

Ciência & Tecnologia: Modelagem de Doenças Infecciosas



História e processos da modelagem de doenças infecciosas

O aparecimento e reemergência de doenças infecciosas são continuamente documentados na história da humanidade com o delineamento de suas transmissões, sendo fortemente relacionados à globalização econômica, mudanças ambientais, demográficas e tecnológicas. Alguns agentes infecciosos causaram inúmeras mortes antes do seu desaparecimento; outros se tornaram endêmicos (epidemias periódicas) em uma população e, conseqüentemente, traçou um perfil de desigualdades, destacando regiões pobres e com precários sistemas de saúde. Patologias como a malária, cólera, esquistossomose, tuberculose, arbovírus de maneira geral (enfermidades transmitidas por vetores como mosquitos), problemas respiratórios, dentre outras, têm um impacto significativamente negativo na expectativa de vida da população, e afeta a economia de um país, uma vez que uma população doente tem sua força de trabalho reduzida. Neste cenário, a modelagem tem se tornado uma aliada, contribuindo para a compreensão das complexidades da transmissão e evolução dos patógenos, predição de tendências e seu controle. São modelos que se harmonizam aos contextos variados, capturam características individuais até um nível macro - levando em consideração características regionais, sociais e econômicas da população. Os resultados obtidos guiam áreas da epidemiologia e saúde pública no estudo da dinâmica de transmissão e controle de contágio de doenças infecciosas.

Breve histórico das doenças infecciosas

por A. RIOS, J. OLIVEIRA, M. RODRIGUES, P. RAMOS

As doenças infecciosas (DIs) são causadas por agentes patogênicos como bactérias, vírus, parasitas, fungos ou príons. A transmissão das DIs pode ocorrer de forma direta (através do contato ou vias respiratórias como tosse, espirros) ou indireta (por intermédio de reservatórios, água, alimentos, materiais biológicos, vetores), de

humanos para humanos ou de animais para humanos, sendo esta denominada como doenças zoonóticas [1]. Uma vez que um desses agentes infecciosos é introduzido em uma população, a sua disseminação pode ganhar grandes proporções e causar uma epidemia, ou seja, causar infecções em um número desproporcionalmente grande de indivíduos em um curto período de tempo [2, 3].

Apesar da moderna definição das DIs, a existência de micro-organismos somente foi demonstrada no século

XVII pelo cientista Van Leeuwenhoek, advinda da aparição do primeiro microscópio. Ainda sem a criação e formalização da teoria microbiana das doenças, uma das primeiras contribuições da modelagem deu-se pelo trabalho de John Graunt (comerciante, considerado também um dos primeiros demográficos e epidemiologistas na história) (1620 - 1674) em seu livro *Natural and Political Observations made upon the bills of Mortality* (Observações Naturais e Políticas feitas nas Contas da Mortalidade,

em português), em 1662, baseado em dados sobre DIs e mortalidade [4]. A partir da análise de seus dados, o autor forneceu um método para estudar a hipótese de quais fatores de risco causados pelas infecções de várias doenças competem pela vida dos indivíduos (ou seja, para cada indivíduo, um desses riscos 'vencerá', e o indivíduo morrerá por causa desse risco). Suas ideias iniciaram a teoria do risco competitivo que vai além das causas de mortes motivadas por doenças. Desde então, somente no final do século XIX e início do século XX, temos uma melhor compreensão dos agentes que provocam doenças infecciosas com o princípio da teoria microbiana das doenças, descrita por Jacob Henle (médico, patologista e anatomista) (1809 - 1885), em 1840, e que foi desenvolvida por Robert Koch (médico e microbiologista) (1843 - 1910), Joseph Lister (cirurgião) (1827 - 1912) e Louis Pasteur (biólogo, microbiologista e químico) (1822 - 1875).

Até o final do século XX, as DIs foram responsáveis por grande parte da taxa de mortalidade e malformações na população. Pandemias globais como varíola, cólera e gripe ameaçaram periodicamente a sobrevivência de populações inteiras. Com a melhoria das condições sanitárias, abastecimento de água e mudanças na qualidade de vida da população, as DIs deram espaço para as doenças não infecciosas, responsáveis na atualidade por maior parte das causas de mortes no mundo [2, 3].

De fato, a epidemia da varíola foi uma das piores doenças na história da humanidade, sendo também a única doença do mundo a ser eradicada e uma das conquistas mais triunfantes do mundo na medicina [5]. Coincidentemente, o primeiro modelo associado à epidemiologia matemática foi agregado ao trabalho de Daniel Bernoulli (matemático e físico) (1700 - 1782) [6]. Em 1760, Bernoulli escreveu um artigo intitulado *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it* (Uma tentativa de uma nova

análise da mortalidade causada pela varíola e das vantagens da inoculação para preveni-la, em português) [6, 7]. Na ausência de vacina, o trabalho de Bernoulli tivera como objetivo compreender se a inoculação (a introdução voluntária de uma pequena quantidade de varíola menos virulenta no corpo para protegê-lo contra infecções posteriores) seria uma técnica a ser considerada, mesmo que às vezes fosse uma operação mortal. Com uso de dados sobre os casos de varíola, da Tábua de mortalidade de Halley e modelagem matemática, Bernoulli estimou que a inoculação poderia ser vantajosa se os fatores de risco associados à morte fossem menores do que 11% e, conseqüentemente, a inoculação pudesse aumentar a expectativa de vida ao nascer em 3 anos.

Outra contribuição notável da modelagem de DIs relaciona-se à epidemia de cólera. A cólera teve grande impacto no século XIX, sendo reduzida significativamente em países desenvolvidos devido à melhoria de saneamento básico e tratamento da água. Contudo, a doença ainda é endêmica em mais de 47 países em todo o mundo, causando aproximadamente 2,9 milhões de casos e matando cerca de 95.000 pessoas [8]. Na epidemia de 1855 em Londres, o médico John Snow (1813 - 1858) contribuiu para a descoberta da fonte causal da epidemia de cólera na cidade [9]. Para tal efeito, Snow desenvolveu um estudo para analisar padrões espaço-temporais de casos de cólera durante a epidemia, obtendo assim, a localização de uma fonte de abastecimento de água responsável pelas infecções.

