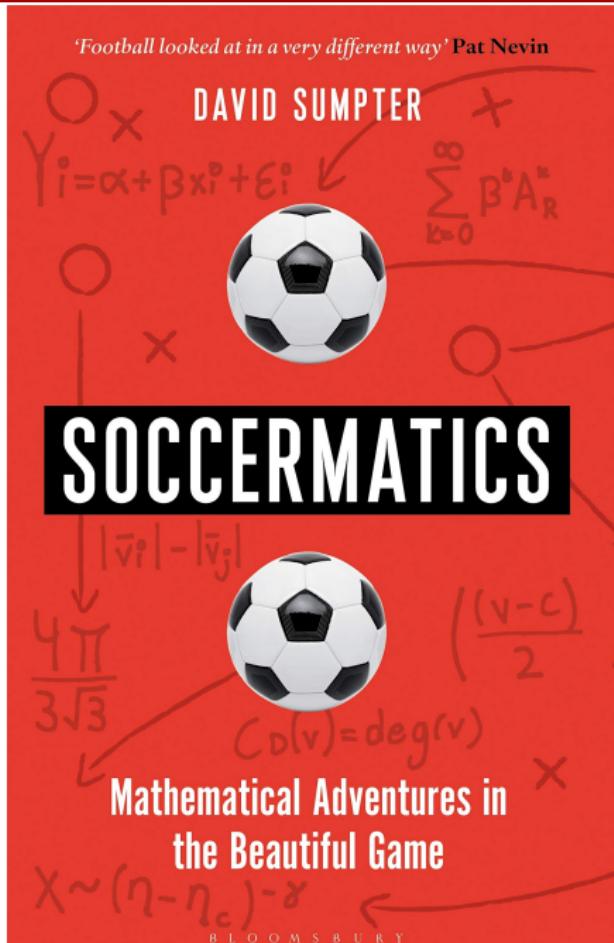
Tratam-se de desempenhos extraordinários, extremos.

Desempenhos extraordinários são observados de tempos em tempos, tanto em esportes individuais quanto em esportes coletivos.

Tratam-se de desempenhos extraordinários, extremos.

Desempenhos extraordinários são observados de tempos em tempos, tanto em esportes individuais quanto em esportes coletivos.

Como tais desempenhos podem ser comparáveis/mensuráveis? O quanto um determinado recorde é extraordinário?



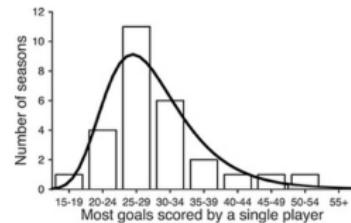


Figure 4.2 Histogram of number of goals scored by the winners of the Pichichi Trophy between 1986/87 and 2013/14 (boxes) compared to the extreme-value distribution (solid line).

New Extremes?

In January 2015, Messi was awarded FIFA World Player of the Year for the fifth time, moving ahead of Ronaldo who has won it on three occasions. Messi was not, however, the first player to win the award for the fifth time. I had the pleasure of watching the Brazilian five-times winner in a Champions League final in 2007, in a match played at a football ground less than a kilometre from where I lived at the time. The venue was Gammliaallen in Umeå, Sweden, and the player was the striker Marta. She is a truly exceptional player. When play-

A Teoria de Valores Extremos trata de fenômenos, como o nome diz, extremos.

A Teoria de Valores Extremos trata de fenômenos, como o nome diz, extremos.

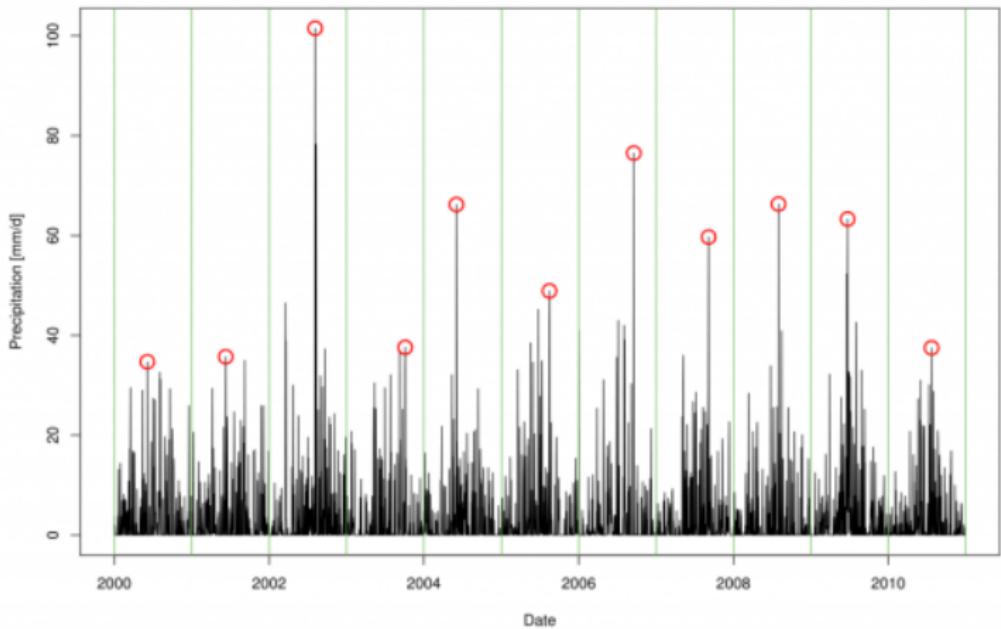
Eventos extremos: dias mais quentes ou mais chuvosos, terremotos, recordes no esporte.

A Teoria de Valores Extremos trata de fenômenos, como o nome diz, extremos.

Eventos extremos: dias mais quentes ou mais chuvosos, terremotos, recordes no esporte.

Mesmo que um evento extremo ainda não tenha sido observado, essa metodologia permite, por exemplo, calcular a probabilidade de sua ocorrência.

Block Maxima Method – Annual Maxima Series



Análise de máximo de blocos

Seja X_1, X_2, \dots, X_N uma sequência de variáveis aleatórias i.i.d. com função distribuição F . Sejam $X_{1:N} \leq X_{2:N} \leq \dots \leq X_{N:N}$ as estatísticas de ordem associadas.

Seja X_1, X_2, \dots, X_N uma sequência de variáveis aleatórias i.i.d. com função distribuição F . Sejam $X_{1:N} \leq X_{2:N} \leq \dots \leq X_{N:N}$ as estatísticas de ordem associadas.

Se existirem sequências reais $a_N > 0$ e $b_N \in \mathbb{R}$ tais que $(X_{N:N} - b_N)/a_N$ converge em distribuição para alguma $G(x)$ não degenerada, ou seja,

Seja X_1, X_2, \dots, X_N uma sequência de variáveis aleatórias i.i.d. com função distribuição F . Sejam $X_{1:N} \leq X_{2:N} \leq \dots \leq X_{N:N}$ as estatísticas de ordem associadas.

Se existirem sequências reais $a_N > 0$ e $b_N \in \mathbb{R}$ tais que $(X_{N:N} - b_N)/a_N$ converge em distribuição para alguma $G(x)$ não degenerada, ou seja,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\frac{X_{N:N} - b_N}{a_N} \leq x \right) = \lim_{N \rightarrow \infty} F^N(a_N x + b_N) = G(x),$$

então G está na família de G_ξ , em que G_ξ é a função distribuição da distribuição de valor extremo generalizada (GEV), dada por

então G está na família de G_ξ , em que G_ξ é a função distribuição da distribuição de valor extremo generalizada (GEV), dada por

$$G_\xi(x) = \begin{cases} \exp\{-(1 + \xi x)^{-1/\xi}\}, & \text{se } \xi \neq 0 \\ \exp\{\exp(-x)\}, & \text{se } \xi = 0, \end{cases}$$

em que $1 + \xi x > 0$ e ξ está relacionado com o peso da cauda direita de G_ξ .

então G está na família de G_ξ , em que G_ξ é a função distribuição da distribuição de valor extremo generalizada (GEV), dada por

$$G_\xi(x) = \begin{cases} \exp\{- (1 + \xi x)^{-1/\xi}\}, & \text{se } \xi \neq 0 \\ \exp\{\exp(-x)\}, & \text{se } \xi = 0, \end{cases}$$

em que $1 + \xi x > 0$ e ξ está relacionado com o peso da cauda direita de G_ξ .

Tal resultado, conhecido como teorema de Fisher-Tippet, é considerado uma versão análoga ao Teorema Limite Central para máximos de variáveis aleatórias i.i.d. (Beirlant et al., 2004).

O parâmetro de forma ξ determina o peso da cauda da distribuição.

O parâmetro de forma ξ determina o peso da cauda da distribuição.

As sub-famílias definidas por $\xi = 0$, $\xi > 0$ e $\xi < 0$ correspondem, respectivamente, às famílias Gumbel, Fréchet e Weibull.

Eventos extremos no futebol

Eventos extremos no futebol

1986–87	Hugo Sánchez (3)	Real Madrid	34	41	0.829
1987–88	Hugo Sánchez (4)	Real Madrid	29	36	0.806
1988–89	Baltazar	Atlético Madrid	35	36	0.972
1989–90	Hugo Sánchez (5)	Real Madrid	38	35	1.086
1990–91	Emilio Butragueño	Real Madrid	19	35	0.543
1991–92	Manolo	Atlético Madrid	27	36	0.75
1992–93	Bebeto	Deportivo La Coruña	29	37	0.784
1993–94	Romário	Barcelona	30	33	0.909
1994–95	Iván Zamorano	Real Madrid	28	38	0.737
1995–96	Juan Antonio Pizzi	Tenerife	31	41	0.756
1996–97	Ronaldo	Barcelona	34	37	0.919
1997–98	Christian Vieri	Atlético Madrid	24	24	1
1998–99	Raúl	Real Madrid	25	37	0.676
1999–00	Salva	Racing Santander	27	36	0.75
2000–01	Raúl (2)	Real Madrid	24	36	0.667
2001–02	Diego Tristán	Deportivo La Coruña	21 ^[v]	35	0.6
2002–03	Roy Makaay	Deportivo La Coruña	29	38	0.763
2003–04	Ronaldo (2)	Real Madrid	25 ^[w]	32	0.781
2004–05	Diego Forlán	Villarreal	25 ^[x]	38	0.658
2005–06	Samuel Eto'o	Barcelona	26	34	0.765
2006–07	Ruud van Nistelrooy	Real Madrid	25	37	0.676
2007–08	Daniel Güiza	Mallorca	27	37	0.73
2008–09	Diego Forlán (2)	Atlético Madrid	32	33	0.97
2009–10	Lionel Messi	Barcelona	34	35	0.971
2010–11	Cristiano Ronaldo	Real Madrid	41 ^{[5][y]}	34	1.206
2011–12	Lionel Messi (2)	Barcelona	50	37	1.351
2012–13	Lionel Messi (3)	Barcelona	46	32	1.438
2013–14	Cristiano Ronaldo (2)	Real Madrid	31	30	1.033

attack and defence can change over the seasons. We'll come back to this second assumption later, but for now we'll assume that it holds and see what the extreme-value theory tells us.

