

Modelação estatística de valores extremos: probabilidades de excedência, quantis extremais, limite superior do suporte e níveis de retorno no lançamento do disco do atletismo

Domingos Silva¹, Frederico Caeiro², Manuela Oliveira³

Resumo: A teoria de valores extremos é um ramo da Estatística especializada em lidar com eventos extremos e raros cuja probabilidade de ocorrência é muito reduzida, mas quando ocorrem têm um impacto muito significativo na sociedade. Neste estudo, usamos o método dos máximos de blocos anuais, a distribuição generalizada de valores extremos, a estimação por máxima verossimilhança e a otimização pelo algoritmo de Nelder-Mead para estimar parâmetros de extremais relativos ao lançamento do disco do atletismo de homens e mulheres, no período de 1960-2016. Os principais resultados apontam para a existência de um parâmetro de forma negativo, representativo da cauda superior leve e finita. Neste caso, o limite superior do suporte é de 74.32m para homens e 77.37m para mulheres. A probabilidade de obter um novo máximo mundial é de 0.334% nos homens e 0.208% nas mulheres. Atendendo às condições atuais e à metodologia seguida neste estudo, com base no intervalo com 95% de confiança, o nível de retorno a 50 anos contém um novo recorde do mundo para homens e mulheres. A validação destes resultados necessita de ser feita, recorrendo quer ao aumento do tamanho da amostra quer a outras metodologias de valores extremos, ou ambas.

Palavras-chave: teoria de valores extremos; distribuição generalizada de valores extremos; probabilidade de excedência; quantis extremais; limite superior do suporte; níveis de retorno; lançamento do disco.

Abstract: The extreme value theory is a branch of Statistics, that addresses extreme and rare events that have a very low occurrence probability and yet have a very significant impact if they indeed occur. In this study, we used the annual block maximum method, the generalized extreme value distribution, the maximum likelihood estimation, and the optimization by Nelder-Mead algorithm to estimate the extremal parameters relatives to track and field discus throw (men and women), in the period between 1960 and 2016. The main results indicate the existence of a negative shape parameter, representative of a short and finite right tail. In this case, the right endpoint is 74.32m for men and 77.37m for women. The exceedance probability to have a new world record is 0.334% for men and 0.208% for women. In the current conditions, according to this methodology, based on 95% confidence interval, a new world record for men and women may occur every 50-years. The validation of this results should be made, either by increasing the sample size, by using other extreme value methodologies, or both.

Key-words: extreme value theory; generalized extreme value distribution; exceedance probability; extremal quantile; right endpoint; return levels; discus throw.

¹ Universidade de Évora, Escola de Ciências e Tecnologia & CIMA, Depto. de Matemática e-mail: domingosjsilva@gmail.com.

² Universidade Nova de Lisboa, Escola de Ciências e Tecnologia & CIMA, Depto. de Matemática.

³ Universidade de Évora, Escola de Ciências e Tecnologia & CIMA, Depto. de Matemática.

1 Introdução

O lançamento do disco é uma das mais antigas provas desportivas. Data da antiguidade e ficou imortalizado pela sua representação na estátua do Discóbolo, uma obra da autoria de Míron (Séc. V a.C.). Os registos existem desde 1892 para homens e 1912 para mulheres (IAAF, 2015). Em termos de presença olímpica, foi introduzido nos primeiros Jogos Olímpicos da Era Moderna, em 1896 (Atenas) para os homens, e em 1928 (Amsterdão) para as mulheres. Contudo, os registos oficiais da *International Amateur Athletic Federation* (IAAF) consideram apenas as marcas obtidas a partir de 1912 para homens e de 1923 para mulheres.

Tabela 1: Evolução de alguns recordes.

Atletas	Marca	Ano	Observações
+ HOMENS			
Harald Andersson-Arbin (SUE)	26.20m	1892	1º recorde do mundo antes da IAAF
James Duncan (EUA)	47.58m	1912	1º recorde do mundo Era IAAF
Virgilijus Alekna (LIT)	69.89m	2004	Atual recorde olímpico (Atenas)
Jurgen Schult (RDA)	74.08m	1986	Atual recorde do mundo
Francisco Belo (POR)	62.01m	2017	Atual recorde de Portugal
Ronald Julião (BRA)	65.55m	2013	Atual recorde do Brasil
+ MULHERES			
Margorie Hilas (EUA)	20.72m	1912	1º recorde do mundo antes da IAAF
Yvonne Tembouret (FRA)	27.39m	1923	1º recorde do mundo Era IAAF
Martina Hellmann (RDA)	72.30m	1988	Atual recorde olímpico (Seul)
Gabriele Reinsch (RDA)	76.80m	1988	Atual recorde do mundo
Teresa Machado (POR)	65.40m	1998	Atual recorde de Portugal
Andressa de Moraes (BRA)	64.68m	2017	Atual recorde do Brasil

URL: https://en.wikipedia.org/wiki/Discus_throw

URL: https://en.wikipedia.org/wiki/National_records_in_athletics

IAAF (2015)

Desde que existem registos, o recorde do mundo dos homens aumentou cerca de 183%, e o das mulheres cerca de 271%! Porém, há mais de 30 anos nos homens e há cerca de 30 anos nas mulheres que não se consegue um novo máximo mundial (Tabela 1). A razão do recorde do mundo das mulheres ser mais alto do que o dos homens deve-se ao peso e dimensões do engenho. Para os homens, o disco pesa 2kg (219-221mm de diâmetro e 44-46mm de espessura), e para as mulheres 1kg (180-182mm de diâmetro e 37-39mm de espessura). O lançamento é feito dentro de um círculo com 2.5 metros de diâmetro, com um anteparo de 2cm, onde o pé apenas pode tocar na sua face interior. O lançamento é feito com a extensão vigorosa da mão e do cotovelo, depois do atleta rodar rapidamente sobre si mesmo. O atleta, se pretender, pode usar uma substância (regulamentar) para ajudar no agarre do disco, bem como um cinturão para proteger a coluna. São fatores importantes para um bom lançamento: deslizamento e velocidade de deslizamento (avanço dos apoios), posição do corpo no deslizamento e no momento da saída do engenho, ângulo de saída, altura de saída e velocidade de saída (Schmolinsky, 1982).

Para ajudar a compreender este fenómeno, existe a Teoria de Valores Extremos (EVT, *Extreme Value Theory*), por se tratar de uma potente e robusta técnica de modelação estatística do comportamento assintótico de eventos extremos e raros, baseada no teorema dos tipos extremos de Fisher-Tippett-

Gnedenko. Tem como objetivos descrever, quantificar e estimar uma possível distribuição limite para o comportamento assintótico das estatísticas ordinais extremas da amostra, composta por variáveis aleatórias independentes e identicamente distribuídas, capaz de modelar as caudas da distribuição, particularmente o máximo (ou o mínimo). Portanto, a EVT lida com a ocorrência de acontecimentos atípicos que são mais extremos do que quaisquer outros já observados, estimando a sua probabilidade de ocorrência mesmo quando não existem dados, ou se existem são escassos. Nesta medida, assume particular importância a inferência para além da amostra disponível. Historicamente, as primeiras de aplicações foram nas áreas do ambiente e da força/resistência dos materiais. Na atualidade, a utilização da EVT difunde-se por praticamente todos os setores da vida e todas as ciências. No caso do desporto, especificamente no domínio do lançamento do disco do atletismo, apenas se conhece o trabalho de Einmah & Magnus (2006, 2008). Dentro da EVT existem vários métodos. Um dos mais conhecidos é o método dos máximos de blocos. Este método lida com a observação máxima registada num certo período de tempo (e.g., ano). Não obstante algumas desvantagens (e.g., exclusão de observações extremas de um bloco, mesmo sendo mais extremas do que as observações máximas de outros blocos), a principal vantagem reside no facto de ser natural coletar os dados máximos. E tratando-se de desporto, especificamente quando se pretende um novo recorde do mundo no lançamento do disco, tal ganha ainda mais importância. Com este estudo pretendemos usar a teoria de valores extremos e o método dos máximos de blocos anuais para estimar alguns parâmetros extremos, tais como, a probabilidade de excedência $P[X > x]$, o limite superior do suporte x^F (quando $\gamma < 0$), os quantis extremos q_{1-p} e os níveis de retorno $U(t)$.