Os exemplos anteriores foram pontos marcantes na história das doenças infecciosas, dando uma síntese do que foi desenvolvido no início da modelagem de DIs. Contudo, as contribuições na área de modelagem de DIs ocorreram majoritariamente a partir do século XX, tendo hoje em dia ramos de pesquisas separados em modelagem estatística, matemática e computacional. Essas áreas têm como interseção o conceito de modelo, que

pode ser traçado de várias formas, tendo diferentes níveis de complexidade, com objetivos de estudos variados, dependendo da avaliabilidade de dados, recursos computacionais, precisão e generalidade requerida e da janela de tempo disponível para a demanda de resultados. Porém, também, não podemos esquecer, condizente com a definição da palavra, um "Modelo" é uma representação, imitação e, portanto, não tem o poder de descrever fielmente um fenômeno natural. Assim sendo, um modelo sempre virá acompanhado de hipóteses que correspondam às condições postas e facilitem e viabilizem a sua manipulação matemática-estatística-computacional.

As três áreas se comunicam e baseiam-se em fundamentos epidemiológicos, gerando assim um nicho interdisciplinar, que desenvolve teorias e aplicações que surtem efeitos imediatos nas populações.

Uma ponte para o futuro: comparação entre inoculação e políticas de imunidade coletiva contra COVID-19 em um cenário ainda sem a vacina

por NATÁLIA TAVARES

A estratégia de inoculação se assemelha às estratégias de imunidade natural por considerarem a exposição ao agente infeccioso como método preventivo da doença em questão. No entanto, ambas são estratégias que incorrem em risco de desenvolvimento da doença em sua forma grave com evolução para óbito. Por isso, a avaliação de risco é crucial para estimar as potenciais vantagens e desvantagens destas estratégias, tanto no âmbito individual quanto populacional. Diferentemente do que ocorre com uma vacina, onde o inóculo é composto apenas por partes do microorganismo ou por um vetor não-infeccioso, gerando imunidade sem doença, nas estratégias de inoculação ou imunidade natural, as respostas individuais são desconhecidas.

das e os riscos para a evolução de formas graves, consideráveis.

Diante disso, é importante considerar que o termo “imunidade coletiva” define uma resistência à disseminação de uma doença contagiosa numa determinada população. Isso ocorre porque uma parcela significativa da população está imune ao microorganismo, reduzindo as chances de transmissão entre um indivíduo infectado para outro susceptível [10]. Isto é, os indivíduos imunizados formam uma barreira de contenção, protegendo aqueles não-imunizados. No entanto, a imunidade coletiva é baseada na imunidade individual. Uma vez que o indivíduo é exposto à uma substância exógena ou agente infeccioso, responde a estes estímulos e os elimina para manter seu estado saudável, ele é considerado imune. Além disso, a imunidade pode ser induzida pela vacinação. Dessa forma, a imunidade coletiva pode ser alcançada através de vacinação ou por muitas pessoas sendo infectadas pelo agente infeccioso.

A imunidade coletiva já foi observada para o sarampo nos EUA na década de 30 do século passado, onde surtos foram contidos após 68% das crianças terem sido infectadas [11]. Isso se manteve pelas décadas seguintes, com números de casos em níveis baixos até a introdução da vacina na década de 70 [12].

Existem também alguns exemplos relevantes de doenças infecciosas que foram erradicadas ou tiveram sua disseminação controlada através da imunidade coletiva. A varíola, por exemplo, foi oficialmente declarada erradicada em 1979 com base em imunidade coletiva alcançada através de intensas campanhas de vacinação (Lane, 2006). Apesar de doenças como sarampo, rubéola e coqueluche não terem sido erradicadas ainda, a imunidade coletiva é mantida através das proporções altas de indivíduos imunes que protegem aqueles susceptíveis [13].

Em se tratando de COVID-19, que não há vacina aprovada ainda, a imunidade coletiva não pode ser alcançada através de vacinação. Neste

sentido, a imunidade coletiva seria alcançada através da infecção natural. Então, qual seria a proporção de pessoas que precisam estar imunizadas contra a COVID-19 para gerar barreira de contenção e proteção dos não-imunizados? Existem diferentes métodos de projeção destes números, mas, em média, as estimativas indicam que cerca de 56% da população precisaria estar imune para alcançar imunidade coletiva contra COVID-19. No entanto, é importante considerar os possíveis desfechos e riscos que decorrem da infecção natural ou inoculação. Um indivíduo naturalmente infectado por SARS-CoV-2 pode se recuperar da doença ou pode ir a óbito em decorrência da evolução para a forma grave da doença.

De acordo com dados epidemiológicos, a COVID-19 se manifesta com casos graves em cerca de 15% dos indivíduos infectados com taxa de mortalidade de 2%, em média. Considerando esses valores, é importante ponderar sobre os “custos” que a infecção natural podem trazer. Estes “custos” envolvem desde perdas de vidas pela forma grave da doença, bem como onera os sistemas de saúde que manejam os cuidados desses pacientes. Diante disso, a modelagem matemática tem sido fundamental para o desenvolvimento de estudos que avaliam os riscos decorrentes da imunidade coletiva alcançada através de infecção natural ou inoculação.

Processos da modelagem Matemática: modelos de compartimentos

por JULIANE OLIVEIRA

A modelagem matemática de DIs teve seu início no finalzinho do século XIX, em que parte dos estudos apresentados desenvolveu-se por profissionais em saúde pública [14]. Até então várias perguntas estavam sem respostas: qual o potencial de um agente infeccioso causar uma epidemia? Qual a proporção e duração da epidemia? A doença atingirá toda

a população? Pode desaparecer? Reaparecer? Se tornar periódica? Como a disseminação da doença se comportará em diferentes cenários?

