

Aplicação da Teoria de Valores Extremos no estudos de dados da Covid-19: análise probabilística do número de novos óbitos diários

(Relatório Parcial)

Bolsista: Ana Carolina Matussi

Orientador: Prof. Dr. Gilberto Rodrigues Liska - DTAiSeR

Universidade Federal de São Carlos (UFSCar)
Campus Araras - CCA



Introdução

- ▶ A **incidência do Coronavírus** no Brasil é preocupante, pois a estimativa de infectados e mortos concorre diretamente com o **impacto sobre o sistema de saúde**.
- ▶ No Brasil, o **primeiro caso confirmado** ocorreu em 26 de fevereiro, após 20 dias, houve o **primeiro caso de morte** em 17 de março de 2020 e com 121 novos casos registrados. É evidente que a doença apresenta um crescimento diário significativo.
- ▶ Nessa lógica, é possível **prever a sua ocorrência** e compreender o padrão de eventualidade desses valores, quantificando a **probabilidade desse caso acontecer**.
- ▶ **Prevenção** ou a redução dos potenciais **efeitos drásticos**, como um estado de calamidade pública.

Introdução

- ▶ A **Teoria dos Valores Extremos** (TVE) é muito utilizada para o estudo de **eventos raros** de um determinado fenômeno, sendo eles epidemiológicos, econômicos ou climáticos.
- ▶ Em muitas aplicações estatísticas o **interesse é direcionado** para a **estimação de características centrais**, por exemplo, o valor médio de precipitação ou o valor médio de temperatura. Em registros epidemiológicos, **o valor médio de mortes** ou o **valor médio de casos**.
- ▶ O que os extremos podem nos mostrar e quais informações esses dados contêm?
- ▶ Mas qual a consequência desses valores extremos no sistema de saúde pública?

Objetivos

- ▶ O projeto tem como objetivo analisar os dados do **número de óbitos** decorrentes da **COVID-19** no Brasil e, com base nesses dados, utilizar a **Teoria de Valores Extremos (TVE)** para **avaliar a probabilidade** de eventos extremos **ocorrerem futuramente**.
- ▶ **Compreender** as distribuições de variáveis aleatórias ao longo do tempo, obtendo a **probabilidade de ocorrência máxima mensal** do número de **novos óbitos** diários causados pela COVID-19.
- ▶ Modelagem dos dados das metodologias de **Blocos Máximos (BM)** e **Picos Acima do Limiar** (POT - *Peaks Over Threshold*) com suas respectivas distribuições.

Dados

- ▶ Os dados foram obtidos de publicações oficiais e diárias do ***Our World in Data***, ligada à Universidade de Oxford – Reino Unido.
- ▶ Separados em dois intervalos:
 - ▶ O período de 26 de fevereiro de 2020 a 31 de março de 2021 como série de dados de treinamento.
 - ▶ O períodos de 1 de abril de 2021 a 31 de janeiro de 2022 como série de dados para comparação.
- ▶ Os dados diários foram organizados em **períodos mensais**:
 - ▶ Considerando como variável o número de mortalidade e obtendo o seu valor máximo.

Materiais e Métodos

▶ Modelos probabilísticos

▶ Metodologia de **Blocos Máximos** (BM):

▶ Distribuição generalizada de valores extremos (GVE) e distribuição Gumbel

▶ Metodologia dos **Picos Acima do Limiar** (POT - *Peaks Over Threshold*):

▶ Distribuição generalizada de Pareto (DGP) e distribuição Exponencial

▶ Estimação dos parâmetros

▶ Método da máxima verossimilhança (MV)

Materiais e Métodos

▶ Testes de Hipóteses (valor $p > 0,01$)

▶ Teste **Ljung-Box**

▶ Teste de **Mann-Kendall**

▶ Teste de aderência de **Kolmogorov-Smirnov** (KS)

▶ Teste da **razão de verossimilhança** (TRV)

▶ Avaliação dos modelos

▶ Raiz quadrada do erro quadrático médio (RMSE), Erro percentual absoluto médio (MAPE), Erro Absoluto Médio (MAE), índice de concordância de Willmott modificado (md). Amplitude Média Intervalar (AMI) e o Critério de eficiência de Nash-Sutcliffe (NSE).