2 Dados e Métodos

Os dados foram obtidos em três Websites: (i) https://en.wikipedia.org/wiki/Discus_throw para dados de 1968 a 2015, (ii) <http://www.all-athletics.com/en-us/all-time-lists> para obter o melhor registo de 2016, e (iii) <http://trackfield.brinkster.net/> para os registos anteriores a 1968. Igualmente, em certas situações consultou-se o compêndio da IAAF (2015). Apenas foram aceites as marcas homologadas pela IAAF. Neste estudo, utilizamos as observações máximas de 1960-2016. Assim, as amostras de homens e mulheres são formadas por 57 observações máximas. Por recurso à EVT iremos usar o método dos máximos de blocos anuais, a estimação por máxima verosimilhança e a otimização pelo algoritmo de Nelder-Mead. Todos os cálculos foram efetuados no programa R.

3 Resultados e Discussão

3.1 Análise Exploratória

A Tabela 2 mostra as estatísticas descritivas básicas das amostras. O valor máximo corresponde ao atual recorde do mundo que nos homens foi obtido em 1986 por Jurgen Schult (RDA) e nas mulheres em 1988 por Gabriele Reinsch (RDA). Por sua vez, o valor mínimo nos homens foi obtido em 1960 por Rink Babka (EUA) e nas mulheres em 1962 por Doris Muller (RDA). Nas duas amostras, a média está mais próxima do valor máximo do que do mínimo, embora mais próxima do quantil 0.25 do que do quantil 0.75. Igualmente, a mediana é mais alta do que a média. O coeficiente de variação (CV) revela fraca dispersão nas duas amostras (homens: CV=4.2%; mulheres: CV=6.9%).

Tabela 2: Caracterização das amostras.

	n	ξ	s	Min	Máx	Q1	Q2	Q3	Skewness	Kurtosis
Homens	57	69.29	2.941	59.91	74.08	68.40	69.91	71.14	-1.316	1.732
Mulheres	57	68.04	4.668	56.39	76.80	66.81	68.96	70.72	-0.915	0.253

A Figura 1 apresenta o diagrama de extremos e quartis e o histograma com a curva de densidade referentes ao salto em comprimento. Percebe-se, como seria de esperar, e conforme as estimativas da assimetria e da curtose apresentadas na Tabela 2, que nas duas amostras existe uma assimetria à esquerda.

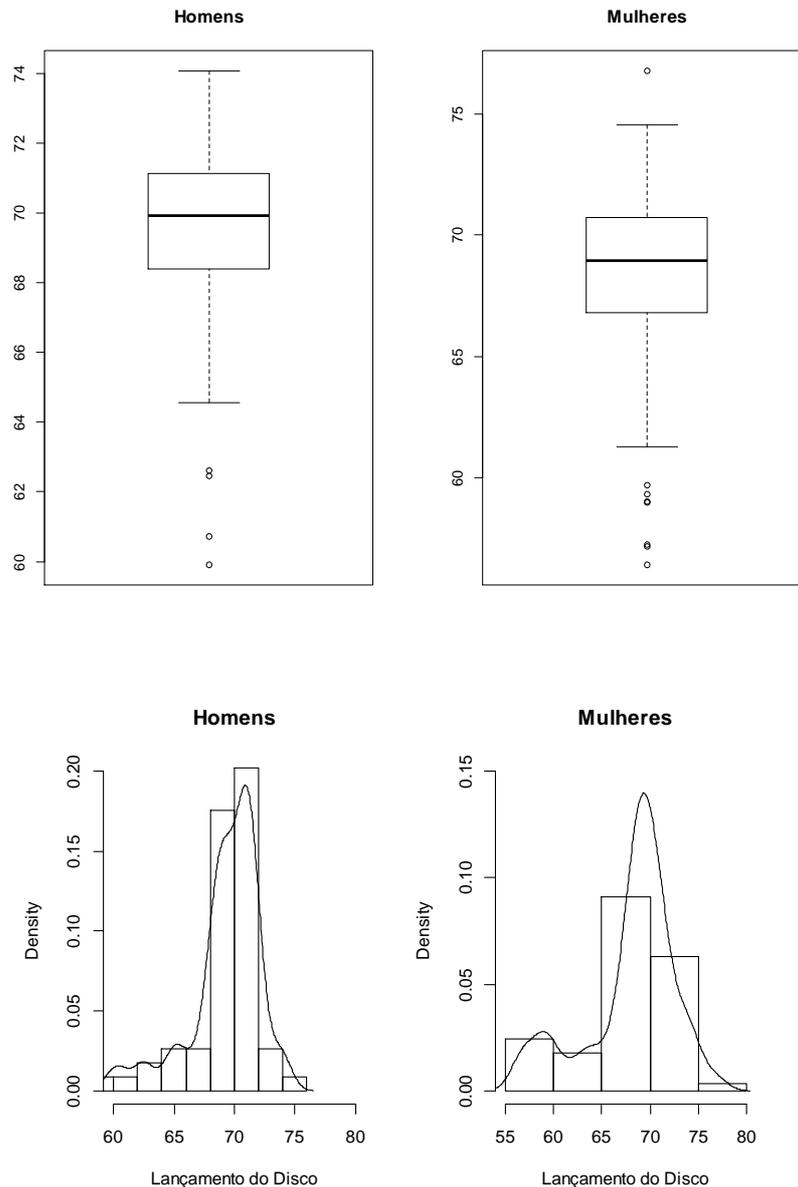


Figura 1: Diagramas de extremos e quartis e histogramas com a curva de densidade, das observações referentes ao lançamento do disco, de homens e mulheres, no período de 1960-2016.

3.2 Aleatoriedade

A aleatoriedade foi verificada pelo Turning Point Test (Moore & Wallis, 1943). Seja $X = \{X_1, \dots, X_n\}$, uma sequência aleatória. Diz-se que existe “ponto de viragem” (turning point test) no tempo i , $1 < i < n$, se $X_{i-1} < X_i$ e $X_i > X_{i+1}$ ou se

$X_{i-1} > X_i$ e $X_i < X_{i+1}$. Sendo T o número de “pontos de viragem” da sequência $X_i, i=1, \dots, n$, suficientemente grande, a probabilidade de ocorrer um “ponto de viragem” no tempo i é de $2/3$. O valor esperado de T é dado por $\mu_T := E(T) = \frac{2n-4}{3}$ e a variância por $\sigma_T^2 := Var(T) = \frac{16n-29}{90}$.