Para compreender como desenvolveu-se a busca por respostas às questões acima, podemos observar no trabalho do médico Pyotr Dimitrievich EN'KO (1844 - 1916) uma discussão que direcionava para os fundamentos da modelagem de uma epidemia. EN'KO inicialmente simplificou o problema, considerando que em uma população podemos ter indivíduos suscetíveis, ou seja, que podem se infectar quando expostos a um agente infeccioso, assim como a possibilidade de uma parte dessa população ser imune, neste caso ele assumiu que o indivíduo nunca contrairia a doença. Uma vez que a pessoa infectada adentra nessa população, conseqüentemente, ela entra em contato com um indivíduo saudável. Se este for suscetível, ele se contagia e se torna transmissor da doença. No seu trabalho [15] de 1889, EN'KO discute que os elementos básicos para descrever o processo da transmissão da doença depende da probabilidade de uma pessoa saudável entrar em contato com um infectado, da probabilidade daquela se infectar e do número de suscetíveis na população. Assim, comunidades pequenas e isoladas dificultam a transmissão de uma doença; em contrário, sociedades numerosas, e em particular com uma grande taxa de contato entre os indivíduos, são mais propícias a terem epidemias.

Em 1906, o médico e epidemiologista William Heaton Hamer (1862-1936) defendeu que a disseminação de uma doença depende do número de indivíduos suscetíveis e infectados, reconhecendo assim que apenas a diminuição da densidade de pessoas suscetíveis poderia interromper uma epidemia. Para descrever a taxa de novas infecções em uma população, Hamer sugeriu o uso da Lei de ação das massas: uma proposição que afirma que a taxa da reação química é diretamente proporcional ao produto das interações ou concentrações dos reagentes, fazendo assim uma analo-

gia entre “taxa de reação química” e “taxa de novas infecções” [16]. Esse foi o primeiro passo que possibilitou o formalismo da transmissão de doenças através de equações matemáticas, dando origem aos modelos de compartimentos.

Um compartimento é um agrupamento esquemático da população em que se divide grupo de indivíduos que apresentam características similares. Quando essas características estão relacionadas à alguma doença infecciosa, o exemplo mais simples de divisão da população de estudo é entre três compartimentos: um grupo de indivíduos suscetíveis (S), um grupo de infectados (I) e outro de imunizados (R). Um modelo de compartimento descreve em termos matemáticos a variação no tempo dos indivíduos de um compartimento para o outro. Além das contribuições de Hamer, a formalização e descrição matemática dos modelos compartimentais foram também dadas pelos trabalhos do médico Sir Ronald Ross (1857 – 1932), pelo médico e epidemiologista Anderson Gray McKendrick (1876 – 1943) e pelo bioquímico William Ogilvy Kermack (1898 – 1970) [14]. Em 1902, Ross foi agraciado com o Prêmio Nobel em Medicina graças ao seu trabalho que descreveu a dinâmica de transmissão da malária entre mosquitos e humanos. Ross mostrou que reduzindo-se a população de mosquitos até um certo nível, seria possível eliminar a transmissão da malária, dando assim o primeiro conceito do número básico de reprodução [17].

Até a data atual, dada a sua simplicidade, os modelos compartimentais formam a base para a construção conceitual da modelagem de doenças infecciosas. A fim de construir tal modelo, o primeiro passo consiste em compreender o fenômeno biológico que se deseja estudar (modelar): qual é o tipo de doença que deseja modelar? Como ela é transmitida? Se existem fatores que influenciam sua transmissão como idade, clima, condições sociais etc; subsequentemente, traduzir os aspectos biológicos, comportamentais, imunológicos

e demográficos em termos que nos permitam analisar o fenômeno com rigor matemático, tecnicamente descrito por equações. As equações matemáticas, posteriormente, serão calibradas com dados disponíveis, quer dizer, medidas e observações do fenômeno. Por fim, uma vez que as condições de análises são biológica e matematicamente plausíveis, desenhamos as conclusões, interpretando a linguagem matemática em termos biológicos. Uma vez que o modelo é validado, é possível simular cenários que direcionem medidas de políticas públicas que contribuem para o controle e até eliminação de uma doença na população.

No atual momento, com a disseminação do SARS-COV-2, agente causal da COVID-19, e na ausência de uma vacina, vários modelos estão sendo construídos para descrever a evolução da doença em diversos contextos sociais. No início da epidemia, modelos simples como o SIR (Suscetível- Infectado- Recuperado) ou até SEIR (E se refere à exposto) são adequados para analisar o potencial de disseminação da doença na região. Para direcionar medidas de controle da doença, compartimentos que simulam quarentenas e isolamentos de infectados podem ser agregados aos modelos anteriores, dando assim uma ideia do impacto destas medidas no achatamento da curva de infectados. Em cenários mais críticos, em que a aplicação de medidas mais restritas se tornam inviáveis, estratégias para proteger o sistema de saúde podem ser consideradas ao adicionar a dinâmica hospitalar no modelo [18, 19, 20]. Essas são algumas das variadas técnicas de modelagem que foram e estão sendo desenvolvidas para conter a disseminação da COVID-19, gerando temas além do objetivo desta Newsletter.

Modelagem estatística

por NIVEA SILVA E ROSEMEIRE FIACCONE

“Estatística”, palavra derivada do latim *status*, significa originalmente

“o estudo do estado”. Sua história inicia com os antigos egípcios, por volta do ano 5000 antes da era cristã, que já nesta época mantinham um sistema de registro dos seus prisioneiros de guerra, e com os chineses, que foram os primeiros por volta do ano 2000 antes da era cristã a se preocupar com o crescimento populacional, realizando um censo da sua população e lavouras cultivadas. Já na Era Cristã, deve-se ao matemático e astrônomo Menelaus de Alexandria a criação das primeiras tabelas de estatística cruzadas e aos Constantinos a primeira agência de Estatística. Os árabes, por volta dos anos 695, utilizaram a ideia de média ponderada para fazer contagem de moedas e, em 826, fizeram uso de cálculos estatísticos como estratégia na tomada de Creta (http://redeabe.org.br/historia_estatistica/).