[Figure 4.2](#) shows the histogram for the top goalscorers from the 1986/87 season to the 2013/14 season. The solid line is the curve for the extreme-value distribution.⁵ Using the extreme-value model, I can now work out just how exceptional Messi's 50 goals were. Overall there is a reasonable enough match between the actual data observed so far to reassure us that the extreme-value theory could be useful. But if you look carefully at the bottom-right-hand corner of [Figure 4.2](#), you'll see that Messi's 50 goals are above the curve derived from theory. The area under the curve corresponding to 50 or more goals is the small shaded area. Here the theory curve is quite a bit lower than the data histogram. In fact, the area under the 50+ part of the curve is only 1.36%, or 1/73, of its total area.⁶ So the model tells us that we should expect a performance like Messi's to come along once every 73 years. The average life expectancy in Argentina is 75 years. Put in those terms, Messi really is a once-in-a-lifetime event.

A probabilidade de um jogador marcar 50 gols ou mais em um campeonato, como Lionel Messi fez em 2011-12, é de 1,36%, ou 1/73.

A probabilidade de um jogador marcar 50 gols ou mais em um campeonato, como Lionel Messi fez em 2011-12, é de 1,36%, ou 1/73.

Como há um campeonato por ano, então espera-se que o feito de Messi se repita, em média, uma vez a cada 73 anos.

veja

[ASSINAR](#) [BUSCAR](#)[RADAR](#) [COLUNISTAS](#) [ECONOMIA](#) [POLÍTICA](#) [MUNDO](#) [CIÊNCIA E TECNOLOGIA](#) [PLACAR](#) [ENTRETENIMENTO](#)

Esporte

Uma a cada 69 anos: campanha do Fla no Brasileiro é exceção estatística

Daniel Takata, professor da Escola Nacional de Ciências Estatísticas do Rio de Janeiro, estudou os números quase inalcançáveis do rubro-negro

Por [Alexandre Senedal, Luiz Felipe Castro](#) - Atualizado em 13 nov 2019, 09h54 - Publicado em 13 nov 2019, 09h17



Flamengo, líder com 77 pontos em 32 rodadas Kalo Lakato/VEJA

Anúncio fechado por Google



São 77 pontos em 32 rodadas. Dez de vantagem para o segundo colocado. A marca que coloca o **Flamengo** com uma mão na taça a seis jogos do fim do **Campeonato Brasileiro** é um recorde. Utilizando um modelo matemático para analisar todas as edições do torneio desde 2003 – quando a competição passou a ser disputada por pontos corridos

Essa abordagem utilizando teoria de valores extremos considera o estudo da distribuição de um conjunto de máximos, e, como se vê, pode ser aplicada a contextos esportivos diferentes.

Essa abordagem utilizando teoria de valores extremos considera o estudo da distribuição de um conjunto de máximos, e, como se vê, pode ser aplicada a contextos esportivos diferentes.

Há uma outra abordagem que trata do estudo não de máximos, mas de um conjunto de observações extremas.

Essa abordagem utilizando teoria de valores extremos considera o estudo da distribuição de um conjunto de máximos, e, como se vê, pode ser aplicada a contextos esportivos diferentes.

Há uma outra abordagem que trata do estudo não de máximos, mas de um conjunto de observações extremas.

Utilizando essa abordagem, é possível estimar, por exemplo, o limite humano em provas de atletismo/natação.

Essa abordagem utilizando teoria de valores extremos considera o estudo da distribuição de um conjunto de máximos, e, como se vê, pode ser aplicada a contextos esportivos diferentes.

Há uma outra abordagem que trata do estudo não de máximos, mas de um conjunto de observações extremas.

Utilizando essa abordagem, é possível estimar, por exemplo, o limite humano em provas de atletismo/natação.

Também é possível comparar desempenhos entre diferentes provas.

Michael Johnson - Olimpíada de Atlanta/1996

Como quantificar qual desempenho é mais expressivo?

Para comparação de desempenhos em diferentes provas, a ideia é considerar a distribuição de probabilidade das marcas dos atletas em cada prova.

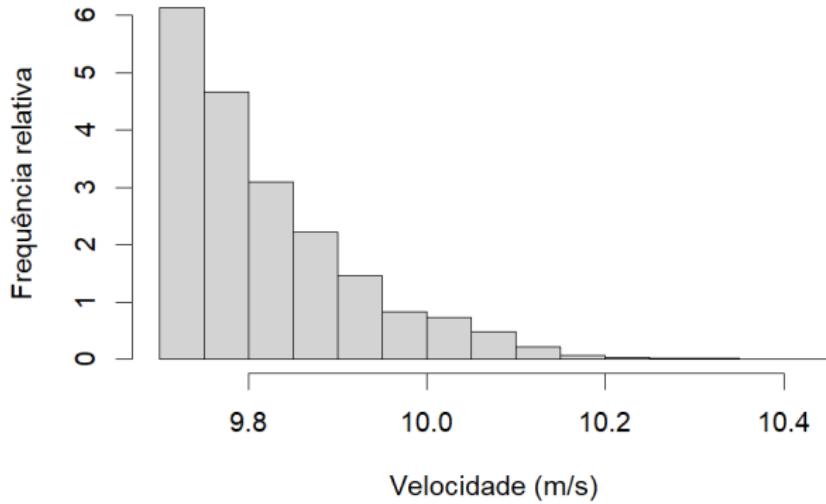
Para comparação de desempenhos em diferentes provas, a ideia é considerar a distribuição de probabilidade das marcas dos atletas em cada prova.

Ao invés de considerar apenas o recorde mundial (que é a melhor marca obtida na história), será considerada a distribuição das melhores marcas da história em cada prova, limitada a uma marca por atleta (para garantir independência).

Problema!

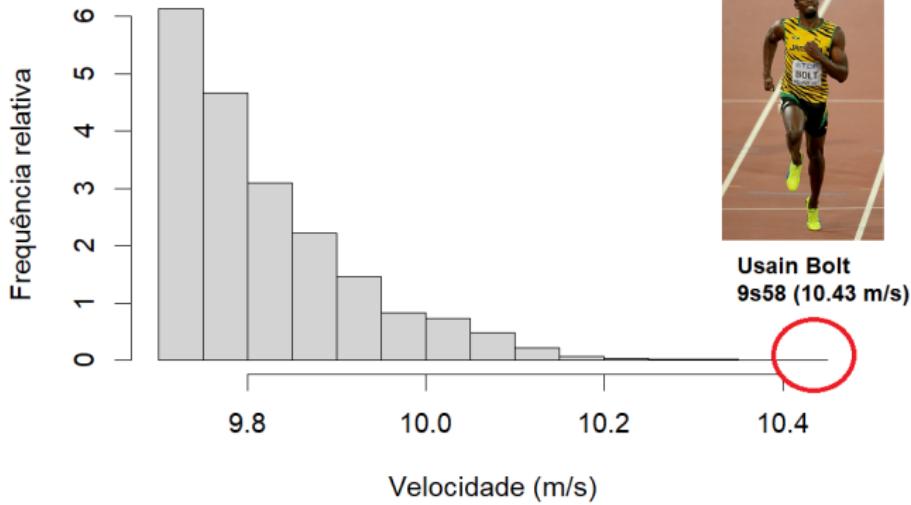
Problema!

100 m rasos masculino



Problema!

100 m rasos masculino



O histograma representa as melhores marcas dos 500 nadadores mais rápidos da história dos 100 m livre masculino.

O histograma representa as melhores marcas dos 500 nadadores mais rápidos da história dos 100 m livre masculino.

É impossível determinar um tempo máximo alcançado por um atleta. E, obviamente, a distribuição de todas as marcas observadas por todos os atletas possíveis é impossível de se determinar.

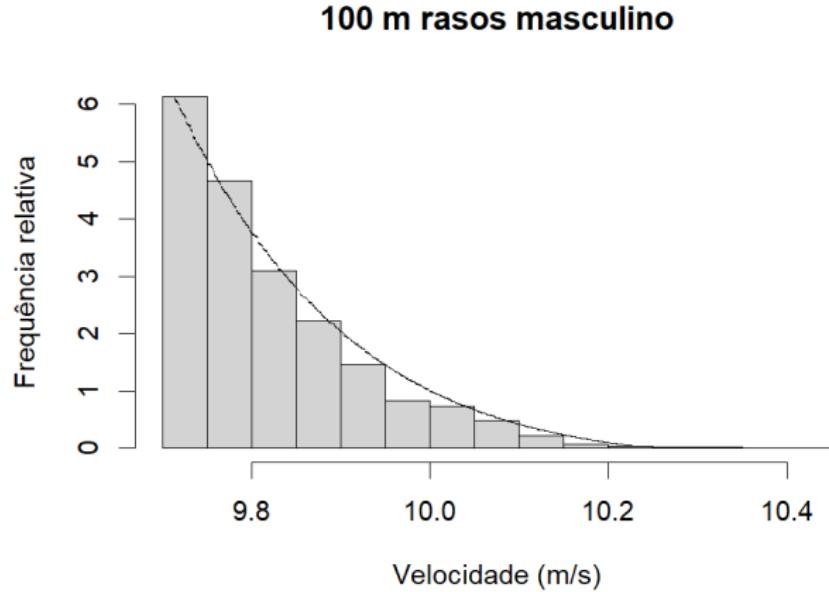
O histograma representa as melhores marcas dos 500 nadadores mais rápidos da história dos 100 m livre masculino.

É impossível determinar um tempo máximo alcançado por um atleta. E, obviamente, a distribuição de todas as marcas observadas por todos os atletas possíveis é impossível de se determinar.