Brockwell & Davis (2002) mostraram que para uma amostra aleatória $\{X_1, \dots, X_n\}$, com n grande, $T \sim N(\mu_T, \sigma_T^2)$. A decisão de rejeitar “ H_0 : a série é aleatória”, a um certo nível α , ocorre se $z = |T - \mu_T| / \sigma_T > \Phi_{1-\alpha/2}$, onde $\Phi_{1-\alpha/2}$ é o quantil $(1-\alpha/2)$ da distribuição Normal padrão.

Considere-se $\alpha=5\%$. Para uma hipótese bilateral, nas duas amostras, dado que $z=-1.2888$; $p=0.1975$ nos homens e $z=-1.1706$; $p=0.2418$, tal sugere a presença de aleatoriedade nas amostras, não-rejeitando a hipótese nula. Ou seja, as observações são i.i.d.’s.

3.3 Tendência

A informação na Figura 2 sugere, nas duas amostras, a ausência de tendência significativa das performances no lançamento do disco ao longo dos anos, embora nos anos iniciais parece existir uma ligeira tendência linearmente crescente, especialmente entre 1960-1971 nos homens e entre 1966-1976 nas mulheres.

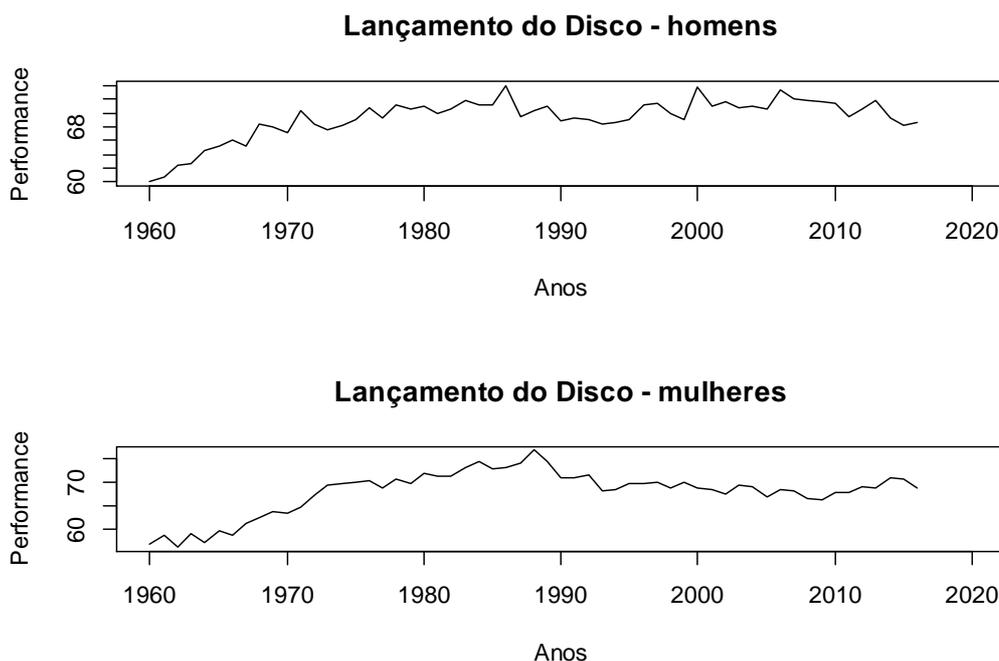


Figura 2: Tendência das observações no lançamento do disco (homens e mulheres) ao longo do tempo: 1960-2016.

3.4 Análise clássica de valores extremos

Seja $\{X_1, X_2, \dots, X_n\}$ uma amostra aleatória (a.a.) formada por uma sequência de variáveis aleatórias (v.a.’s) independentes e identicamente distribuídas (i.i.d.’s) ou fracamente estacionárias, com função de distribuição (f.d.) comum $F(\cdot)$, desconhecida. Denota-se por $\{X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}\}$ as estatísticas de ordinais (e.o.’s) ascendentes associadas à amostra $\{X_1, X_2, \dots, X_n\}$, de tal forma que $X_{1:n}$ representa o mínimo e $X_{n:n}$ o máximo (i.e., o atual recorde do mundo).

Uma vez que neste estudo a variável de interesse é o máximo, então,

$$M_n = X_{n:n} = \max\{X_1, X_2, \dots, X_n\} \xrightarrow[n \rightarrow \infty]{P} x^F \quad [1]$$

Sendo que a f.d. exata do Mn (também desconhecida) é dada por,

$$F_{M_n}(x) = P[M_n \leq x] = P[X_1 \leq x, \dots, X_n \leq x] = P[X_1 \leq x] \times \dots \times P[X_n \leq x] \\ = F(x) \times \dots \times F(x) = \{F(x)\}^n = F^n(x), \quad x \in \mathfrak{R}, \quad n \in \mathbb{N} \quad [2]$$

que converge para zero quando $x < x^F$ e converge para 1 quando $x \geq x^F$, em que

$$x^F := \sup\{x \in \mathfrak{R} : F(x) < 1\} \quad [3]$$

é o limite superior do suporte do modelo F subjacente aos dados disponíveis.

Sendo F desconhecido, a expressão $P[M_n \leq x] = F^n(x)$ não tem grande utilidade prática. Além disso, o limite de $F^n(x)$ é degenerado, uma vez que está concentrado no ponto definido em [3], ou seja,

$$\lim_{n \rightarrow \infty} F^n(x) = \begin{cases} 0, & \text{se } F(x) < 1 \\ 1, & \text{se } F(x) = 1 \end{cases} \quad [4]$$

comprometendo qualquer inferência para o máximo, onde é necessário que este tenha um comportamento assintótico não-degenerado. Por esta razão, torna-se necessário normalizar o M_n de modo a obter uma possível lei limite não-degenerada, à semelhança do que acontece com o Teorema do Limite Central (TLC) para as médias (ou somas), onde o máximo linearmente normalizado converge para uma v.a. não-degenerada. Assim, recorre-se a uma sucessão de constantes reais $a_n > 0$ e b_n ($n \in \mathbb{N}$), tal que $M_n^* = \frac{M_n - b_n}{a_n}$, resulta do facto da distribuição do máximo, por si só, convergir para uma f.d. degenerada no limite superior do suporte, x^F . Portanto, analogamente ao TLC, o máximo normalizado terá a forma,

$$\frac{M_n - b_n}{a_n},$$

com as sequências reais $\{a_n\}_{n \geq 1}$, $a_n > 0$ (cujo efeito é a estabilização da escala) e $\{b_n\}_{n \geq 1}$, $b_n \in \mathfrak{R}$ (que visa estabilizar a localização). Assim, tem-se o Teorema dos Tipos Extremais de Fisher-Tippett-Gnedenko (Fisher & Tippett, 1928; Gnedenko, 1943), tal que