Foi o acadêmico alemão Gottfried Achenwall, considerado um dos pais da Estatística, um dos primeiros a introduzir, no século XVII, o termo “Estatística” e usá-lo para analisar informações econômicas, sociais e políticas. Assim, as primeiras aplicações do pensamento estatístico eram, na época, voltadas para a formulação de políticas públicas, fornecendo dados demográficos e econômicos. Ainda no século XVII, os matemáticos Blaise Pascal e Pierre de Fermat também estudaram Estatística, através do desenvolvimento da teoria das probabilidades, para solucionar problemas relacionados com jogos de azar. Um dos teoremas mais importantes na área de probabilidade, conhecido como a Lei dos grandes números (LGN), foi formulado no século XVIII pelo matemático suíço Jacob Bernoulli. Esta e outras contribuições de Jacob foram publicadas alguns anos após sua morte no livro *Ars Conjectand*, editado por seu sobrinho Nicholas Bernoulli. Outra contribuição importante deste século, e que revolucionou a Estatística nos dias atuais, é atribuída ao matemático e pastor presbiteriano inglês Thomas Bayes. O reverendo Bayes, como era conhecido, demonstrou o famoso teorema que

leva seu nome: o Teorema de Bayes. Este e outros resultados atribuídos a Bayes foram reunidos em um ensaio chamado *Essay Towards Solving a Problem in the Doctrine of Chances*, publicado em 1763, após sua morte, no periódico *Philosophical Transactions* da Royal Society of London. O raciocínio construído por Bayes neste teorema deu origem, anos mais tarde, a um novo paradigma em Estatística: o paradigma Bayesiano.

Foi, contudo, a partir do século XIX, através das contribuições de nomes como o matemático, astrônomo e físico francês Pierre-Simon de Laplace (princípio da probabilidade inversa, Teorema central do Limite de Moivre-Laplace, dentre outros), do matemático e físico alemão Carl Friedrich Gauss (método de mínimos quadrados), do antropólogo, meteorologista, matemático e estatístico inglês Francis Galton (Teoria da regressão, noção inicial de correlação, demonstrada mais tarde por Pearson), do estatístico britânico Karl Pearson (fundador do *Biometrika* journal, demonstrou o coeficiente de correlação e foi um dos primeiros a afirmar que correlação não implica causalidade; propôs o teste qui-quadrado e método dos momentos, dentre outras contribuições), do estatístico, biólogo evolutivo e geneticista britânico Ronald Fisher (propôs a Verossimilhança, função que possui um papel central na Estatística, produzindo resultados teóricos de inferência estatística que englobam desde estimação de parâmetros, via método da máxima verossimilhança, até a realização de testes de hipótese e construção de regiões de confiança. Desenvolveu também a ANOVA e a área de planejamento de experimentos), dentre outros, que a teoria Estatística se consolidou e passou a ser enunciada a partir de generalizações das propriedades observadas em grandes amostras [21, 22, 23].

Na primeira metade do século XX vários resultados de inferência foram demonstrados por Fisher e matemáticos como Jerzy Neyman, Egon Pearson e Abraham Wald. Neste mesmo

período deu-se o desenvolvimento dos delineamentos experimentais e levantamentos por amostragem, assim como as ideias fundamentais sobre séries temporais [21]. Na década de 1972, Nelder e Wedderburn propuseram a classe de modelos conhecida como modelos lineares generalizados (MLG), que além de incluir o modelo linear normal como caso particular, engloba os modelos de regressão logística e de Poisson, além dos modelos de análise de variância e covariância, dentre outros. A importância dos MLG's e das extensões propostas a partir desta classe vai além das diferentes possibilidades de aplicação. Por constituir uma abordagem que unifica vários modelos estatísticos, ela também promove o papel central da verossimilhança na teoria de inferência. Em 1977, o conhecido algoritmo EM (Expectation-Maximization), comumente utilizado para estimação dos parâmetros de diversos modelos estatísticos, foi proposto por Dempster e colaboradores como uma estratégia de maximização baseada na ideia de dados faltantes. Antes da década de 1990, contudo, os métodos baseados em Monte Carlo [28], como, por exemplo, as cadeias de Markov Monte Carlo (MCMC), ainda não haviam ganhado força na comunidade estatística e uma possível justificativa deve-se à falta de desenvolvimento computacional na época, apesar de já existir uma teoria consolidada dos algoritmos baseados em Metropolis-Hastings e no amostrador de Gibbs [25].

A partir do século XX, os métodos estatísticos foram sendo desenvolvidos, mesclando lógica, ciência e tecnologia para a solução e investigação de problemas em áreas diversas do conhecimento humano [24]. Uma destas áreas é a Epidemiologia, ramo da medicina que estuda os diferentes fatores que intervêm na difusão e propagação de doenças, e possui, desde o século XVII, uma estreita relação com a Estatística, quando o cientista e demógrafo britânico John Graunt, precursor na construção das tábuas de mor-

talidade, percebeu a importância da análise quantitativa dos chamados eventos vitais (nascimentos, óbitos e perdas fetais). Graunt usou dados de uma série temporal anual (de 1604 a 1660), coletados nas paróquias de Londres, e concluiu, dentre outras coisas, que a maior parte dos nascimentos eram de crianças do sexo masculino, apesar da distribuição por sexo ser homogênea na população geral, que havia altas taxas de mortalidade nos primeiros anos de vida e que a mortalidade era maior nas áreas urbanas em comparação às áreas rurais [21]. No século XIX, William Farr, considerado o primeiro médico estatístico do escritório de registros gerais da Inglaterra e país de Gales, fez uso do registro civil para o estudo de doenças e propôs uma classificação para as causas de morte, que anos mais tarde serviria como base estrutural para a atual classificação internacional de doenças. Foi, contudo, em meados do século XX, com a consagração da teoria de multicausalidade (segundo esta teoria, a causa de determinada doença não é única, ou seja, no seu aparecimento coexistem várias causas devidas a vários fatores causais), que os epidemiologistas começaram a adotar a Estatística como metodologia analítica em seus estudos [26].