Felizmente, isso não é necessário.

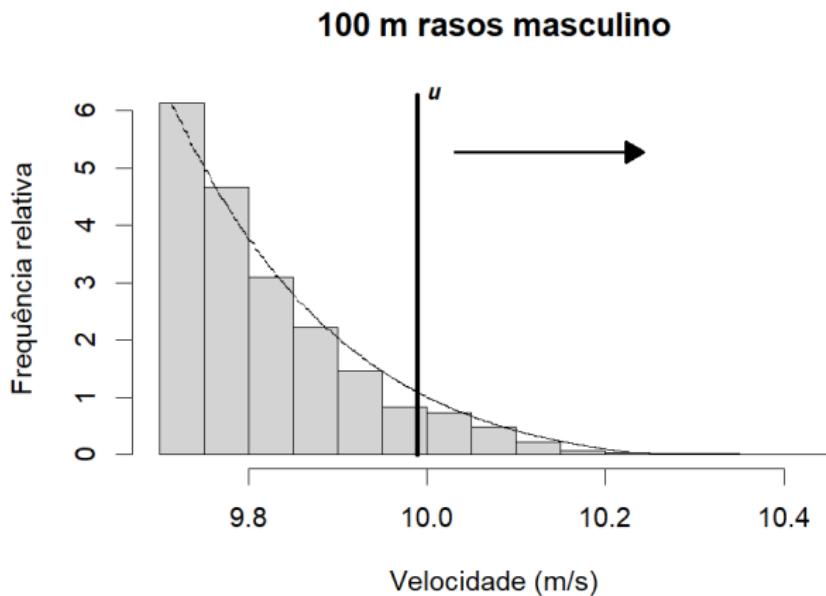
Suponha que a curva represente a distribuição de probabilidade das marcas de todos os atletas da história (que não conhecemos).

Suponha que a curva represente a distribuição de probabilidade das marcas de todos os atletas da história (que não conhecemos).



Há resultados em Teoria de Valores Extremos que garantem que, à direita de um limiar u , a exata distribuição de probabilidade pode ser determinada. Ou seja, a distribuição da cauda da distribuição original pode ser determinada.

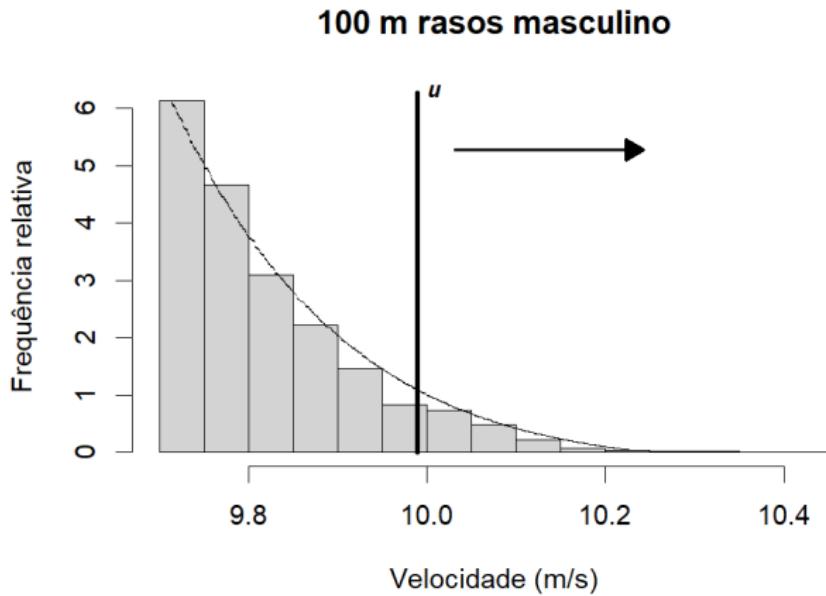
Há resultados em Teoria de Valores Extremos que garantem que, à direita de um limiar u , a exata distribuição de probabilidade pode ser determinada. Ou seja, a distribuição da cauda da distribuição original pode ser determinada.



Metodologia POT (*Peak-Over-Threshold*).

Digamos que X_1, X_2, \dots representem as velocidades das marcas dos atletas representados no gráfico a seguir, e que sigam a distribuição de uma variável aleatória X com função distribuição F . O objetivo é encontrar a distribuição de X acima de um certo limiar u :

Digamos que X_1, X_2, \dots representem as velocidades das marcas dos atletas representados no gráfico a seguir, e que sigam a distribuição de uma variável aleatória X com função distribuição F . O objetivo é encontrar a distribuição de X acima de um certo limiar u :



Ou seja, deseja-se obter a distribuição condicional

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}.$$

Teorema de Pickands-Balkema-de Hann: a distribuição condicional F_u , à medida que o valor do limiar u cresce, converge para a distribuição de Pareto Generalizada (GPD), sob a validade do teorema de Fisher-Tippet.

Teorema de Pickands-Balkema-de Hann: a distribuição condicional F_u , à medida que o valor do limiar u cresce, converge para a distribuição de Pareto Generalizada (GPD), sob a validade do teorema de Fisher-Tippet.

$$G_{\theta,\xi}(x) = \begin{cases} 1 - (1 + \xi x / \theta)^{-1/\xi}, & \text{se } \xi \neq 0, \\ 1 - \exp\{-x/\theta\}, & \text{se } \xi = 0, \end{cases}$$

em que $x \geq 0$ se $\xi \geq 0$ e $0 \leq x \leq -\theta/\xi$ se $\xi < 0$ (Beirlant et al., 2004).

Desafio: determinar o limiar u .

Desafio: determinar o limiar u .

A escolha do limiar u é um problema em aberto em Teoria de Valores Extremos.

Desafio: determinar o limiar u .

A escolha do limiar u é um problema em aberto em Teoria de Valores Extremos.

Alguns métodos têm sido sugeridos na literatura. Um dos mais utilizados é a função de excesso médio (*mean excess function*) (Davison & Smith, 1990).

Desafio: determinar o limiar u .

A escolha do limiar u é um problema em aberto em Teoria de Valores Extremos.

Alguns métodos têm sido sugeridos na literatura. Um dos mais utilizados é a função de excesso médio (*mean excess function*) (Davison & Smith, 1990).

$$e(u) = \mathbb{E}(X - u | X > u) = \frac{\theta + \xi u}{1 - \xi}.$$

Nessa aplicação, a escolha do limiar não é um problema.

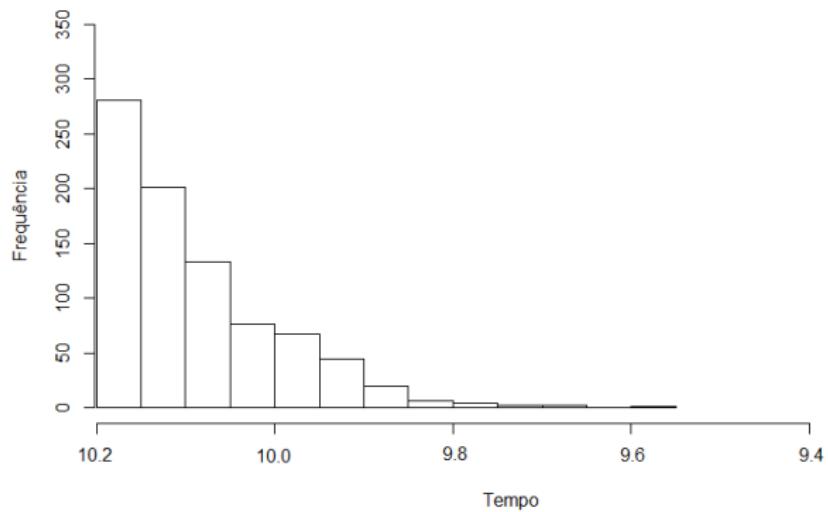
Nessa aplicação, a escolha do limiar não é um problema.

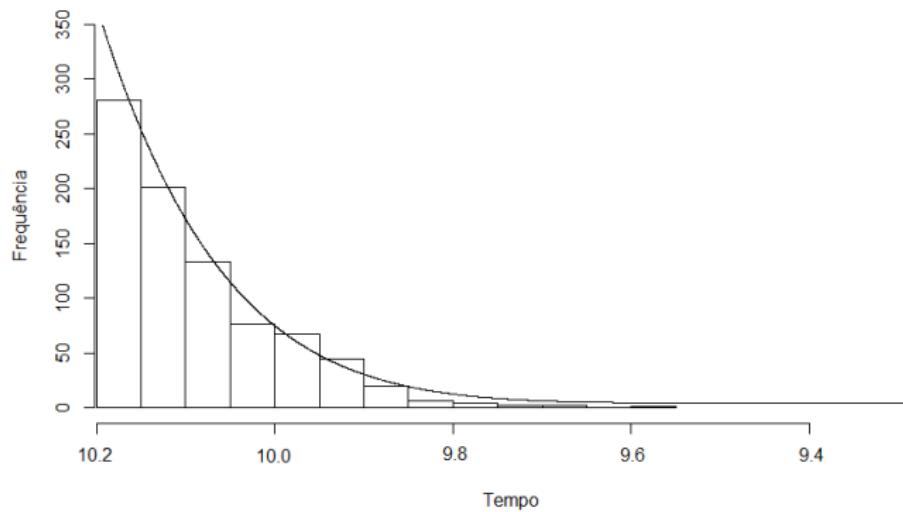
Na prática, temos acesso aos tempos dos atletas mais rápidos da história, que correspondem a um pedaço da cauda da distribuição de todos os atletas do mundo.