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{d} G(x), \quad [5]$$

sendo $G(x)$ uma f.d. não-degenerada, então G é do mesmo tipo de uma das seguintes distribuições de valores extremos (I, II e III) – Gumbel, Fréchet e Weibull (Fréchet, 1927; Gumbel, 1935, 1958; Weibull, 1939, 1951), suportadas pelo Teorema dos Tipos Extremais e dadas respectivamente por,

$$\text{Tipo I : } G(x) = \exp(-\exp(-x)) \quad -\infty < x < \infty \quad [6]$$

$$\text{Tipo II : } G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \quad \alpha > 0 \end{cases} \quad [7]$$

$$\text{Tipo III : } G(x) = \begin{cases} \exp(-(-x)^\alpha) & x < 0, \quad \alpha > 0 \\ 1 & x \geq 0 \end{cases} \quad [8]$$

von Mises (1936, 1954) e Jenkinson (1955) unificaram as distribuições [6], [7] e [8] numa forma paramétrica única, a designada distribuição generalizada de

valores extremos (GEV, *Generalized Extreme Value*), que surge como distribuição limite do valor máximo de uma sequência de v.a.'s i.i.d.'s, dependente unicamente do parâmetro de forma. A estimativa dos parâmetros de localização (λ), escala (δ) e forma (γ) da distribuição GEV, com $x_+ = \max(x, 0)$, $\lambda \in \mathfrak{R}$, $\delta > 0$ e $\gamma \in \mathfrak{R}$, obtém-se através,

$$GEV_\gamma(x; \lambda, \delta) = \left\{ \begin{array}{l} \exp\left\{-\left[1 + \gamma\left(\frac{x-\lambda}{\delta}\right)\right]_+^{-\frac{1}{\gamma}}\right\}, \quad x \in \mathfrak{R}, 1 + \gamma\left(\frac{x-\lambda}{\delta}\right) > 0 \quad \text{se } \gamma \neq 0 \\ \exp\left\{-\exp\left[-\frac{x-\lambda}{\delta}\right]\right\}, \quad x \in \mathfrak{R}, \quad \text{se } \gamma = 0 \end{array} \right\} \quad [9]$$

O parâmetro de forma γ (ou índice de valores extremos), determina o comportamento da cauda direita da distribuição. Se $\gamma=0$ ($\gamma \rightarrow 0$), tem-se o domínio de atração de Gumbel, com distribuições de cauda direita do tipo exponencial e com limite superior do suporte de F finito ($x^F < \infty$) ou infinito ($x^F = \infty$); se $\gamma > 0$, tem-se o domínio de atração de Fréchet ($\gamma = 1/\alpha > 0$) com distribuições de cauda direita pesada do tipo polinomial negativo e com limite superior do suporte de F infinito ($x^F = \infty$); se $\gamma < 0$, tem-se o domínio de atração de max-Weibull ($\gamma = -1/\alpha < 0$) com distribuições de cauda direita leve e com limite superior do suporte de F finito ($x^F < \infty$). Mais detalhes podem ser observados em Coles (2001, secção 3).

Portanto, é conveniente transformar a equação [5] em,

$$P\{M_n \leq x\} \approx G(x - b_n)/a_n = G^*(x) \quad [10]$$

Assim, a distribuição resultante, $G^*(x)$, é também membro da família da distribuição GEV e permite que as constantes normalizadoras $a_n > 0$ e b_n ($n \in \mathbb{N}$) e os parâmetros extremais (λ, δ, γ) possam ser estimados simultaneamente.

3.5 Método dos máximos de blocos

O método dos máximos de blocos está relacionado com a distribuição assintótica não-degenerada do máximo, $M_n = \max(X_1, \dots, X_n)$, linearmente normalizado, $(M_n - b_n)/a_n$. A amostra de tamanho n (X_1, \dots, X_n) é dividida em k blocos (neste estudo, em k anos) de igual dimensão m ($M_{n,1m}, M_{n,2m}, \dots, M_{n,km}$). Recordamos que neste estudo, este método foi aplicado à marca do lançamento do disco mais elevada obtida em cada ano, no período de 1960-2016 (portanto, $k=57$).

3.6 Estimação de parâmetros

Na ausência de tendência e com $\hat{\gamma} \neq 0$, como é o caso do presente estudo (Tabela 3), a função log-verosimilhança para a distribuição [9], define-se da seguinte forma,

$$\lambda(\lambda, \delta, \gamma | x_1, \dots, x_k) = -k \log \delta - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^k \log \left[1 + \gamma \left(\frac{x_i - \lambda}{\delta}\right)\right]_+ - \sum_{i=1}^k \left[1 + \gamma \left(\frac{x_i - \lambda}{\delta}\right)\right]_+^{-\frac{1}{\gamma}}, \quad [11]$$

$$\text{com a restrição } 1 + \gamma \left(\frac{x_i - \lambda}{\delta}\right) > 0, \quad i = 1, \dots, k, \quad [12]$$

onde k é o número de máximos de blocos, $\{M_{n,1}, M_{n,2}, \dots, M_{n,k}\}$. Com a substituição habitual quando $\hat{\gamma} = 0$.

O estimador de máxima verosimilhança (ML, *maximum likelihood*) maximiza as funções de verosimilhança e de log-verosimilhança, uma vez que a função logarítmica é monótona. Se as condições assintóticas são válidas, a distribuição dos parâmetros estimados $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$ tem distribuição normal multivariada com média e matriz de variância-covariância igual ao inverso da matriz de informação observada. Todavia, uma vez que não há solução analítica para obter as estimativas de ML para os parâmetros da distribuição GEV, terão que ser usadas técnicas iterativas de otimização numérica. Neste estudo, usamos o método de Nelder-Mead (Nelder & Mead, 1965), o qual considera a restrição [12].

Smith (1985) estabeleceu que o comportamento limite na estimação por ML depende do valor do parâmetro de forma γ , ou seja,

- quando $\hat{\gamma} < -1$, geralmente os estimadores de ML são inexistentes.
- quando $\hat{\gamma} \leq -0.5$, corresponde a distribuições com cauda superior muito leve.
- quando $-1 < \hat{\gamma} < -0.5$, os estimadores de ML podem ser aplicáveis e os parâmetros são geralmente obtidos, mas não são nem consistentes, nem eficientes, nem normalmente assintóticos.
- quando $\hat{\gamma} > -0.5$, os estimadores de ML são regulares, i.e., consistência, eficiência, invariância e normalidade assintótica.

Os erros-padrão (se) para $\theta = (\lambda, \delta, \gamma)$, foram calculados pela matriz de informação de Fisher. Detalhes sobre este procedimento podem ser observados em Coles (2001, p.32). A estimativa do intervalo com $100(1-\alpha)\%$ de confiança, foi obtida por,

$$IC_{100(1-\alpha)\%}(\theta) = \hat{\theta} \pm z_{1-\alpha/2} \times se(\hat{\theta}) \quad [13]$$

A Tabela 3 apresenta as estimativas dos parâmetros da distribuição GEV, pelo método de estimação de máxima verosimilhança.

Nas duas amostras a estimativa do parâmetro de forma foi negativo, representativo da presença de cauda superior leve e finita, homens: $\hat{\gamma} = -0.555066$ (0.069), mulheres: $\hat{\gamma} = -0.4729238$ (0.061). À semelhança do presente estudo, e também com lançadores do disco, Einmah & Magnus (2006, 2008) apresentam estimativas negativas do índice de valores extremos para homens $\hat{\gamma} = -0.23$ e para mulheres $\hat{\gamma} = -0.16$.