Falando especificamente da modelagem de doenças infecciosas, é notório o crescente uso de modelos matemáticos determinísticos para representação de epidemias, como descrito nas seções anteriores. No entanto, os modelos estatísticos (ou estocásticos) têm ganhado bastante espaço nesta área, devido, dentre outros fatores, à melhora dos recursos computacionais nas últimas duas décadas, além do uso e disseminação de procedimentos de estimação baseados, por exemplo, em métodos MCMC. As ferramentas estatísticas para modelagem de doenças infecciosas podem fornecer evidências, em termos de probabilidade, se uma infecção está se espalhando ou sendo controlada em uma população, se as taxas de infecção variam de acordo com fa-

tores demográficos importantes, ou ainda se determinadas intervenções de saúde pública estão ou não tendo o impacto desejado [27]. As abordagens propostas na literatura vão desde a aplicação de MLG's (ou extensões desta classe), a utilização de modelos usuais de séries temporais, bem como suas extensões para respostas não normais, ao uso de modelos espaciais para mapeamento de doenças, que consistem, por exemplo, em encontrar padrões (correlações) espaciais da doença sob estudo. Mais recentemente, a pandemia causada pelo novo coronavírus trouxe à tona a importância das modelagens estatística e matemática dentro deste cenário, e muitas aplicações baseadas, por exemplo, em modelos de crescimento logístico, modelo de Poisson com sobredispersão, modelos de pontos de mudança (change-point models), dentre outros, têm sido propostas na tentativa de modelar a curva de novos casos e/ou óbitos causados pela COVID-19.

É importante destacar que as metodologias existentes na literatura para modelagem estatística geral e, em especial, a modelagem de doenças infecciosas, possuem pressupostos, além de pontos positivos e negativos, e tais fatores precisam ser levados em consideração na pesquisa científica. Vale ressaltar que no caso de doenças infecciosas, o indivíduo depende de outros indivíduos ao seu redor para se tornar infectado e dificilmente se conhece o vetor transmissor, ou seja, por quem a pessoa foi infectada, nem o tempo exato em que o sujeito ficou infeccioso. Dessa forma, a qualidade e a disponibilidade da informação é fundamental para que determinadas hipóteses científicas possam ser verificadas. Por fim, outro ponto importante é que dados apropriados precisam estar disponíveis para análise, pois sem eles, as hipóteses de interesse podem não ser adequadamente testadas.

Modelagem computacional: panorama histórico

e papel na pandemia

por ROBESPIERRE PITA E DANIELA ALMEIDA

A definição mais simples para o termo "modelagem computacional" corresponde a descrição de um fenômeno através de uma linguagem de programação [29]. Desta maneira, o esforço para criar um modelo computacional passa por observar uma porção do mundo real, descrever as interações entre seus diferentes componentes e tentar expressá-las em dados e operações que possam ser usados por um computador. Vamos a um exemplo simples. Suponha que você deseja desenvolver um simulador para cobrança de pênaltis. Assim, você sabe que precisará levar em conta diversas características para cada cobrança (gravidade, força e direção do vento, modelo e material da chuteira, tamanho e peso da bola, força do chute, etc) e que diversos modelos matemáticos deverão ser empregados para simular bem cada cenário de acordo com as intenções do jogador e goleiro. É muito provável que estes modelos sejam complexos demais para permitir uma solução analítica legível ou rápida, e, por isso, derivar seus passos em operações computacionais e dispor suas variáveis em estruturas de dados talvez seja a única forma de construir o seu simulador. A escrita deste programa deve garantir um comportamento de direção e efeito da bola que correspondam aos comandos do jogador e um resultado que também depende dos comandos do goleiro, sem mencionar que a qualidade e aceitação deste programa estão muito associados a modelagem matemática usada para reproduzir as leis da física em cada chute. É importante pontuar que a relação entre as modelagens matemática e computacional são ainda mais amplas e profundas que a demonstrada neste exemplo.

Com uma complexidade e volume absurda de cálculos necessários para a construção de bombas nucleares no Projeto Manhattan, o matemático John von Neumann participou da

criação do *Electronic Numerical Integrator and Computer* (ENIAC - em português: computador integrador numérico eletrônico) em 1944, primeiro computador eletrônico que trazia em sua arquitetura componentes similares aos computadores de hoje: uma unidade de memória, uma unidade de processamentos e periféricos de entrada e saída [30]. John von Neumann, em 1945, também se envolveu no esboço do primeiro computador que executava programas previamente armazenados em memória, o *Electronic Discrete Variable Automatic Computer* (ED-VAC - em português: Computador Eletrônico de Variáveis Discretas). Este segundo projeto foi bastante influenciado pela máquina de Turing universal, proposta por Alan Turing em 1936. Turing foi o matemático e criptoanalista responsável por criar um computador capaz de decifrar as mensagens codificadas que eram trocadas pelos Nazistas durante a segunda guerra mundial através da Máquina Enigma. Hoje, Alan Turing é considerado o pai da Ciência da Computação. Este breve histórico já deve ser suficiente para que possamos entender o quanto a modelagem computacional é íntima da matemática desde sua gênese. Perceba que até mesmo as arquiteturas dos computadores clássicos foram desenhadas para atender aos propósitos da modelagem matemática. Desde então a modelagem computacional se tornou uma área do conhecimento muito relevante, principalmente na resolução de problemas não-lineares complexos.

Hoje os cientistas da computação refinam e aplicam os fundamentos da sua área para tentar responder diversas questões científicas que irão contribuir com o nosso conhecimento sobre as relações entre fenômenos do mundo real - tais como naturais ou sociais. A eficácia destes modelos é maior principalmente nos contextos em que existem estruturas matemáticas capazes de permitir abstração, generalização e interpretação [29], não sendo eles, contudo, dependentes destas estruturas. Os estudos

que mais atraem os interesses hoje estão inseridos no contexto de Ciência de Dados e Big Data, temas que impulsionam a evolução da modelagem computacional para lidar com estruturas de dados cada vez maiores e mais complexas. A contribuição da ciência da computação nestes tópicos reforça ainda mais a importância desta epistemologia e faz transcender as aplicações de suas descobertas. Não é difícil encontrar exemplos práticos advindos destes estudos em nosso cotidiano: pesquisas que visam entender a relevância ou popularidade numa rede social [31], robôs de conversação (*chatbots*) capazes de ser confundidos com pessoas reais [32, 33], reconhecimento facial [34], entre outros. O advento da Ciência de Dados impulsionou ainda mais o uso do aprendizado de máquina para resolver questões do nosso cotidiano - automação de tarefas, predição de eventos, diagnósticos, etc. -, e existe muita colaboração da ciência da computação nisso.