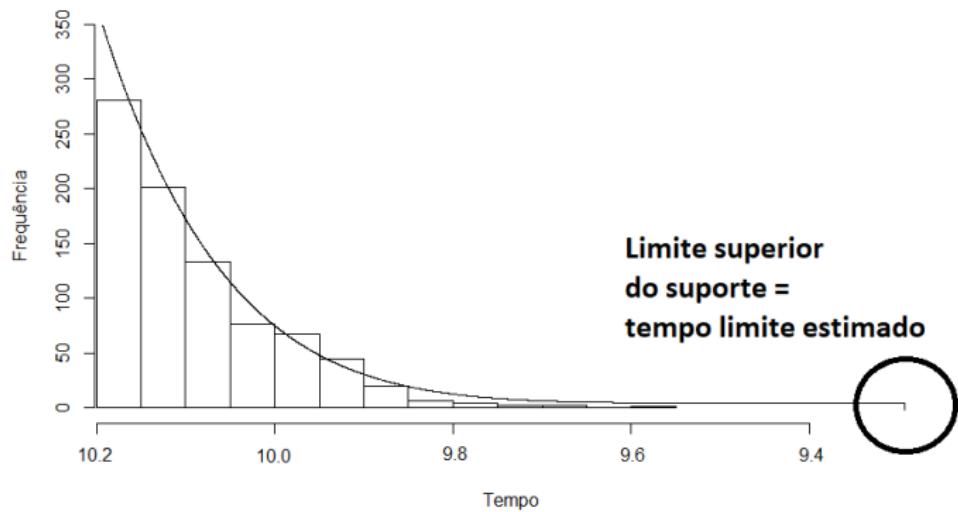
Nessa aplicação, a escolha do limiar não é um problema.

Na prática, temos acesso aos tempos dos atletas mais rápidos da história, que correspondem a um pedaço da cauda da distribuição de todos os atletas do mundo.

Logo, trata-se de uma parte muito pequena da distribuição, e a aproximação para a GPD não é problema.







Quando o limite superior do suporte da distribuição é finito, é possível estimar o tempo limite alcançável pelos atletas em cada prova.

Quando o limite superior do suporte da distribuição é finito, é possível estimar o tempo limite alcançável pelos atletas em cada prova.

Einmahl, J. H., Magnus, J. R. (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association*, 103(484), 1382-1391.

Quando o limite superior do suporte da distribuição é finito, é possível estimar o tempo limite alcançável pelos atletas em cada prova.

Einmahl, J. H., Magnus, J. R. (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association*, 103(484), 1382-1391.

Os pesquisadores estimaram os tempos limite nas provas de atletismo

Einmahl

Event	Men			Women		
	endpoint	prec.	world rec.	endpoint	prec.	world rec.
100m	9.29	0.39	9.77	10.11	0.40	10.49
110/100m h.	12.38	0.35	12.88	11.98	0.19	12.21
200m	18.63	0.88	19.32	20.75	0.57	21.34
400m	—	—	43.18	45.79	1.83	47.60
800m	1:39.65	1.44	1:41.11	1:52.28	1.39	1:53.28
1500m	3:22.63	3.31	3:26.00	3:48.33	2.78	3:50.46
10,000m	—	—	26:17.53	—	—	29:31.78
marathon	2:04:06	57	2:04:55	2:06:35	10:05	2:15:25
shot put	24.80	1.25	23.12	23.70	0.86	22.63
javelin throw	106.50	10.30	98.48	72.50	2.99	71.70
discus throw	77.00	2.85	74.08	85.00	8.10	76.80
long jump	—	—	8.95	—	—	7.52
high jump	2.50	0.05	2.45	2.15	0.05	2.09

Table 3: Ultimate world records, their precisions, and the current world records.

Prova	$\hat{\theta}$	$EP(\hat{\theta})$	$\hat{\xi}$	$EP(\hat{\xi})$	Limite	Rec. mundial
F 100m rasos	0.160	0.007	-0.148	0.029	10.12	10.49
F 200m rasos	0.173	0.007	-0.168	0.028	20.68	21.34
F 400m rasos	0.165	0.008	-0.177	0.030	46.37	47.60
F 800m rasos	0.150	0.006	-0.265	0.022	1:52.30	1:53.28
F 1500m rasos	0.158	0.006	-0.273	0.022	3:47.97	3:50.07
F 5000m rasos	0.130	0.007	-0.182	0.040	13:36.47	14:06.62
F 10000m rasos	0.124	0.006	-0.163	0.031	28:10.54	29:01.03
F 100m barreiras	0.171	0.011	-0.307	0.042	11.85	12.20
F 400m barreiras	0.191	0.014	-0.308	0.052	51.51	51.46
M 100m rasos	0.129	0.005	0.142	0.026	9.36	9.58
M 200m rasos	0.127	0.005	-0.117	0.024	18.50	19.19
M 400m rasos	0.116	0.005	-0.178	0.027	42.35	43.03
M 800m rasos	0.103	0.005	-0.224	0.030	1:39.92	1:40.91
M 1500m rasos	0.105	0.005	-0.200	0.029	3:22.49	3:26.00
M 5000m rasos	0.111	0.005	-0.223	0.027	12:26.67	12:35.36
M 10000m rasos	0.098	0.004	-0.214	0.025	26.00.82	26:11.00
M 110m barreiras	0.146	0.008	-0.202	0.040	12.49	12.80
M 400m barreiras	0.146	0.007	-0.183	0.031	45.46	46.78

Gomes, D. T., Henriques-Rodrigues, L. (2019). Swimming performance index based on extreme value theory. *International Journal of Sports Science & Coaching*, 14(1), pp. 51-62.

Gomes, D. T., Henriques-Rodrigues, L. (2019). Swimming performance index based on extreme value theory. *International Journal of Sports Science & Coaching*, 14(1), pp. 51-62.

Utilizaram teoria de valores extremos para comparar diferentes desempenhos em provas de natação e determinar o mais significativo.

Gomes, D. T., Henriques-Rodrigues, L. (2019). Swimming performance index based on extreme value theory. *International Journal of Sports Science & Coaching*, 14(1), pp. 51-62.

Utilizaram teoria de valores extremos para comparar diferentes desempenhos em provas de natação e determinar o mais significativo.

O método pode ser adaptado para provas do atletismo.

Ideia: verificar o quanto cada recorde é extremo, em relação à distribuição dos outros atletas da história da respectiva prova.

Ideia: verificar o quanto cada recorde é extremo, em relação à distribuição dos outros atletas da história da respectiva prova.

Quanto mais improvável é o recorde, mais extremo ele é (mais "forte").

Ideia: verificar o quanto cada recorde é extremo, em relação à distribuição dos outros atletas da história da respectiva prova.

Quanto mais improvável é o recorde, mais extremo ele é (mais "forte").

Há alguns recordes antigos considerados muito difíceis de serem superados, e a metodologia pode quantificar o quanto eles são expressivos.

2009



2019



2009



2019



Vídeos dos 100m rasos masculino.

Evolução do recorde mundial dos 100m rasos masculino:

Evolução do recorde mundial dos 100m rasos masculino:

- Jim Hines 9s95 (1968)

Evolução do recorde mundial dos 100m rasos masculino:

- Jim Hines 9s95 (1968)
- Asafa Powell 9s77 (2006)

Evolução do recorde mundial dos 100m rasos masculino:

- Jim Hines 9s95 (1968)
- Asafa Powell 9s77 (2006)
- Usain Bolt 9s58 (2009)

Evolução do recorde mundial dos 100m rasos masculino:

- Jim Hines 9s95 (1968)
- Asafa Powell 9s77 (2006)
- Usain Bolt 9s58 (2009)

Comparações:

Evolução do recorde mundial dos 100m rasos masculino:

- Jim Hines 9s95 (1968)
- Asafa Powell 9s77 (2006)
- Usain Bolt 9s58 (2009)

Comparações:

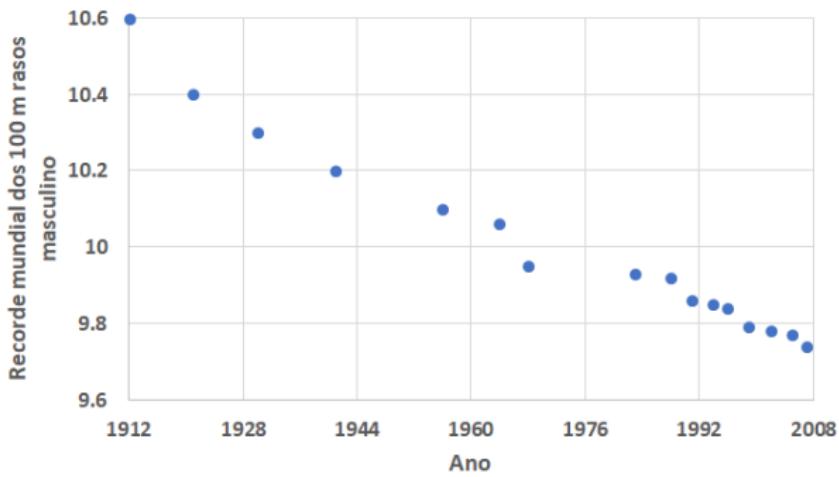
- De 1968 a 2006 (38 anos) o recorde caiu 18 centésimos.

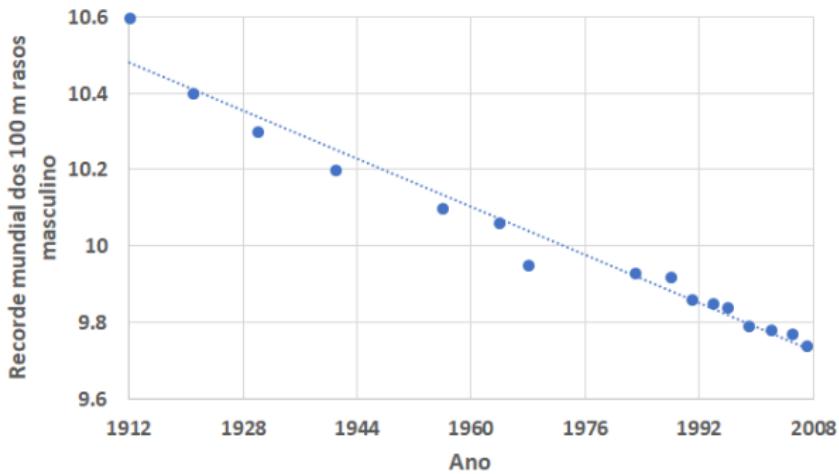
Evolução do recorde mundial dos 100m rasos masculino:

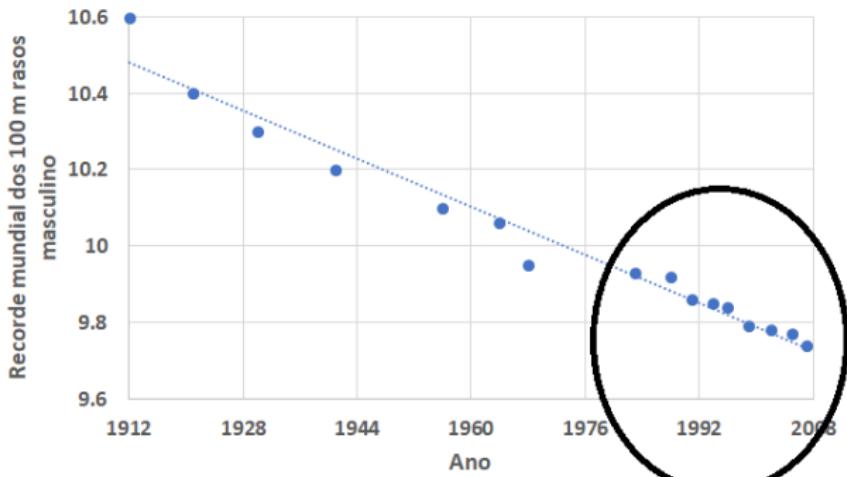
- Jim Hines 9s95 (1968)
- Asafa Powell 9s77 (2006)
- Usain Bolt 9s58 (2009)

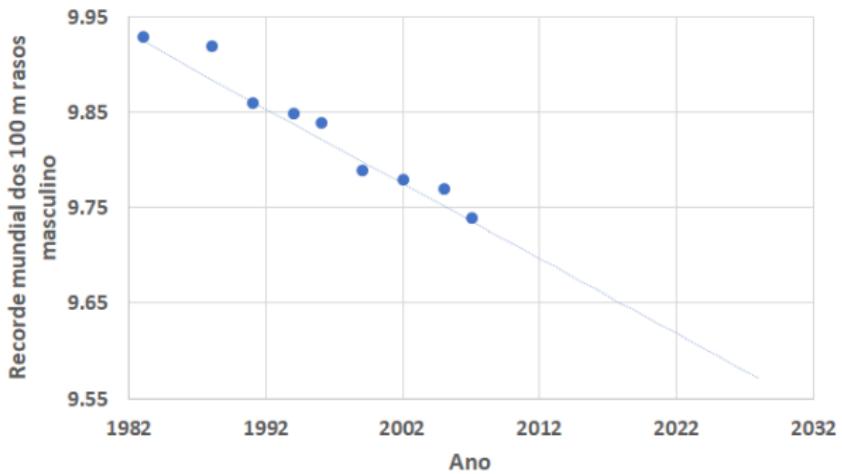
Comparações:

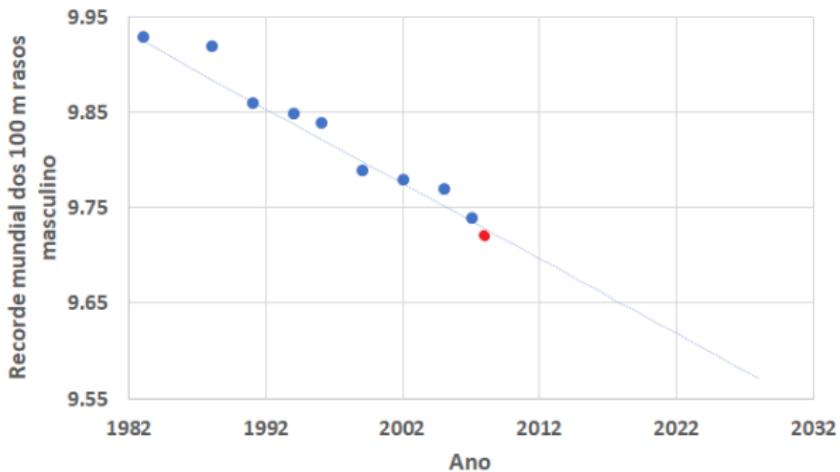
- De 1968 a 2006 (38 anos) o recorde caiu 18 centésimos.
- De 2006 a 2009 (3 anos) o recorde caiu 19 centésimos.

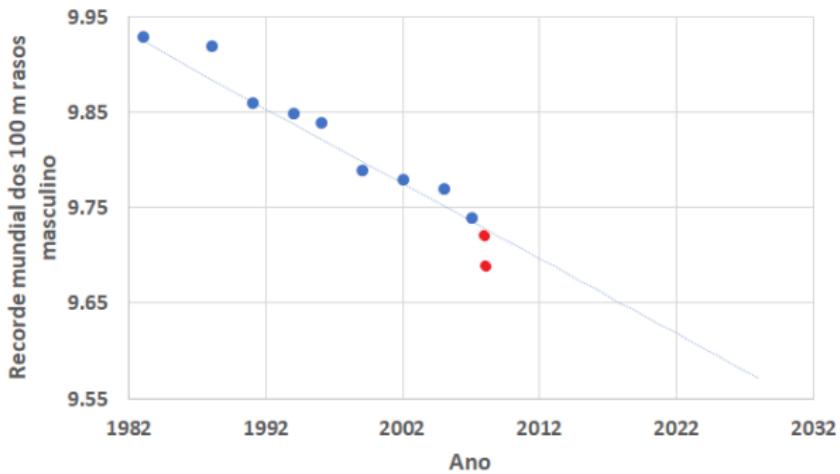


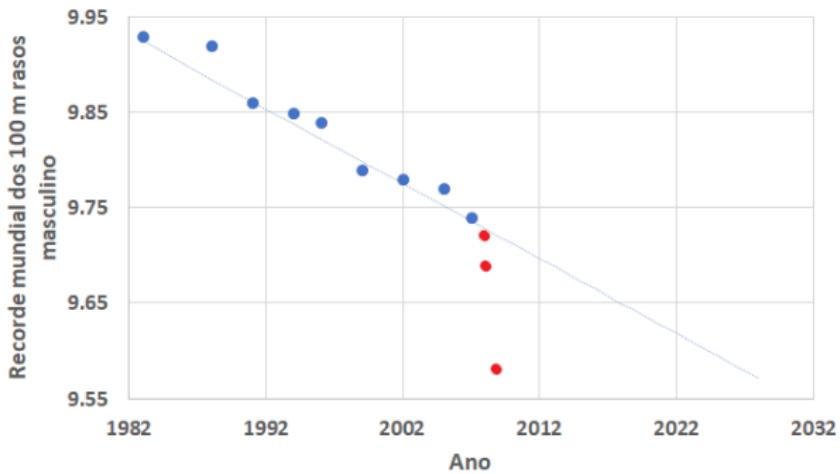


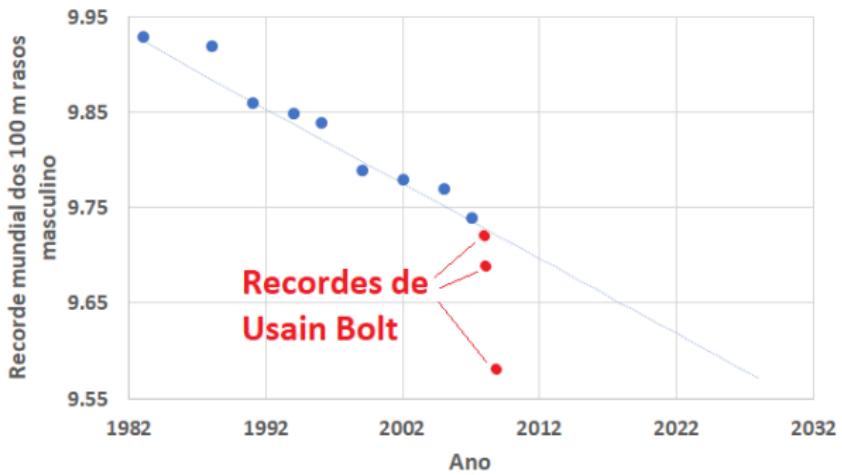












Recordes de
Usain Bolt

O modelo é otimista, e ainda assim o tempo de 9s58 de Usain Bolt obtido em 2009 seria esperado somente após 2025!

O modelo é otimista, e ainda assim o tempo de 9s58 de Usain Bolt obtido em 2009 seria esperado somente após 2025!

O recorde não foi apenas excepcional. Foi surpreendentemente excepcional!

O modelo é otimista, e ainda assim o tempo de 9s58 de Usain Bolt obtido em 2009 seria esperado somente após 2025!

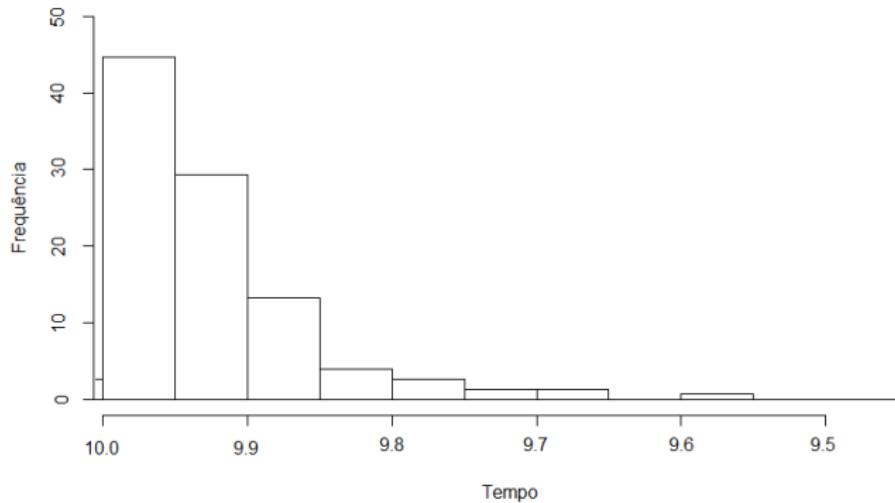
O recorde não foi apenas excepcional. Foi surpreendentemente excepcional!