▶ **Recurso computacional:** Software R com os pacotes *evd*, *extRemes*, *hydroGOF*

Resultados Parciais

Tabela 1: Resultados das estimativas dos parâmetros das distribuições da metodologia BM e seus respectivos testes de hipótese para o número de novos óbitos de Covid-19

Metodologia dos Blocos Máximos (BM)							
Análise Mensal							
Distribuição	Parâmetros			TRV	Mann-Kendal	Ljung-Box	KS
	μ	σ	ξ				
GVE	1064,78	679,60	-0,0407	0,7903	0,583	0,01118	0,5841
Gumbel	1057,39	690,22	-				0,5682

Tabela 2: Resultados das estimativas dos parâmetros das distribuições da metodologia POT e seus respectivos testes de hipótese para o número de novos óbitos de Covid-19

Metodologia dos Picos Acima do Limiar (POT)							
Análise Mensal							
Distribuição	Parâmetros			TRV	Mann-Kendal	Ljung-Box	KS
	μ	σ	ξ				
Exponencial	2800	577,71	-	0,1191	0,0162	0,0896	0,6596
DGP	2800	726,71	-0,5963				0,4912

Resultados Parciais

Tabela 3: Probabilidade de ocorrência de novos óbitos de Covid-19 da metodologia BM e o valor máximo esperado

Metodologia dos Blocos Máximos (BM)										
Análise Mensal										
	Probabilidades					Esperado (meses)				
Distribuição	>1000	>2000	> 3000	> 4000	>5000	2	3	4	5	6
GVE	66,71	21,54	4,73	0,86	0,14	1312	1667	1890	2053	2182
Gumbel	66,27	22,52	5,82	1,40	0,33	1310	1680	1917	2093	2232

Tabela 4: Probabilidade de ocorrência de novos óbitos de Covid-19 da metodologia POT e o valor máximo esperado

Metodologia dos Picos Acima do Limiar (POT)										
Análise Mensal										
	Probabilidades					Esperado (meses)				
Distribuição	>1000	>2000	> 3000	> 4000	>5000	2	3	4	5	6
Exponencial	17,71	3,14	0,56	0,10	0,02	2328	2562	2728	2857	2963
DGP	5,61	0,00	0,00	0,00	0,00	2035	2461	2707	2870	2988

Próximas etapas

- ▶ Em uma segunda análise, será feito a análise para a 1ª quinzena e 2ª quinzena do mês, totalizando **três séries de observações** para as duas metodologias propostas.
- ▶ Realizar a análise para o **número de casos** confirmados de COVID-19, seguindo a mesma dinâmica da primeira análise.
- ▶ Utilizar outras medidas de **qualidade de ajuste** para avaliar possíveis incertezas nas estimativas de ambos os modelos.

Referências Bibliográficas

- ▶ ANTUNES, J. L. F.; CARDOSO, M. R. A. **Uso da análise de séries temporais em estudos epidemiológicos.** Epidemiologia e Serviços de Saúde, v. 24, no 3, p. 565–576, 2015. Disponível em: <https://doi.org/10.5123/S1679-49742015000300024>.
- ▶ ATKESON, Andrew. **What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios.** National Bureau of Economic Research, 2020. Disponível em: <http://www.nber.org/papers/w26867.pdf>.
- ▶ COLES, Stuart. **An Introduction to Statistical Modeling of Extreme Values.** Springer series in statistics, v. 1, no 1, p. 1-219, 2001. Disponível em: <https://doi.org/10.1007/978-1-4471-3675-0>.
- ▶ GUILLOU, A.; KRATZ, M.; STRAT, Y. Le. **An extreme value theory approach for the early detection of time clusters. A simulation-based assessment and an illustration to the surveillance of Salmonella.** Statistics in Medicine, v. 33, no 28, p. 5015–5027, 2014. Disponível em: <http://doi.wiley.com/10.1002/sim.6275>.
- ▶ OPAS, O. P.-A. de S. **Folha informativa sobre COVID-19 - OPAS/OMS | Organização Pan-Americana da Saúde.** paho.org. 2021. Disponível em: <https://www.paho.org/pt/covid19>.
- ▶ R CORE TEAM. **R: A language and environment for statistical computing.** Foundation for Statistical Computing. 2021. Disponível em: <https://www.r-project.org/>.
- ▶ THOMAS, M. et al. **Applications of extreme value theory in public health.** PLoS ONE, v. 11, no 7, p. e0159312, 2016. Disponível em: <https://doi.org/10.1371/journal.pone.0159312>.

Agradecimentos