Machine Learning (em português: aprendizado de máquina) é um ramo da Inteligência Artificial. Pode ser classificado como modelo computacional de aprendizado de máquina todo programa que melhora o desempenho em realizar uma determinada tarefa ao passo que sua experiência aumenta. Esta experiência pode ser caracterizada por um dado de treino maior. Considere como exemplo prático um programa capaz de jogar cartas que vence mais de acordo com o número de partidas que joga. Embora os fundamentos destes programas estejam alicerçados na matemática e estatística, a modelagem computacional conta com duas colaborações essenciais. A primeira é implementação de programas cada vez mais otimizados e escaláveis, ou seja, com tempo de execução

reduzido ou capazes de lidar com grandes quantidades de dados em múltiplos processadores ou computadores. A segunda está na idealização e implementação de modelos computacionais inovadores e cada vez mais complexos, como o Deep Learning (em português: aprendizagem profunda). Modelos de Deep Learning simulam o funcionamento de nossas células cerebrais, montando redes neurais profundas, dispostas em camadas hierarquicamente interconectadas que conduzem o dado de entrada a diversas derivações para permitir a predição de eventos futuros ou agrupamento dos dados atuais. É comum utilizar estas técnicas para análise de discurso, reconhecimento de fala e reconhecimento visual, por exemplo. Não seria surpresa que, assim como as ciências matemática, estatística e de saúde, os modelos computacionais pudessem ser utilizados para contribuir ativamente no entendimento e monitoramento da pandemia do novo coronavírus (Sars-CoV-2).

Dezenas de milhares de artigos científicos tornaram público os esforços de vários cientistas da computação que, em colaboração com pesquisadores da matemática, estatística e saúde, se debruçaram sobre temas relacionados ao estado de emergência causados pela pandemia para obter conhecimento sobre sua dinâmica e salvar vidas. Em resposta às limitações de diagnóstico em massa para monitoramento dos infectados, o domínio e uso das técnicas de machine learning permitiram que dados clínicos fossem utilizados para prever quem poderia estar doente mesmo sem apresentar nenhum sintoma [35, 36]. A falta de testes também impulsionou que Redes Neurais Profundas fossem montadas para apoiar o diag-

nóstico através do reconhecimento de padrões em exames de imagem dos pulmões [37]. Ou, para evitar uma pressão ainda maior ao sistema de saúde, modelos de Deep Learning foram utilizados para prever o risco de um paciente precisar de internação em enfermaria ou UTI [38]. Muitas outras iniciativas estão sendo publicadas a cada instante, existem até potenciais colaborações na corrida pela vacina [39].

O potencial da ciência da computação na modelagem de doenças infecciosas é enorme. O avanço das técnicas utilizadas no monitoramento e resposta aos eventos do cotidiano, especialmente durante pandemia, também é considerável. As colaborações e o trabalho em rede dos pesquisadores das mais diversas áreas do conhecimento, porém, foram fundamentais para alavancar o conhecimento e diminuir as nossas limitações. É de estrita importância ressaltar que ainda há muito o que se entender e que muitas outras questões sobre este mesmo tema podem vir à tona, trazendo desafios ainda maiores. Precisamos estar prontos para enfrentar a ignorância com ainda mais conhecimento.

Agradecimentos

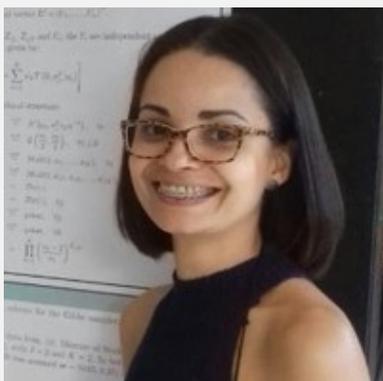
O projeto tem apoio do Programa Fiocruz de Fomento à Inovação: ideias e produtos inovadores - COVID-19, encomendas e estratégias - INOVA-FIOCRUZ (Processo VPPIS-005-FIO-20-2-40). Trabalho de colaboração entre o "Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS)" e a "Fundação Oswaldo Cruz de Rondônia". Revisão: Paloma Fonseca de Oliveira. Divulgação científica: Raíza Tourinho dos Reis Silva Lima e Karina de Souza Costa.

Biografia dos convidados



Natália Machado Tavares

Possui graduação em Ciências Biológicas (UFBA, 2005), mestrado (UFBA-FIOCRUZ/BA, 2008) e doutorado (UFBA-FIOCRUZ/BA, 2013) em Patologia Experimental. Participou da equipe coordenada pelo Dr. Nicolas Glaichenhaus no doutorado sanduíche (Bolsista CNPq-SWE, 2012) na área de Imunidade de Mucosas e Inflamação (IPMC-Université de Nice/França). Fez pós-doutorado no CPqGM-FIOCRUZ/BA, investigando a interação entre células dendríticas e neutrófilos humanos na leishmaniose cutânea. Também participou da equipe coordenada pelo Dr. Carlos Henrique Serezani durante pós-doutorado na Indiana University - Purdue University of Indiana (Indianapolis, EUA). Atualmente é Pesquisadora do Instituto Gonçalo Moniz (FIOCRUZ-BA), atuando em temas relacionados à imunologia, biologia celular, molecular, parasitologia e análise de dados globais, com ênfase no papel da imunidade inata. Tem desenvolvido projetos nos temas: microRNAs expressos em doenças cutâneas, comorbidades na Leishmaniose humana e receptores de ativação da imunidade inata. Endereço para acessar este CV: <http://lattes.cnpq.br/8380706398761910>