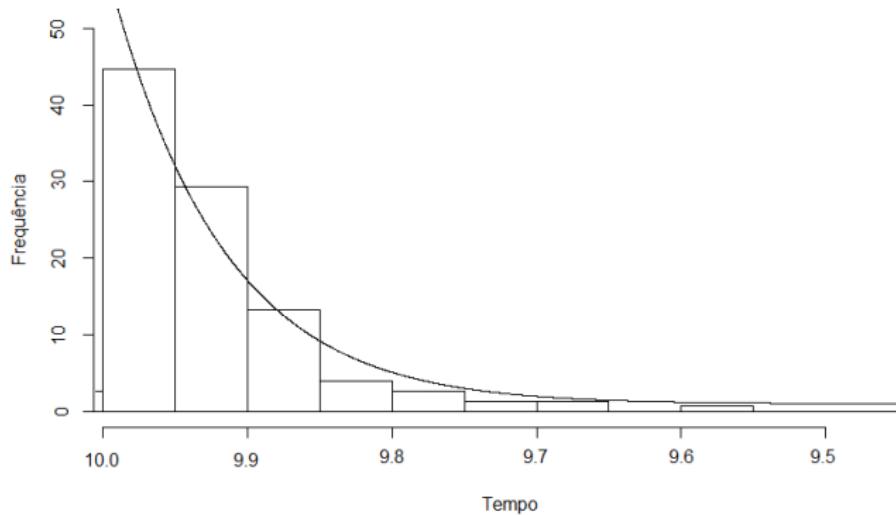
Será esse o recorde mais expressivo entre todas as provas do atletismo?

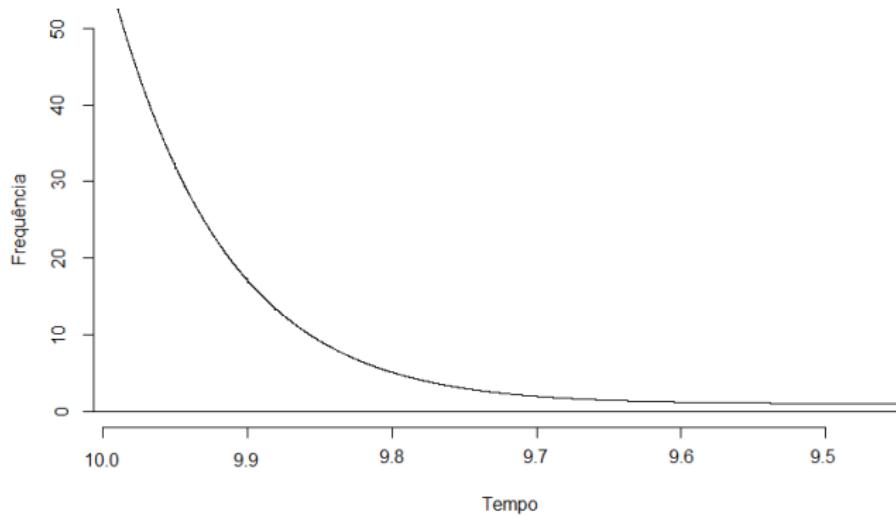
Novamente aqui, utiliza-se a metodologia POT.

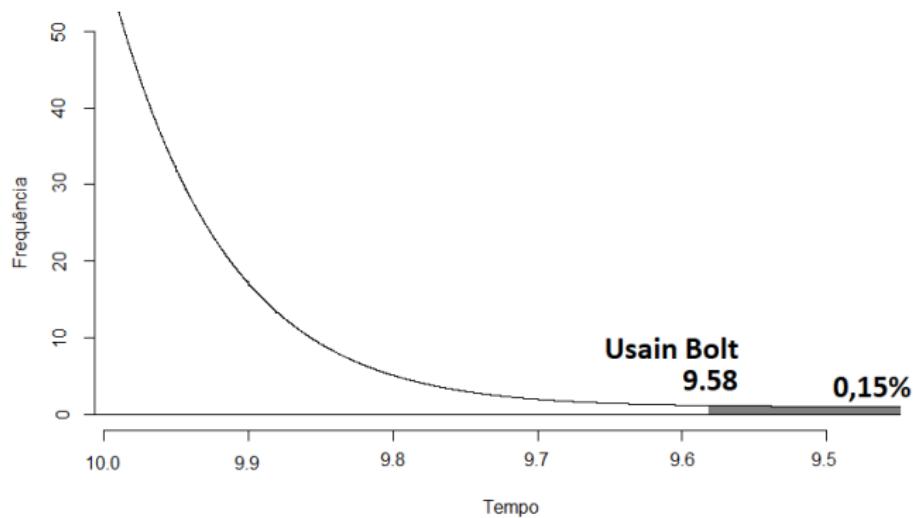
Novamente aqui, utiliza-se a metodologia POT.

Para as análises a seguir, em todas as provas, tomam-se limiares de tal modo que 150 observações sejam consideradas. Ter o mesmo número de observações para todas as provas é de fundamental importância para as comparações que se seguirão.









Prova	Rec. mundial	Ano	Prob_exced
F 100m rasos	10.49	1988	0.18%
F 200m rasos	21.34	1988	0.29%
F 400m rasos	47.60	1985	0.20%
F 800m rasos	1:53.28	1983	0.24%
F 1500m rasos	3:50.07	2015	0.54%
F 5000m rasos	14:06.62	2008	0.23%
F 10000m rasos	29:01.03	2008	0.40%
F 100m barreiras	12.20	2016	0.50%
F 400m barreiras	51.46	2019	0.30%
M 100m rasos	9.58	2009	0.15%
M 200m rasos	19.19	2009	0.24%
M 400m rasos	43.03	2016	0.48%
M 800m rasos	1:40.91	2012	0.23%
M 1500m rasos	3:26.00	1998	0.16%
M 5000m rasos	12:35.36	2004	0.16%
M 10000m rasos	26:11.00	2005	0.17%
M 110m barreiras	12.80	2012	0.60%
M 400m barreiras	46.78	1992	0.25%

Vamos ao que interessa!

Vamos ao que interessa!



Usain Bolt x Michael Phelps

A metodologia descrita foi utilizada para comparar resultados em provas diferentes da natação, mas pode ser utilizada para comparar resultados de esportes diferentes.

A metodologia descrita foi utilizada para comparar resultados em provas diferentes da natação, mas pode ser utilizada para comparar resultados de esportes diferentes.

No caso, natação e atletismo.

A metodologia descrita foi utilizada para comparar resultados em provas diferentes da natação, mas pode ser utilizada para comparar resultados de esportes diferentes.

No caso, natação e atletismo.

Serão feitas comparações entre dois dos melhores resultados das carreiras dos dois atletas, com o objetivo de se verificar, com base no critério proposto, qual é o mais espetacular.

Usain Bolt:

Usain Bolt:

100 m rasos no Campeonato Mundial de Berlim (2009).

Usain Bolt:

100 m rasos no Campeonato Mundial de Berlim (2009).

Recorde mundial de 9s58 (recorde mundial até aquele ano: Usain Bolt, 9s69).

Michael Phelps:

Michael Phelps:

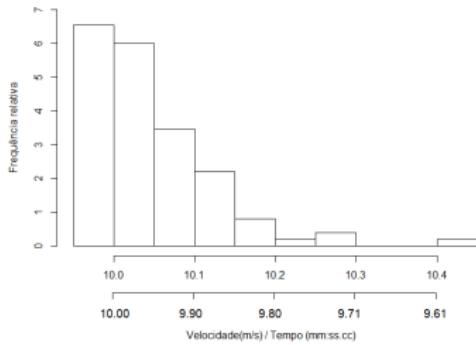
200 m medley no Campeonato Mundial de 2003

Michael Phelps:

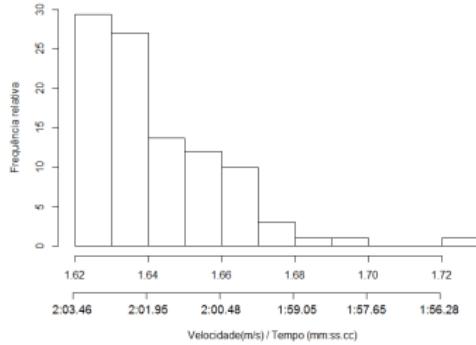
200 m medley no Campeonato Mundial de 2003

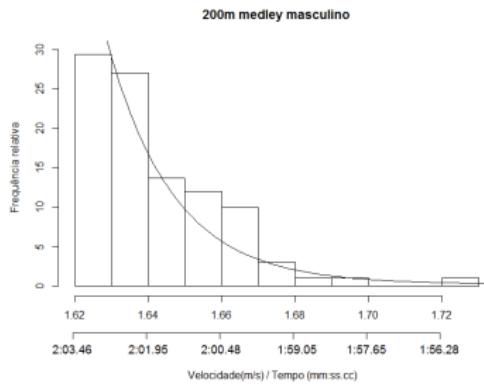
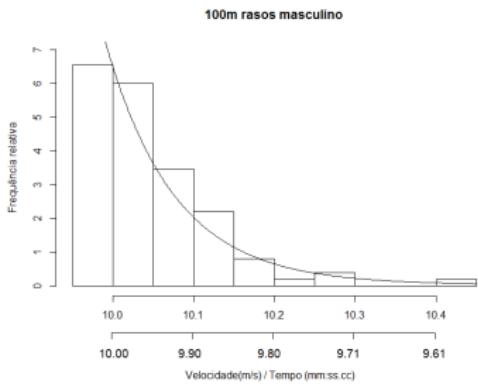
Recorde mundial de 1min56s04 (recorde mundial até aquele ano: Jani Sievinen, 1min58s16).

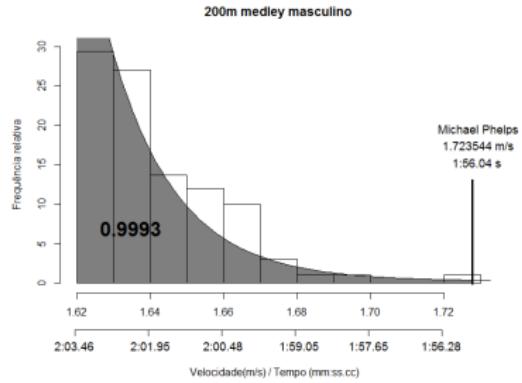
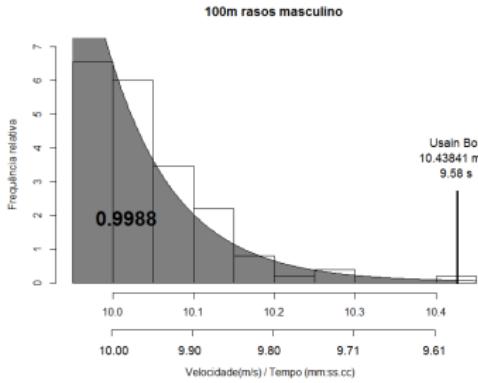
100m rasos masculino



200m medley masculino







No masculino, o recorde mais expressivo é o de Usain Bolt nos 100 m rasos.

