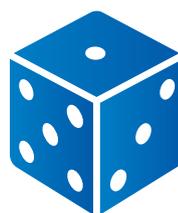


# Boletim



# SPE

Sociedade Portuguesa  
de Estatística

Publicação semestral

primavera de 2021



## Especial Covid: a Estatística ao serviço da sociedade

### O Papel da Estatística no Combate à Pandemia COVID-19

João A. Branco ..... 17

### De passagem pela mortalidade...

Clara Cordeiro ..... 22

### Algumas considerações sobre a COVID-19

Isabel Natário ..... 25

### CoronaSurveys na estimação de casos ativos de Covid-19

Raquel Menezes e Carlos Baquero ..... 35

### Estatística no CEAUL em tempos de COVID

Tiago A. Marques e Lisete Sousa ..... 40

### Indicadores para a pandemia COVID-19 em Portugal/ Uma leitura do contexto demográfico e da expressão territorial

Instituto Nacional de Estatística ..... 43

Editorial .....	2
Mensagem do Presidente .....	4
Notícias .....	6
<i>Enigmística</i> .....	12
Episódios na História da Estatística .....	13
SPE e a Comunidade .....	15
Especial Covid: a Estatística ao serviço da sociedade .....	17
Ciência Estatística .....	45
Prémio SPE 2020 .....	46

### Informação Editorial

**Endereço:** Sociedade Portuguesa de Estatística,  
Campo Grande. Bloco C6. Piso 4.  
1749-016 Lisboa. Portugal.

**Telefone:** +351.217500120

**e-mail:** [spe@spestatistica.pt](mailto:spe@spestatistica.pt)

**URL:** <https://www.spestatistica.pt>

**ISSN:** 1646-5903

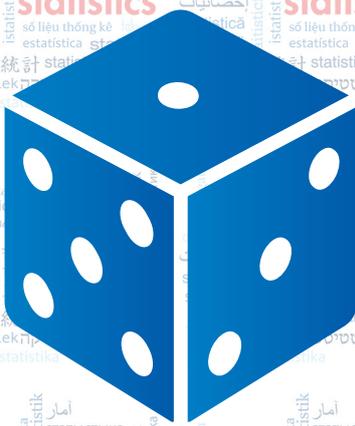
**Depósito Legal:** 249102/06

**Tiragem:** Edição digital

**Execução Gráfica e Impressão:** Gráfica Sobreirese

**Editor:** Fernando Rosado, [fernando.rosado@fc.ul.pt](mailto:fernando.rosado@fc.ul.pt)

Sociedade Portuguesa de Estatística desde 1980



# SPE

## Sociedade Portuguesa de Estatística



### XXV Congresso

## Sociedade Portuguesa de Estatística

### 2021 Évora

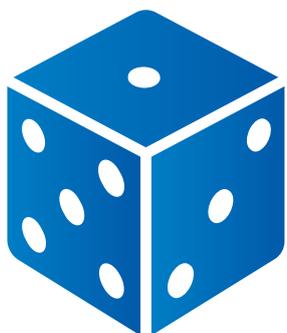


# SPE 2021

XXV Congresso da Sociedade Portuguesa de Estatística  
13 a 16 de Outubro de 2021 | Évora

<http://www.spe2021.uevora.pt/>

# Junta-te à



# SPE

Sociedade Portuguesa  
de Estatística

*“Se já és sócio da SPE, incentiva os teus colegas,  
colaboradores e alunos a juntarem-se à SPE”*

## A SPE

- Oferece descontos em congressos e outros eventos organizados pela SPE.
- Oferece distinções e reconhecimento através dos seus prémios.
- Oferece oportunidades para ampliares a tua rede de contactos através da comunidade SPE.
- Oferece aos sócios acesso a um sistema de acreditação internacional.
- Valoriza sócios, comunidade e profissão apostando na educação, formação e inovação.
- Defende a profissão e molda o seu futuro.

## Junta-te à SPE:

<https://www.spestatistica.pt/socios/admissao-formulario>

## Quota anual\*

- Regular: 30€
- Estudante: 15€
- Promotor: 0€
- Recém-licenciado<sup>†</sup>: grátis!

<sup>†</sup> Até um ano após terminar a licenciatura.

\* Um “sócio promotor” angaria 2 novos sócios por ano.

# Editorial

## ... e, como Covid-19 requereu o aumento da Literacia Estatística...

Este Boletim SPE primavera 2021 surge em “tempos novos”.

É tempo novo, sempre que se realiza mais um passo de uma continuidade que, no caso da SPE e no presente, se manifesta através da entrada em funções dos novos Órgãos Sociais. É um momento forte festejado pelo rejuvenescimento que representa e também porque é um acontecimento que produz frutos: novas ideias, novos projetos e novas visões com novas equipas. É este tempo novo que também propicia maior disponibilidade para agregar vontades de todos os que, como apanágio da SPE, se reúnem em torno dos novos órgãos para mais e ainda melhor sucesso da Sociedade Portuguesa de Estatística; a associação científica cuja missão, em Portugal e no mundo, é patrocinar a Ciência Estatística e os estatísticos.