Nivea Bispo da Silva

Possui graduação (UFBA, 2007) em Estatística, mestrado (UFMG, 2013) e doutorado (UFMG, 2017) em Estatística. Atual professora adjunta A e pesquisadora no Departamento de Estatística da Universidade Federal da Bahia. É colaboradora no Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), da Fundação Oswaldo Cruz, desenvolvendo pesquisas na plataforma da Coorte de 100 milhões de brasileiros. Possui mais de dez anos de experiência em pesquisa na área de Probabilidade e Estatística, com ênfase em Estatística Aplicada. Atua principalmente com os seguintes temas: Análise de dados longitudinais, Análise de dados correlacionados, Modelos de mistura finita e Estatística Bayesiana. Endereço para acessar este CV: <http://lattes.cnpq.br/2365687224418330>



Rosemeire Leovigildo Fiaccone

Possui graduação em Bacharelado Em Estatística (UFBA, 1988), mestrado em Estatística (UNICAMP, 1998) e doutorado em Phd In Statistics - Lancaster University (2006). Atualmente é Professora Associada II e pesquisadora da Universidade Federal da Bahia e colaboradora no Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), da Fundação Oswaldo Cruz, desenvolvendo pesquisas na plataforma da Coorte de 100 milhões de brasileiros. Com mais de 30 anos de carreira, tem experiência na área de Probabilidade e Estatística, com ênfase em Bioestatística, atuando principalmente nos seguintes temas: análise de dados longitudinais, análise de dados categorizados correlacionados e modelos de equações estruturais. Endereço para acessar este CV: <http://lattes.cnpq.br/1839882342448396>



Robespierre Dantas da Rocha Pita

Possui Bacharelado (Universidade Salvador, 2010) em Sistemas de Informação, mestrado (UFBA, 2015) e doutorado (UFBA, 2019) em Ciência da Computação. É especialista em Redes de Computadores e professor substituto e convidado na Universidade Federal da Bahia, Serviço Nacional de Aprendizagem Industrial da Bahia, Faculdade Maurício de Nassau e na Universidade de Salvador. Atualmente é pesquisador no Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS) da Fundação Oswaldo Cruz, atuando nas áreas de Sistemas Distribuídos, Aprendizado de Máquina e Ciência de Dados, com particular ênfase na área de Computação Aplicada à Saúde. Endereço para acessar este CV: <http://lattes.cnpq.br/9031151666715654>



Daniela Santos Almeida

Graduada em Sistemas Para Internet pela Universidade Salvador - UNIFACS (2016). Almeida foi premiada com Diploma de Honra ao Mérito devido ao seu desempenho acadêmico de excelência em 2017 e duas vezes obteve o Prêmio Mérito Acadêmico em 2014 e 2015 na UNIFACS. Já produziu pesquisa e tecnologias na plataforma da Coorte de 100 milhões de brasileiros e na Plataforma de vigilância de longo prazo para a Zika, no Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS) da Fundação Oswaldo Cruz. Atualmente é Mestranda em Ciência da Computação pela Universidade Federal da Bahia, desenvolvendo pesquisa sobre Processamento de Linguagem Natural, estando também em processo de Especialização em Big Data e Business Intelligence. É engenheira de Dados Sênior na Magnetis Gestora de Investimentos. Tem experiência na área de Ciência da Computação, com ênfase em Engenharia e Ciência de Dados. Endereço para acessar este CV: <http://lattes.cnpq.br/9651739731722097>

References

- [1] Barreto ML, Teixeira MG, Carmo EH. Infectious diseases epidemiology. *Journal of Epidemiology Community Health*. Mar 1;60(3):192-5, 2006.
- [2] Saker, L., Lee, K., Cannito, B., Gilmore, A., Campbell-Lendrum, D. H. Globalization and infectious diseases: a review of the linkages. (No. TDR/STR/SEB/ST/04.2). World Health Organization, 2004.
- [3] Holmes, King K., Stefano Bertozzi, Barry R. Bloom, Prabhat Jha, Hellen Gelband, Lisa M. DeMaria, and Susan Horton. Major infectious diseases: key messages from disease control priorities. 2017.
- [4] Smith, David and Keyfitz, Nathan. *Mathematical demography: selected papers*. Springer Science & Business Media, Vol. 6., 2012.
- [5] Fenner, Frank, Donald Ainslie Henderson, Isao Arita, Zdenek Jezek, and Ivan D. Ladnyi. *Smallpox and its eradication*. Vol. 6. Geneva: World Health Organization, 1988.
- [6] Bacaër, Nicolas. Daniel Bernoulli, d'Alembert and the inoculation of smallpox (1760). In *A short history of mathematical population dynamics*, pp. 21-30. Springer, London, 2011.