No masculino, o recorde mais expressivo é o de Usain Bolt nos 100 m rasos.

No feminino, é o da americana Florence Griffith-Joyner, também nos 100 m rasos.

No masculino, o recorde mais expressivo é o de Usain Bolt nos 100 m rasos.

No feminino, é o da americana Florence Griffith-Joyner, também nos 100 m rasos.

São recordes antigos: o de Bolt dura 10 anos e o de Griffith-Joyner, 31.

No masculino, o recorde mais expressivo é o de Usain Bolt nos 100 m rasos.

No feminino, é o da americana Florence Griffith-Joyner, também nos 100 m rasos.

São recordes antigos: o de Bolt dura 10 anos e o de Griffith-Joyner, 31.

Será que estamos próximos de ver esses recordes serem quebrados?

Pergunta: essas comparações não poderiam ter sido feitas nos termos das comparações feitas utilizando dado do futebol?

Pergunta: essas comparações não poderiam ter sido feitas nos termos das comparações feitas utilizando dado do futebol?

Por exemplo, considerar o melhor desempenho registrado a cada ano, utilizar a distribuição de valor extremo generalizada para determinar a distribuição desses máximos, e calcular a probabilidade de o recorde ser superado?

Pergunta: essas comparações não poderiam ter sido feitas nos termos das comparações feitas utilizando dado do futebol?

Por exemplo, considerar o melhor desempenho registrado a cada ano, utilizar a distribuição de valor extremo generalizada para determinar a distribuição desses máximos, e calcular a probabilidade de o recorde ser superado?

Resposta: essa não seria a abordagem mais adequada, pois, pela evolução do esporte, esses máximos poderiam ter distribuições diferentes ao longo dos anos.

Qual é a chance de um determinado recorde ser batido em um futuro próximo?

Qual é a chance de um determinado recorde ser batido em um futuro próximo?

No masculino, quem mais chegou perto do recorde de Bolt, de 9s58, foi o americano Tyson Gay, com 9s69, em 2009.

Qual é a chance de um determinado recorde ser batido em um futuro próximo?

No masculino, quem mais chegou perto do recorde de Bolt, de 9s58, foi o americano Tyson Gay, com 9s69, em 2009.

Desde 2016, a melhor marca foi de outro americano, Christian Coleman, com 9s79 em 2018.

Qual é a chance de um determinado recorde ser batido em um futuro próximo?

No masculino, quem mais chegou perto do recorde de Bolt, de 9s58, foi o americano Tyson Gay, com 9s69, em 2009.

Desde 2016, a melhor marca foi de outro americano, Christian Coleman, com 9s79 em 2018.

No feminino, por muito tempo o recorde de Griffith-Joyner, de 10s49, ficou intocável.

Qual é a chance de um determinado recorde ser batido em um futuro próximo?

No masculino, quem mais chegou perto do recorde de Bolt, de 9s58, foi o americano Tyson Gay, com 9s69, em 2009.

Desde 2016, a melhor marca foi de outro americano, Christian Coleman, com 9s79 em 2018.

No feminino, por muito tempo o recorde de Griffith-Joyner, de 10s49, ficou intocável.

Até que em 2021 duas jamaicanas, Elaine Thompson-Herah (10.54) e Shelly-Ann Fraser-Pryce (10.60), chegaram mais perto do que nunca.

Qual é a chance de um determinado recorde ser batido em um futuro próximo?

No masculino, quem mais chegou perto do recorde de Bolt, de 9s58, foi o americano Tyson Gay, com 9s69, em 2009.

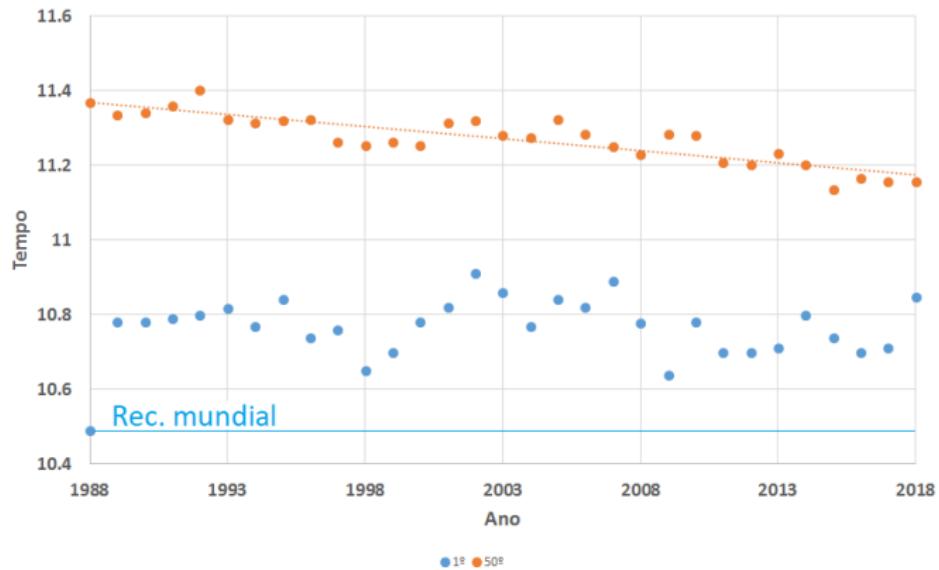
Desde 2016, a melhor marca foi de outro americano, Christian Coleman, com 9s79 em 2018.

No feminino, por muito tempo o recorde de Griffith-Joyner, de 10s49, ficou intocável.

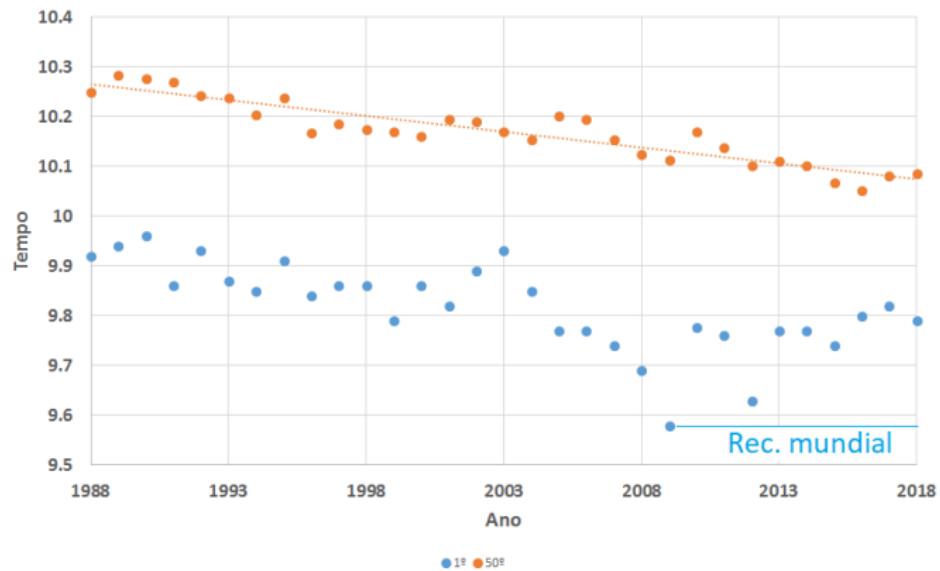
Até que em 2021 duas jamaicanas, Elaine Thompson-Herah (10.54) e Shelly-Ann Fraser-Pryce (10.60), chegaram mais perto do que nunca.

Devemos esperar que esses recordes sejam quebrados quando?

Evolução do 50º e do 1º tempos do ranking mundial dos 100 m feminino, ano a ano



Evolução do 50º e do 1º tempos do ranking mundial dos 100 m masculino, ano a ano



Ideia: apesar de não parecer que nos últimos anos os líderes do ranking mundial estão chegando perto do recorde, a distribuição geral dos dados indica que há, sim, evolução ao longo do tempo.

Ideia: apesar de não parecer que nos últimos anos os líderes do ranking mundial estão chegando perto do recorde, a distribuição geral dos dados indica que há, sim, evolução ao longo do tempo.

Chegará um momento em que essa evolução tornará inevitável a quebra do recorde.

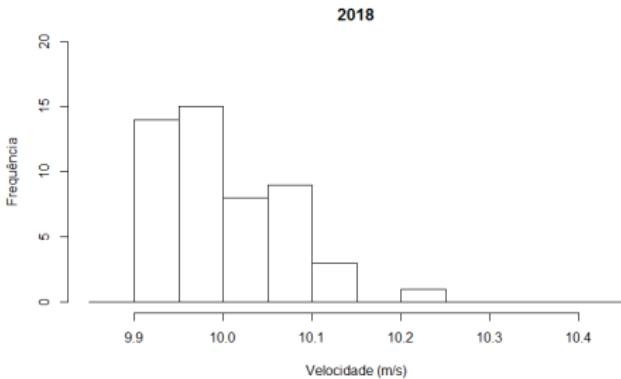
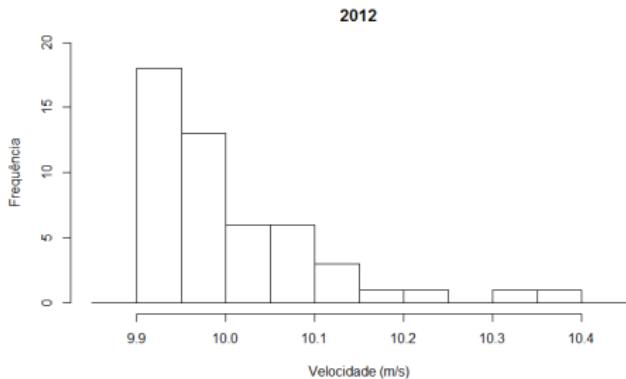
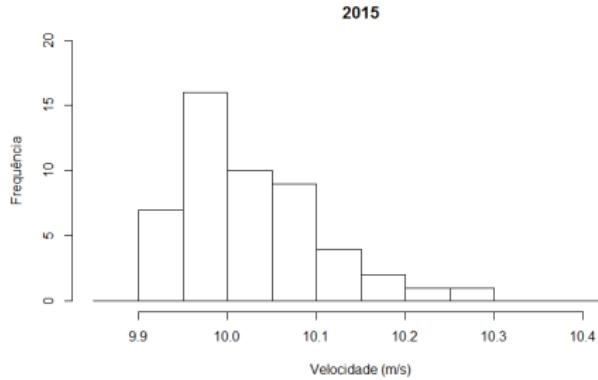
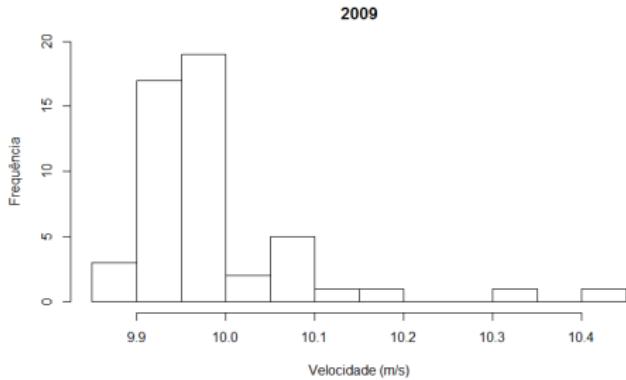
Ideia: apesar de não parecer que nos últimos anos os líderes do ranking mundial estão chegando perto do recorde, a distribuição geral dos dados indica que há, sim, evolução ao longo do tempo.