Nas duas amostras, a estimativa do $IC_{95\%}(\gamma)$ situa-se no lado negativo, sugerindo por um lado que a sequência de dados é melhor descrita por um parâmetro de forma negativo (correspondendo à distribuição de Weibull na notação original), e por outro que o modelo Gumbel não deverá ser considerado. Contudo, quisemos verificar esta possibilidade comparando o desempenho dos dois modelos (GEV vs Gumbel), tendo para tal usado o teste de razão de verosimilhanças (TRV), baseado na estatística *deviance* (D). Assim, considere-se a função log-verosimilhança maximizada (NLLH, *negative log-likelihood*) (com $NLLH = -\log(l)$). Sendo a *deviance* dada por, $D = 2 \times (w - z)$, onde w a NLLH=156.481 (homens) e NLLH=178.9214 (mulheres) do modelo Gumbel e z é a NLLH=136.0865 (homens) e NLLH=165.0294 (mulheres) do modelo GEV. Rejeita-se $H_0 : \gamma = 0$ se $D \geq \chi^2_{\nu; 1-\alpha}$ (ν , são os graus de liberdade da distribuição quiquadrado, dados pelo módulo da diferença entre os graus de

liberdade dos dois modelos). Assim, a um nível $\alpha=0.05$, $\chi_{1;0.95}^2 = 3.841459$, confirmamos a rejeição do modelo Gumbel, pois $D = 40.789 > 3.841$ (homens) e $D = 27.784 > 3.841$ (mulheres).

Tabela 3: Estimativa pontual, erro-padrão e intervalo com 95% de confiança para os parâmetros de localização, escala e forma, com base na distribuição GEV, segundo a estimação por máxima verosimilhança e o método de otimização de Nelder-Mead.

	$\hat{\lambda}$	$\hat{\delta}$	$\hat{\gamma}$
Homens	68.663346 (0.445) (67.790; 69.535)	3.139214 (0.329) (2.495; 3.783)	-0.555066 (0.069) (-0.690; -0.420)
Mulheres	66.8302549 (0.706) (65.445; 68.215)	4.9835719 (0.503) (3.998; 5.969)	-0.4729238 (0.061) (-0.593; -0.353)

Estimados os parâmetros da distribuição GEV $(\lambda, \delta, \gamma)$, estamos em condições de obter as probabilidades de excedência de um nível x elevado $P[X > x]$, o limite superior do suporte (uma vez que $\hat{\gamma} < 0$), os quantis extremais \hat{q}_{1-p} e os níveis de retorno $U(t)$ para alguns períodos de retorno $t(U)$.

Sabendo que o máximo da amostra $\{X_i\}_{i=1,\dots,n}$ corresponde ao atual recorde do mundo de 74.08m de Jurgen Schult (RDA) em 1986 e de 76.80m de Gabriele Reinsch (RDA) em 1988, será interessante conhecer, dado o atual estado de arte, a probabilidade destas marcas serem ultrapassadas. Neste estudo, a probabilidade de excedência é dada diretamente pela função de cauda da distribuição GEV estimada,

$$P(X > x) := 1 - G_{\hat{\gamma}}(x; \hat{\lambda}, \hat{\delta}) = 1 - \exp \left\{ - \left[1 + \hat{\gamma} \left(\frac{x - \hat{\lambda}}{\hat{\delta}} \right) \right]^{-\frac{1}{\hat{\gamma}}} \right\} \quad [14]$$

Portanto, a probabilidade do recorde de mundo ser batido, i.e., $P[X > 74.08]$ e $P[X > 76.80]$, é de, aproximadamente, 0.334% nos homens e 0.208% nas mulheres.

Uma vez que $\hat{\gamma} < 0$, a cauda superior é finita, então,

$$\hat{x}^F := \hat{q}_0 = \hat{U}(\infty) = \hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}}, \quad [15]$$

com $U(t)$ dado em [17].

Assim, nos homens $\hat{x}^F = 74.32$ é superior ao atual recorde do mundo em apenas 0.32%, e nas mulheres $\hat{x}^F = 77.37$ ultrapassa o atual recorde do mundo em 0.74%.

Os quantis extremais obtêm-se invertendo a f.d. GEV [9]. Desta forma, o quantil extremal de probabilidade $1-p$ (também chamado de quantil extremal de probabilidade de excedência p), designado por q_{1-p} , é o valor que será excedido com muito pequena probabilidade, que neste estudo é dado por,

$$\hat{q}_{1-p} := G_{\hat{\gamma}}^{\leftarrow}(1-p | \hat{\lambda}, \hat{\delta}) = \hat{U}\left(\frac{1}{p}\right) = \left\{ \hat{\lambda} + \frac{\hat{\delta}}{\hat{\gamma}} \left[(-\log(1-p))^{-\hat{\gamma}} - 1 \right], \hat{\gamma} \neq 0 \right. \quad [16]$$

Substituindo $(\lambda, \delta, \gamma)$ pelas respectivas estimativas de ML $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$. De notar que o quantil extremal de probabilidade $1-p$ também pode ser escrito recorrendo à função quantil de cauda, ou seja,

$$U(t) := F^{\leftarrow}(1-1/t), \quad t \in [1, \infty] \quad [17]$$

Então, considerando $t = 1/p$, tem-se

$$q_{1-p} = G_{\gamma}^{\leftarrow}(1-p | \lambda, \delta) = G_{\gamma}^{\leftarrow}\left(1 - \frac{1}{1/p} | \lambda, \delta\right) = U\left(\frac{1}{p}\right) \quad [18]$$

A Tabela 4 apresenta as estimativas pontuais para alguns quantis extremos. Regista-se que a estimativa de um quantil extremal de probabilidade de excedência $p=0.25\%$ ($\hat{q}_{0,9975}$) nos homens e $p=0.1\%$ ($\hat{q}_{0,999}$) nas mulheres é superior ao atual recorde do mundo.

Tabela 4 – Estimativa pontual de alguns quantis extremos.

	$\hat{q}_{0,90}$	$\hat{q}_{0,95}$	$\hat{q}_{0,99}$	$\hat{q}_{0,995}$	$\hat{q}_{0,9975}$	$\hat{q}_{0,999}$
Homens	72.70	73.23	73.88	74.02	74.12	74.20
Mulheres	73.73	74.78	76.17	76.51	76.75	76.97

Em muitas situações é mais conveniente analisar os modelos considerando os níveis de retorno. Assim, quando se procede à estimação dos valores de retorno, enquanto valores limiares, as questões práticas que frequentemente se colocam são: (i) qual o nível de retorno esperado de um lançamento do disco para um certo período de retorno t ? (ii) qual a probabilidade de um lançamento do disco ultrapassar um certo nível de retorno? (iii) quanto tempo será necessário esperar até que ocorra um evento igual ou ainda mais extremo?

Os valores de retorno compreendem os níveis de retorno e o período de retorno (ou intervalo de recorrência médio). Neste estudo o nível de retorno é entendido numa escala anual. Portanto, o nível de retorno U associado a um dado período de retorno t ($t=1/p$) é o nível $U(t)$ que se espera que seja excedido, em média, uma vez a cada t -anos; por sua vez, o período de retorno t do nível U , i.e., $t(U)$, é o número médio de anos necessários até que um valor maior que U seja observado. O período de retorno é igual ao inverso da probabilidade de excedência, sendo portanto estimado como,

$$\hat{t}(U) = \frac{1}{G_{\hat{\gamma}}(U; \hat{\lambda}, \hat{\delta})} = \frac{1}{P[X > U]} \quad [19]$$

Portanto, estimados os parâmetros de localização, escala e forma, e uma vez que neste estudo $\hat{\gamma} < 0$, a estimativa do nível de retorno de t -anos, $\hat{U}(t)$, é dada pela inversão da f.d. GEV, ou seja,

$$\hat{U}(t) = G_{\hat{\gamma}}^{\leftarrow}\left(1-1/t | \hat{\lambda}, \hat{\delta}\right) = \hat{\lambda} + \frac{\hat{\delta}}{\hat{\gamma}} \left[\left(-\log\left(1 - \frac{1}{t}\right) \right)^{-\hat{\gamma}} - 1 \right] \quad [20]$$

onde $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$ são as estimativas apresentadas na Tabela 3.