- [7] Bernoulli, Daniel, and Sally Blower. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Reviews in medical virology* 14, no. 5:275, 2004.
- [8] WHO. Cholera: The Forgotten Pandemic. <https://www.who.int/cholera/the-forgotten-pandemic/en/>, 2018.
- [9] Snow, John. On the mode of communication of cholera. John Churchill, 1855.
- [10] Smith, David R. Herd Immunity. *The Veterinary clinics of North America. Food animal practice* 35, no. 3:593-604, 2019.
- [11] Fine, Paul EM. Herd immunity: history, theory, practice. *Epidemiologic reviews* 15, no. 2:265-302, 1993.
- [12] McNabb, S. J., R. A. Jajosky, P. A. Hall-Baker, D. A. Adams, P. Sharp, W. J. Anderson, A. J. Javier et al. Summary of notifiable diseases—United States, 2005. *MMWR. Morbidity and mortality weekly report* 54, no. 53:1, 2007.
- [13] Fine, Paul, Ken Eames, and David L. Heymann. Herd immunity: a rough guide. *Clinical infectious diseases* 52, no. 7:911-916, 2011.
- [14] Foppa, Ivo M. *A Historical Introduction to Mathematical Modeling of Infectious Diseases: Seminal Papers in Epidemiology*. Academic Press, 2016.
- [15] En'ko, P.D. (1889). On the course of epidemics of some infectious diseases. *Vrach. St. Petersburg*, X, 1008-10010, 1039-1042, 1061-1063 (in Russian). English translation by Dietz, K. *International Journal of Epidemiology* 18, 749-755, 1989.
- [16] Hamer, William Heaton. The Milroy lectures on epidemic disease in England; the evidence of variability and of persistency of type. <https://archive.org/details/milroylectureson00hameuoft/page/26/mode/2up>, 1906.
- [17] Ross, Sir Ronald. *Memoirs with a full account of The Great Malaria Problem and its Solution*. Albemarle Street, W. London: John Murray. p. <https://archive.org/details/b29825738/page/n13/mode/2up>, 1923.
- [18] Amad, Alan, Aureliano Sancho Souza Paiva, Caio Porto de Castro, Daniel Cardoso Pereira Jorge, Diego Santos Souza, Elaine Cristina Cambui Barbosa, Gabriel Bertolino et al. Boletim CoVida – Acompanhamento da pandemia por Covid-19 no Brasil: destaque para a situação na Bahia. <https://redecovida.org/relatorios/boletim-covida/>, 2020.
- [19] Amad, Alan, Aureliano Sancho Souza Paiva, Caio Porto de Castro, Daniel Cardoso Pereira Jorge, Diego Santos Souza, Elaine Cristina Cambui Barbosa, Gabriel Bertolino et al. Boletim CoVida – Pandemia de Covid-19. <https://redecovida.org/relatorios/boletim-covida-ed-02/>, 2020.
- [20] Amad, Alan, Aureliano Sancho Souza Paiva, Caio Porto de Castro, Daniel Cardoso Pereira Jorge, Diego Santos Souza, Elaine Cristina Cambui Barbosa, Gabriel Bertolino et al. Boletim CoVida: Pandemia de COVID-19: fortalecer o Sistema de Saúde para proteger a população. https://www.arca.fiocruz.br/bitstream/icict/41472/2/boletim_4_rede_covida_final.pdf, 2020.
- [21] Memória, J. M. P. *Breve história da Estatística*. Embrapa - Informação Tecnológica, 2004.
- [22] Stanton, J. Galton, Pearson, and the Peas. A Brief History of LiM. near Regression for Statistics Instructors. *Journal of Statistics Education*, vol 9, issue 3, 2001.
- [23] Stigler, S. The Epic Story of Maximum Likelihood. *Statistical Science*, vol. 22, n 4, 2007.
- [24] Stigler, S. M. *The history of Statistics: The measurement of uncertainty before 1900*. Harvard University Press, 1986.
- [25] Brooks, S. , Gelman, A., Jones, G. L., Meng, Xiao-Le M. *Handbook of Markov Chain Monte Carlo*. Chapman Hall / CRC press, 2011.
- [26] Szwarcwald, C. L., Castilho, E. A. de. Os Caminhos da Estatística e suas incursões pela Epidemiologia. *Cadernos de saúde pública*, vol 8, p. 05-21, 1992.
- [27] Grant, B., Ozanne, M. Statistical Models for Infectious Diseases: A Useful Tool for Practical Decision-Making. *American Journal of Tropical Medicine and Hygiene*, vol 10, issue, 2019.

- [28] Metropolis, N., Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association*. v. 44, n. 247, p. 335-41, Sep. 1949.
- [29] Mahoney, MS Historical perspectives on models and modeling. XIIIth DHS-DLMPS joint conference on “scientific models: Their historical and philosophical relevance, 2000.
- [30] Stallings, William *Arquitetura e Organização de Computadores 8a Edição*. Prentice-Hall, Pearson, 2010.
- [31] Gabardo, Ademir C *Análise de redes sociais: uma visão computacional*. Novatec Editora, 2015.
- [32] Kuyven, Neiva Larisane and Antunes, Carlos André and de Barros Vanzin, Vinicius João and da Silva, João Luis Tavares and Krassmann, Aliane Loureiro and Tarouco, Liane Margarida Rockenbach. *Chatbots na educação: uma Revisão Sistemática da Literatura*. *RENTE-Revista Novas Tecnologias na Educação*, 16, 1, 2018.
- [33] Maeda, A and Moraes, S. Chatbot baseado em deep learning: um estudo para língua portuguesa. *Symposium on Knowledge Discovery, Mining and Learning*, 5th, 2017.
- [34] Dubey, Arun Kumar and Jain, Vanita. A review of face recognition methods using deep learning network. *Journal of Information and Optimization Sciences*, Taylor & Francis, 40, 2, 547–558, 2019.
- [35] Banerjee, Abhirup and Ray, Surajit and Vorselaars, Bart and Kitson, Joanne and Mamalakis, Michail and Weeks, Simonne and Baker, Mark and Mackenzie, Louise S. Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population. *International immunopharmacology*, 86, 106705, Elsevier, 2020.
- [36] Batista, AF d M and Miraglia, JL and Donato, THR and Chiavegatto Filho, ADP and de Moraes Batista, André Filipe and Miraglia, João Luiz and Donato, Thiago Henrique Rizzi and Chiavegatto Filho, Alexandre Dias Porto. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *Hospital Israelita Albert Einstein-Big Data Analytics M*, São Paulo, SP, Brazil., Department of Epidemiology SoPH, University of Sao Paulo, Sao Paulo, Brazil, editors, 2020.
- [37] Harmon, Stephanie A and Sanford, Thomas H and Xu, Sheng and Turkbey, Evrim B and Roth, Holger and Xu, Ziyue and Yang, Dong and Myronenko, Andriy and Anderson, Victoria and Amalou, Amel and others. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nature communications*, 11, 1, 1–7, Nature Publishing Group, 2020.
- [38] Liang, Wenhua and Yao, Jianhua and Chen, Ailan and Lv, Qingquan and Zanin, Mark and Liu, Jun and Wong, SookSan and Li, Yimin and Lu, Jiatao and Liang, Hengrui and others. Early triage of critically ill COVID-19 patients using deep learning. *Nature communications*, 11, 1, 1–7, 2020.
- [39] Kannan, Shantani and Subbaram, Kannan and Ali, Sheeza and Kannan, Hemalatha. The role of artificial intelligence and machine learning techniques: Race for covid-19 vaccine. *Archives of Clinical Infectious Diseases*, 15, 2, Kowsar, 2020.