É tempo novo, aos mais diversos níveis, políticos, sociais e científicos, em todo o percurso mundial vivido nos últimos meses. O próprio Boletim outono 2020, os seus autores e editor, implicitamente admitiam que a edição seguinte seria a jeito de uma revisão de algo que (no seu pleno) já estaria ultrapassado e de tudo o que teria permitido tirar conclusões. A realidade, a incerteza estatística, de novo comandou e o caminho foi diferente; mais lento no ramo solucionador que mais interessa. E aqui, de facto, uma grande conclusão podemos tirar: a exigência de um maior reforço e intervenção da Ciência Estatística e, por consequência, também da SPE. Nesse sentido as propostas de atuação programática da nova Direção SPE são indicadores apelativos para toda a comunidade e, cada vez mais e com mais presença, a assumir como prioridade a divulgação da Estatística. O tempo novo que se vive, veio salientar, além da maior importância, a necessidade do aumento da capacidade de “ler e escrever”; mas, também e principalmente, de “perceber e interpretar” o que é lido e escrito sobre “dados e indicadores estatísticos”.

De igual modo, reforça e confirma essa importância, a Notícia sobre o Projeto “Competição Internacional de *Posters* dinamizada pelo ISLP (*International Statistical Literacy Project*)” que nesta edição publicamos. O ISLP é um projeto iniciado pelo IASE, secção para a educação estatística constituída no âmbito do *International Statistical Institute* (ISI) do qual a SPE é membro filiado.

Muitos membros da nossa comunidade académica têm desenvolvido as mais diversas intervenções com esse objetivo; mas a urgência e a necessidade de mais ação é palavra de ordem.

É tempo novo, o modo como a comunidade tem aceite e, de algum modo, solicitado aos estatísticos uma maior intervenção que, evidentemente também, é ao mesmo tempo um enorme desafio. Este inicia-se com uma dicotomia, uma oposição tão simples quanto desafiadora, entre “esclarecer sobre a incerteza” enquanto a comunidade “requer indicadores exatos”.

O Tema Central do Boletim, habitualmente, é planeado, às vezes, quase um ano antes. O Título escolhido para este Boletim primavera 21 “com alguma ingenuidade?” mas com o convicto desejo de “melhores dias” admitia implicitamente, sair num futuro onde Covid-19 tivesse sido “um item” do passado onde a Estatística (ainda mais) se afirmara.

Assim não aconteceu.

Tudo decorreu de um modo mais agitado e o tempo mostrou que mais algum tempo deve passar até um tempo de mais tranquilidade.

De facto, a Estatística também aqui ajudou a consolidar a metodologia de “combate à pandemia” mas também confirmou que “o prazo” (o tempo) é “uma (co)variável” a “não perder de olho”. O tempo desde a anterior edição mostrou que devemos continuar um pouco mais nesse tema. Assim, para além

das contribuições ora incluídas no Tema Central fica o convite para outros autores, que desejem incluir as suas contribuições, na secção SPE e a Sociedade do próximo Boletim SPE outono 21.

O Cardeal Tolentino de Mendonça, no seu discurso ao receber o Prémio da Universidade de Coimbra 2021, em fevereiro deste ano, sintetizou de uma forma brilhante, os grandes desafios à Ciência e com base num passado e em alguns pontos que Covid já fez realçar: “A crise pandémica vem desativar muitas modalidades de construção do real e dizer que estão ultrapassadas. Sobre o conhecimento, porém, ela vem reforçar a relevância. As sociedades do futuro terão de potenciar sempre mais a importância e a centralidade do conhecimento”, afirmou.

Eis um novo (velho) desafio em, (e para) tempos novos!

Eis como, Covid-19 também requer um aumento da Literacia Estatística!

O Boletim da SPE, ao longo da sua já longa história de 50 anos, tem conhecido uma evolução característica de uma edição que se deseja afirmar o mais possível junto dos seus leitores e, para tal, já conheceu diversas etapas que, progressivamente, permitiram chegar àquela “bonita idade” com a satisfação de “dever cumprido”. As várias equipas que lhe têm dado vida com o esforço e a dedicação manifestada são o garante de uma continuidade sempre melhorada. O Boletim SPE é uma publicação científica generalista de divulgação de Ciência que deseja estar cada vez mais próximo dos seus leitores. Para podermos ainda melhor partilhar os desafios e colaborar na sua concretização para além do habitual coeditor de cada Tema Central e que desde há vários boletins tem sido um precioso meio de melhorar; também parece muito oportuno que tenhamos “coeditores fixos”. A tarefa editorial poderá assim ser mais partilhada para, sempre mais e melhor, podermos cumprir os objetivos estatutários da SPE: promover, cultivar e desenvolver em Portugal, o estudo da Estatística, suas aplicações e Ciências afins (nº 2 do art 1º).