A Tabela 5 apresenta as estimativas pontuais por ML, bem como os erros-padrão e os IC95% quer pelo método delta quer pela função profile log-likelihood, quanto aos níveis de retorno para os períodos de retorno a 5, 10, 20, 50 e 100 anos. Em termos de estimativa pontual, ou seja, o nível que se espera que seja excedido pelo menos uma vez em cada período de retorno considerado, nas duas amostras, não se afigura que o recorde do mundo seja batido nos próximos 100 anos! Na nossa perspectiva, mesmo considerando que o atual recorde do mundo se mantém há mais de 30 anos, mas atendendo à evolução que o fenómeno desportivo em geral e o lançamento do disco em particular tem vindo a sofrer e que previsivelmente irá continuar, não é expectável que um novo recorde do mundo possa demorar tanto tempo a surgir. Esta afirmação também é apoiada na evolução que o recorde do mundo teve nos últimos 100 anos, ou seja, nos homens: 47.58m em 1912 para

74.08m em 1986 (aumento de quase 56%) e nas mulheres: 27.39m em 1923 para 76.80m em 1988 (aumento de 180%). Contudo, provavelmente, o método dos máximos de blocos anuais, ao considerar apenas a observação máxima de cada ano descurou dentro do mesmo bloco outras importantes performances extremas, para além do valor máximo de um bloco poder ser inferior a alguns valores extremos de outros blocos. Coles (2001, p.54) afirmou que a escolha do tamanho (m) dos k blocos é uma questão delicada dado ser um “jogo” entre a variância e o viés. Assim, a distribuição equilibrada das n observações em blocos de pequena dimensão é suscetível de originar viés e uma aproximação pobre pela distribuição GEV, ao passo que a formação de blocos de grande dimensão produzem uma redução do número máximos de blocos, podendo originar grande variância na estimação. Por sua vez, Castillo (1994) vai mais longe ao afirmar que o método dos máximos (mínimos) de blocos leva a um importante desperdício da informação existente. Entre outras, mas também por estas razões, Coles (2001) e Caeiro et al. (2015) alertam para os cuidados a ter com a extrapolação para longos períodos de retorno. Importa, uma vez mais, realçar que o lançamento do disco lida com máximos porque são estes que garantem a obtenção de uma medalha de ouro.

Porém, uma estimativa pontual por si só é insuficiente. Qualquer estimativa requer os correspondentes erros-padrão e intervalos de confiança. Como os níveis de retorno são uma função dos parâmetros da distribuição GEV, usamos o método delta para estimar os erros-padrão, onde a variância do estimador do nível de retorno $\hat{U}(t)$ é aproximada como,

$$Var[\hat{U}(t)] \approx \nabla U^T(t) V \nabla U(t) \quad [21]$$

onde,

$$\nabla U^T(t) = \left[\frac{\partial x_t}{\partial \lambda}, \frac{\partial x_t}{\partial \delta}, \frac{\partial x_t}{\partial \gamma} \right] = \left[1, -\gamma^{-1}(1 - y_t^{-\gamma}), \delta \gamma^{-2}(1 - y_t^{-\gamma}) - \delta \gamma^{-1} y_t^{-\gamma} \log y_t \right] \quad [22]$$

onde $y_t = -\log(1 - t^{-1})$ e V é a matriz de variância-covariância para as estimativas $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$.

Obtida a estimativa do erro-padrão para cada nível de retorno, os intervalos de confiança podem ser construídos pela aproximação à Normal (Coles, 2001),

$$IC_{100(1-\alpha)\%}[\hat{U}(t)] = \hat{U}(t) \pm \Phi_{1-\alpha/2} \times \sqrt{Var[\hat{U}(t)]} \quad [23]$$

Por exemplo, numa estimativa do nível de retorno a 10 anos tem-se $p = 1/10$.

Pela Tabela 5 sabe-se que $\hat{U}(10) = 72.70$, $se = 0.2684$ nos homens e $\hat{U}(10) = 73.73$, $se = 0.5183$ nas mulheres. Então, um IC95% para $\hat{U}(10)$ é estimado como $72.70 \pm 1.96 \times 0.2684$ (homens) e $73.73 \pm 1.96 \times 0.5183$ (mulheres). Portanto, $\hat{U}(10) = 72.70$ nos homens e $\hat{U}(10) = 73.73$ nas mulheres, é o nível que se espera que seja excedido, em média, uma vez a cada 10 anos.

Segundo Coles (2001), maior precisão dos intervalos de confiança pode ser obtida pela função *profile log-likelihood*. Assim, um $IC_{100(1-\alpha)\%}[\hat{U}(t)]$ é dado por,

$$\left\{ \hat{U}(t) : \log L_p[U(t)] \geq \log L_p[\hat{U}(t)] - \frac{\chi_{1,1-\alpha}^2}{2} \right\}, \quad [24]$$

com $\log L_p U(t) = \max L[\hat{U}(t)]$. De notar que geralmente os intervalos de confiança pela função *profile log-likelihood* são assimétricos, i.e., a distância entre a estimativa pontual por máxima verosimilhança e os seus limites são diferentes. Pela Tabela 5 observa-se que o IC95% suporta a presença de um

novo recorde do mundo pelo método delta e pela *profile log-likelihood function*, para um período de retorno a 50 anos nos homens e nas mulheres, bem como pela função *profile log-likelihood function*, para um período de retorno a 50 anos nas mulheres.

Tabela 5: Estimativas pontuais por máxima verosimilhança, erros-padrão e intervalo com 95% de confiança pelo método delta e pela função *profile log-likelihood*, dos níveis de retorno para os períodos de retorno a 5, 10, 20, 50 e 100 anos.

	$\hat{U}(5)$	$\hat{U}(10)$	$\hat{U}(20)$	$\hat{U}(50)$	$\hat{U}(100)$
Homens	71.86 (0.320)	72.70 (0.268)	73.23 (0.232)	73.67 (0.201)	73.88 (0.211)
Método Delta	(71.23; 72.49)	(72.17; 73.22)	(72.78; 73.69)	(73.26; 74.09)	(73.47; 74.29)
Profile log-likelihood	(71.20; 72.49)	(72.18; 73.26)	(72.75; 73.82)	(73.32; 74.37)	(73.57; 74.60)
Mulheres	72.18 (0.578)	73.73 (0.518)	74.78 (0.472)	75.70 (0.443)	76.17 (0.447)
Método Delta	(71.05; 73.32)	(71.72; 74.75)	(73.86; 75.71)	(74.83; 76.57)	(75.30; 77.05)
Profile log-likelihood	(71.07; 73.34)	(72.74; 74.81)	(73.84; 76.00)	(74.95; 77.12)	(75.47; 77.74)

A Figura 3 apresenta a evolução dos níveis de retorno para vários períodos de retorno. Consta-se que as estimativas pontuais dos níveis de retorno, nas duas amostras, estão situadas dentro da banda com 95% de confiança, em todos os períodos de retorno considerados.