Chegará um momento em que essa evolução tornará inevitável a quebra do recorde.

Pelo histograma essa ideia pode ser mais clara.

Distribuições dos 50

tempos mais rápidos dos 100m rasos masculino em 2009, 2012, 2015 e 2018



Metodologia POT

Metodologia POT

A ideia aqui é considerar as marcas dos 50 primeiros colocados dos rankings mundiais das provas de 100 m e 200 m rasos, ano a ano. Pelo teorema de Pickands-Balkema-de Haan, as distribuições associadas possam ser bem aproximadas por GPDs.

Metodologia POT

A ideia aqui é considerar as marcas dos 50 primeiros colocados dos rankings mundiais das provas de 100 m e 200 m rasos, ano a ano. Pelo teorema de Pickands-Balkema-de Haan, as distribuições associadas possam ser bem aproximadas por GPDs.

Como a forma da distribuição pode mudar ao longo dos anos, propõe-se um modelo dinâmico bayesiano para os parâmetros de forma, ξ_t , e de escala, θ_t , da distribuição.

Metodologia POT

A ideia aqui é considerar as marcas dos 50 primeiros colocados dos rankings mundiais das provas de 100 m e 200 m rasos, ano a ano. Pelo teorema de Pickands-Balkema-de Haan, as distribuições associadas possam ser bem aproximadas por GPDs.

Como a forma da distribuição pode mudar ao longo dos anos, propõe-se um modelo dinâmico bayesiano para os parâmetros de forma, ξ_t , e de escala, θ_t , da distribuição.

$$\xi_t | \xi_{t-1} \sim Beta^* \left(\alpha \frac{\xi_{t-1}}{1 + \xi_{t-1}/k}, \alpha \right), \quad \theta_t | \theta_{t-1} \sim Gama^* \left(\frac{\xi_t}{\beta}, \beta \right)$$

com $k = \theta_{t-1} / (\max\{x_1, \dots, x_{50}\} - u)$.

As distribuições $Beta^*$ e $Gama^*$ são, respectivamente, $Beta$ modificada e $Gama$ deslocada. O suporte da distribuição $Beta^*$ é $[-k, 0]$ e da $Gama^*$ é $[k, \infty)$. Tais distribuições são escolhidas de modo que $\mathbb{E}(\xi_t | \xi_{t-1}) = \xi_{t-1}$ e $\mathbb{E}(\theta_t | \theta_{t-1}) = \theta_{t-1}$. Os hiperparâmetros α e β determinam a variabilidade das distribuições.

As distribuições $Beta^*$ e $Gama^*$ são, respectivamente, $Beta$ modificada e $Gama$ deslocada. O suporte da distribuição $Beta^*$ é $[-k, 0]$ e da $Gama^*$ é $[k, \infty)$. Tais distribuições são escolhidas de modo que $\mathbb{E}(\xi_t | \xi_{t-1}) = \xi_{t-1}$ e $\mathbb{E}(\theta_t | \theta_{t-1}) = \theta_{t-1}$. Os hiperparâmetros α e β determinam a variabilidade das distribuições.

A distribuição $Beta^*$ é assim escolhida para que o limite superior da GPD seja finito.

As distribuições $Beta^*$ e $Gama^*$ são, respectivamente, $Beta$ modificada e $Gama$ deslocada. O suporte da distribuição $Beta^*$ é $[-k, 0]$ e da $Gama^*$ é $[k, \infty)$. Tais distribuições são escolhidas de modo que $\mathbb{E}(\xi_t | \xi_{t-1}) = \xi_{t-1}$ e $\mathbb{E}(\theta_t | \theta_{t-1}) = \theta_{t-1}$. Os hiperparâmetros α e β determinam a variabilidade das distribuições.

A distribuição $Beta^*$ é assim escolhida para que o limite superior da GPD seja finito.

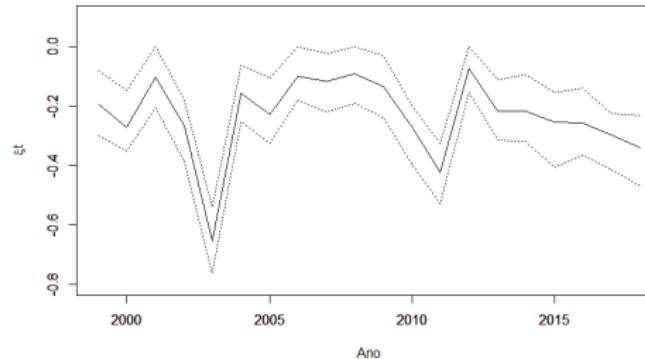
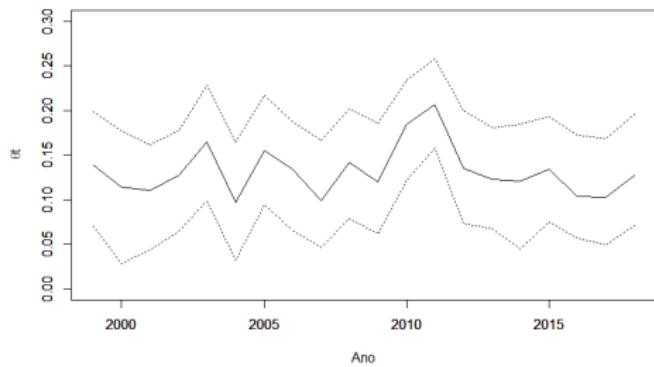
O limiar da GPD, em cada ano, corresponde ao 50º tempo do ranking mundial.

As distribuições $Beta^*$ e $Gama^*$ são, respectivamente, $Beta$ modificada e $Gama$ deslocada. O suporte da distribuição $Beta^*$ é $[-k, 0]$ e da $Gama^*$ é $[k, \infty)$. Tais distribuições são escolhidas de modo que $\mathbb{E}(\xi_t | \xi_{t-1}) = \xi_{t-1}$ e $\mathbb{E}(\theta_t | \theta_{t-1}) = \theta_{t-1}$. Os hiperparâmetros α e β determinam a variabilidade das distribuições.

A distribuição $Beta^*$ é assim escolhida para que o limite superior da GPD seja finito.

O limiar da GPD, em cada ano, corresponde ao 50º tempo do ranking mundial.

As distribuições a posteriori de ξ_t e θ_t são obtidas através do método MCMC, pelo algoritmo de Metropolis-Hastings.



Dessa forma, de acordo com o comportamento e a evolução dos dados (rankings mundiais top 50 nos últimos 20 anos) e dos parâmetros, podemos simular da distribuição preditiva dos dados para os anos futuros.

Dessa forma, de acordo com o comportamento e a evolução dos dados (rankings mundiais top 50 nos últimos 20 anos) e dos parâmetros, podemos simular da distribuição preditiva dos dados para os anos futuros.

Executamos simulações para milhares de cenários, e calculamos em quantas delas o recorde mundial é quebrado. Assim, temos uma probabilidade estimada de o recorde mundial cair em determinado ano.

Feminino						Masculino					
100 m			200 m			100 m			200 m		
Ano	Estim. do 1o tempo	Prob. rec. mundial	Ano	Estim. do 1o tempo	Prob. rec. mundial	Ano	Estim. do 1o tempo	Prob. rec. mundial	Ano	Estim. do 1o tempo	Prob. rec. mundial
2022	10.65	1.50%	2022	21.52	2.05%	2022	9.77	0.73%	2022	19.56	0.58%
2023	10.63	3.91%	2023	21.50	5.25%	2023	9.75	1.99%	2023	19.53	1.59%
2024	10.60	8.14%	2024	21.47	10.75%	2024	9.73	4.01%	2024	19.49	3.21%
2025	10.56	20.76%	2025	21.42	19.45%	2025	9.70	10.54%	2025	19.45	8.43%
2026	10.51	36.21%	2026	21.35	43.12%	2026	9.66	18.21%	2026	19.39	14.57%

Essa mesma análise, feita em 2019, apontava que a probabilidade de o recorde dos 100 m rasos feminino cair em um prazo de cinco anos era baixíssima, menos de 2%.

Essa mesma análise, feita em 2019, apontava que a probabilidade de o recorde dos 100 m rasos feminino cair em um prazo de cinco anos era baixíssima, menos de 2%.

No entanto, o cenário mudou, pois duas atletas ameaçaram perigosamente o recorde mundial recentemente, fazendo a probabilidade aumentar, e muito.

Essa mesma análise, feita em 2019, apontava que a probabilidade de o recorde dos 100 m rasos feminino cair em um prazo de cinco anos era baixíssima, menos de 2%.

No entanto, o cenário mudou, pois duas atletas ameaçaram perigosamente o recorde mundial recentemente, fazendo a probabilidade aumentar, e muito.

O esporte é dinâmico!

Principais referências bibliográficas

-  Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. L. (2004). *Statistics of extremes: theory and applications*. John Wiley Sons.
-  Einmahl, J. H., Magnus, J. R. (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association*, 103(484), pp. 1382-1391.
-  Gomes, D. T., Henriques-Rodrigues, L. (2019). Swimming performance index based on extreme value theory. *International Journal of Sports Science & Coaching*, 14(1), pp. 51-62.
-  Sumpter, D. (2016). *Soccermatics: mathematical adventures in the beautiful game*. Bloomsbury Publishing.