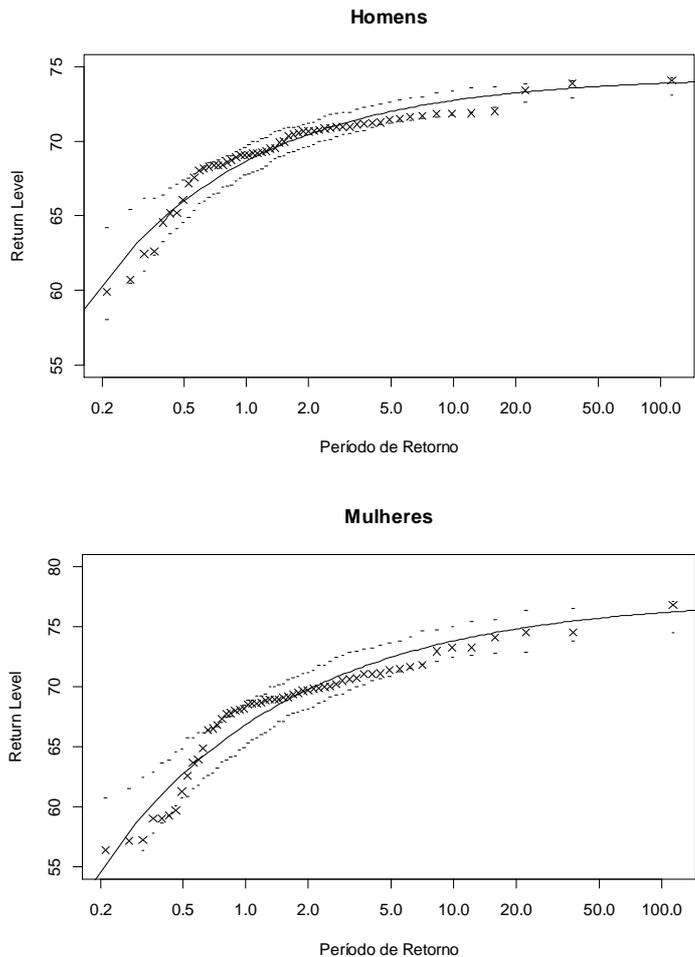


Figura 3: Níveis de retorno e banda de confiança a 95% para o modelo GEV ajustado para o lançamento do disco, homens e mulheres, no período de 1960-2016.

A Tabela 6 apresenta, para cada estimativa pontual de cada nível de retorno observado na Tabela 5 a respectiva probabilidade de excedência. Percebe-se que, nas duas amostras, a probabilidade de excedência diminui com o aumento dos valores de retorno.

Tabela 6 – Probabilidade de excedência para as estimativas pontuais dos níveis de retorno a 5, 10, 20, 50 e 100 anos.

	$\hat{U}(5)$	$\hat{U}(10)$	$\hat{U}(20)$	$\hat{U}(50)$	$\hat{U}(100)$
Homens	$P[X>71.86]$ =0.1999	$P[X>72.70]$ =0.0997	$P[X>73.23]$ =0.0501	$P[X>73.67]$ =0.0200	$P[X>73.88]$ =0.0099
Mulheres	$P[X>72.18]$ =0.2003	$P[X>73.73]$ =0.1001	$P[X>74.78]$ =0.0501	$P[X>75.70]$ =0.0201	$P[X>76.17]$ =0.0100

3.7 Ajustamento do modelo

Uma vasta gama de procedimentos pode ser usada para verificação do ajustamento do modelo. Neste estudo, recorreremos a metodologias gráficas. Dada uma amostra ordenada de observações independentes $\{X_1, \dots, X_n\}$ de uma população com f.d. F . Então, o gráfico de probabilidade consiste nos pontos,

$$\left\{ \left(\hat{F}(x_{(i)}), \frac{i}{n+1} \right) : i=1, \dots, n \right\}, \quad [25]$$

pelo que a distribuição estimada deve coincidir com a f.d. empírica $i/(n+1)$ a um nível razoável e, neste caso, os pontos deverão estar situados na proximidade ou sobre a reta de diagonal $y=x$.

Por sua vez, o gráfico dos quantis consiste nos pontos,

$$\left\{ \left(\hat{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)} \right) : i=1, \dots, n \right\} \quad [26]$$

que compara os quantis do modelo teórico com os quantis empíricos. Novamente, \hat{F} é uma representação válida de F se os pontos estiverem em torno da reta de diagonal igual a 1.

Na Figura 4 (homens) e na Figura 5 (mulheres) apresenta-se o ajuste do modelo pelos gráficos de probabilidade, de quantis, níveis de retorno e histograma com a função densidade.

Nas duas amostras, os gráficos de probabilidade e de quantis empíricos exibem uma tendência quase linear, com a nuvem de pontos a distribuir-se aproximadamente ao longo da reta de diagonal $y=x$. Geralmente, o gráfico de quantis é preferível ao gráfico de probabilidade. Assim, ainda que nas duas amostras existam alguns problemas na cauda inferior, tal não parece inviabilizar o ajustamento do modelo, pois este estudo é de máximos.

O histograma com a função densidade, apesar do caráter subjetivo, não deixa de ser uma importante ferramenta no apoio à avaliação da qualidade do ajustamento do modelo. O ajuste do modelo é observado pela consistência da curva de densidade de probabilidade face aos dados, que parece ser relativamente consistente, evidenciando contudo uma ligeira assimetria à esquerda.

O gráfico dos níveis de retorno, apresentado, numa escala logarítmica, consiste na estimativa de máxima verosimilhança dos níveis de retorno, $\hat{U}(t)$, por oposição a $-\log(1-p)$, $0 < p < 1$. O parâmetro de forma define a forma que o gráfico irá assumir e o comportamento do limite superior. As bandas de confiança, baseadas num IC95%, acopladas às estimativas empíricas são uma ferramenta importante na ajuda à avaliação do modelo.

Dado que neste estudo $\hat{\gamma} < 0$, o gráfico dos níveis de retorno, nas duas amostras, exibe uma forma convexa voltada para cima. Os valores extremos situam-se na proximidade da curva dos níveis de retorno, mas dentro da banda de confiança a 95%, demonstrando, por isso, concordância entre a f.d. empírica com a curva dos níveis de retorno. Nas duas amostras, independentemente do período de retorno, a banda com 95% de confiança tem amplitude similar, não se verificando flutuações pronunciadas para períodos de retorno mais elevados. Em suma, a análise gráfica sugere a presença de um razoável ajustamento e a adequação do modelo GEV.

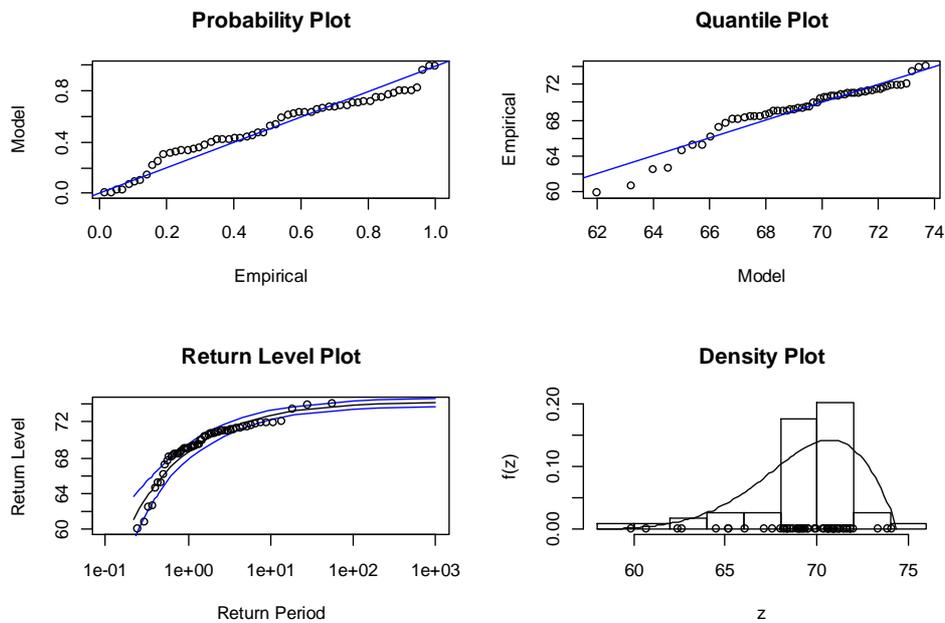


Figura 4: Gráficos de diagnóstico para o modelo GEV ajustado para o lançamento do disco homens, no período de 1960-2016.

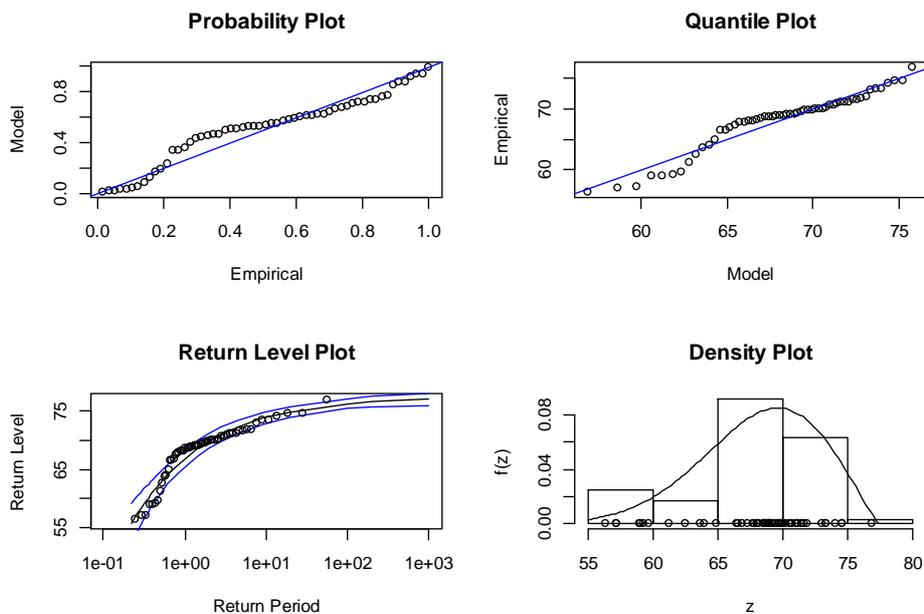


Figura 5; Gráficos de diagnóstico para o modelo GEV ajustado para o lançamento do disco mulheres, no período de 1960-2016.

4 Conclusões

Neste estudo utilizamos o método dos máximos de blocos anuais, a distribuição generalizada de valores extremos e a estimação por máxima verosimilhança com otimização de Nelder-Mead para estimar os principais parâmetros de valores extremos (e com isto, outros parâmetros de interesse) no lançamento do disco, em homens e mulheres, no período de 1960-2016. Nas duas amostras a cauda da distribuição é leve e finita ($\hat{\gamma} < 0$). A probabilidade de excedência do atual recorde do mundo é de, aproximadamente, 0.33% nos homens e 0.21% nas mulheres. Atendendo às condições atuais e à metodologia seguida neste estudo, com base no intervalo com 95% de confiança, o nível de retorno a 50 anos contém um novo recorde do mundo para homens e mulheres. Julgamos ser necessário incluir observações anteriores a 1960, mesmo que tal possa afetar a estacionariedade e/ou a tendência da sequência de dados, bem como realizar uma análise recorrendo a outros métodos (e.g., maiores observações anuais, excessos acima de um limiar). Não obstante, estes resultados são um importante guia para todos os que se interessam pela problemática dos extremos no desporto, especialmente no lançamento do disco.

Agradecimentos

Esta pesquisa foi parcialmente suportada pela Fundação para a Ciência e a Tecnologia através dos projetos UID/MAT/04674/2013 (Centro de Investigação em Matemática e Aplicações, Universidade Évora) e UID/MAT/00297/2013 (Centro de Matemática e Aplicações, Universidade Nova de Lisboa).

Referências Bibliográficas

- [1] Brockwell, P.J. & Davis, R.A. (2002). Time Series and Forecasting. Second Edition, Springer Texts in Statistics, New York: Springer.
- [2] Caeiro, F.; Mateus, A. & Ramos, L. (2015). Extreme value analysis of the sea levels in Venice. AIP Conf Proc, 1648, 540006-1540006-4.
- [3] Castillo, E. (1994). Extremes in Engineering Applications. In Extreme Value Theory, Galambos J, Lechner, J. & Simiu, E. (eds.), pp.15-42, Kluwer Academic Publishers, London.
- [4] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. London: Springer-Verlag.
- [5] Cox, D.R. & Stuart, A. (1955). Some quick sign tests for trend in location and dispersion. Biometrika, 42 (1-2): 80-95.
- [6] Einmahl, J.H.J. & Magnus, J.R. (2006). Records in athletics through Extreme-Value Theory. CentER Discussion Paper Series No. 2006-83.
- [7] Einmahl, J.H.J. & Magnus, J.R. (2008). Records in athletics through extreme-value theory. Journal of the American Statistical Association, 103 (484): 1382-1391.
- [8] Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample, Proceedings of the Cambridge Philosophical Society, 24: 180-190.
- [9] Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. Annales de la Société Polonaise de Mathématique (Cracovie), 6: 93-116.
- [10] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics, 44 (3): 423-453.
- [11] Gumbel, E.J. (1935). Les valeurs extrêmes des distributions statistiques. Annals of Institute Henri Poincaré, 5 (2): 115-158.
- [12] Gumbel EJ (1958). Statistics of Extremes. New York: Columbia University Press, USA.

- [13] IAAF - International Association of Athletics Federations (2015). Progression of IAAF World Records. 2015 edition, Edited by Richard Hymans (ATFS) and Imre Matrahazi (IAAF). Multiprint, Monaco.
- [14] Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81 (348): 158-171.
- [15] Moore GH, Wallis WA (1943). Time series significance tests based on signs of differences. Journal of the American Statistical Association, 38 (222): 153-164.
- [16] Nelder, J.A. & Mead, R. (1965). A simplex method for function minimization. Computer Journal, 7 (4): 308-313.
- [17] Schmolinsky G (1982). Atletismo. Lisboa: Editorial Estampa.
- [18] Smith, RL (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika, 72 (1): 67-90.
- [19] von Mises R (1936). La distribution de la plus grande de n valeurs. Revue Mathématique de l'Union Interbalkanique, 1, 141-160.
- [20] von Mises R (1954). La distribution de la plus grande de n valeurs. Revue Mathématique de l'Union Interbalcanique, 1, 141-160. Reprinted in Selected Papers of Richard von Mises, Am Math Soc, 2, 271-294.
- [21] Weibull W (1939). The Phenomenon of Rupture in Solids. Ingeniors Vetenskaps Akademien Handlinga, 153, Stockholm: Sweden.
- [22] Weibull, W. (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanics American Society of Mechanical Engineers, 18 (3): 293-297.
- [23] URL: https://en.wikipedia.org/wiki/Discus_throw
- [24] URL: <http://www.all-athletics.com/en-us/all-time-lists>
- [25] URL: <http://trackfield.brinkster.net/>
- [26] URL: https://en.wikipedia.org/wiki/National_records_in_athletics