O Tema Central do próximo Boletim SPE será:

*Machine Learning* e Inteligência Artificial

# Mensagem do Presidente

Caros Sócios da SPE

Gostava antes de mais de vos dizer que estou profundamente agradecido a todos vós pelo vosso constante apoio, dedicação e empenho à SPE. Esta é a minha primeira “Mensagem do Presidente”; aproveito assim para agradecer o voto de confiança aos sócios da SPE por elegerem a lista atual de Órgãos Sociais e para agradecer às anteriores Direções pelo seu papel de liderança e contribuições para a nossa comunidade. Connosco houve quem acreditasse que é tempo de fazer mais e melhor numa série de áreas estratégicas. Esta é a vossa vitória; esta é a oportunidade de moldar o futuro.

Tenho algumas novidades para partilhar, mas antes disso tenho um pedido muito importante. Por favor:

*“Incentiva os teus colegas, colaboradores e alunos a juntarem-se à SPE”*

A SPE está em constante transformação e o contributo de todos faz a diferença. A SPE oferece oportunidades, descontos e distinções à sua comunidade, valoriza os seus sócios, defende a nossa profissão e molda o seu futuro. Posso contar com a tua ajuda?

Eis agora alguns outros pontos que gostava de partilhar. O Congresso da SPE (SPE 2021; [www.spe2021.uevora.pt/](http://www.spe2021.uevora.pt/)) vai decorrer on-line; se gostas do Mundo dos Dados este é o teu Congresso: Passa a palavra! Ambicionamos usar oportunidade para fazer com que o Congresso seja o mais amplo e diverso possível.

Gostava de notar ainda que a SPE começou a usar a rede social Twitter como plataforma de comunicação por forma a dinamizar o contacto com o Mundo Estatístico Lusófono, bem como para contribuir para a literacia Estatística:

<https://twitter.com/spestatistica>

Em particular, a SPE vai combater as afrontas diárias à nossa profissão através do hashtag [#PoligrafoEstatistico](https://twitter.com/PoligrafoEstatistico). A pandemia destaca o papel fundamental da Estatística, mas é altura de recordar que a produção e comunicação de estatísticas envolve uma enorme responsabilidade e sentido ético. Sugiro que nos sigas no Twitter para que possas ficar a par dos desenvolvimentos mais recentes neste espaço.

Estamos também em vias de colocar em funcionamento o processo de acreditação de estatísticos profissionais em colaboração com a FENStatS (*Federation of European National Statistical Societies*), o qual é mutuamente reconhecido pela ASA (*American Statistical Association*).

E por forma a elevar o perfil e a visibilidade da “Nossa Estatística” no Mundo, a SPE tem patrocinado recentemente uma série de eventos internacionais, tais como o **EVA 2021**; esta conferência internacional teve origem há quase 40 anos em Vimeiro, Portugal (1983) e é o fórum de discussão mais importante em Estatística de Valores Extremos. A SPE patrocina ainda em conjunto com a Ordem dos Médicos os eventos: *3rd Statistics on Health Decision Making* (U. Aveiro); *Interfaces Between Statistics, Machine Learning and AI* (U. Edimburgo).

Termino com uma reflexão: O que é uma Sociedade, por definição? Será um conjunto de sócios? O conceito Matemático de *conjunto* é rico e profundo, mas uma sociedade, como a SPE, é muito mais sofisticada do que um mero *conjunto*. Somos uma comunidade, um grupo de peritos unidos, de mão dada, com um objetivo e denominador comum:

*“promover, cultivar e desenvolver, o estudo da Estatística, suas aplicações e ciências afins”*

Até breve,

Edimburgo, 12 de Abril de 2021

Cordais saudações

Miguel de Carvalho

# Notícias

## • Novos Órgãos Sociais da Sociedade Portuguesa de Estatística

O Processo Eleitoral para o triénio 2021-2023, realizou-se no dia 7 de dezembro de 2020.

A partir das 12.30, decorreu a Assembleia Eleitoral de apuramento final da votação eletrónica para eleição dos novos Órgãos Sociais da SPE para o mandato 2021-2023.

A votação eletrónica decorreu entre as 18.00 horas do dia 28 de novembro e as 18.00 horas do dia 5 de dezembro, com os seguintes resultados:

nº total de inscritos – 217

nº votantes – 164 (75,6%)

abstenção – 53 (24,4%)

### Mesa Assembleia Geral

Lista A – 85

Lista B – 76

brancos – 3

**Vencedora: Lista A – 85 (52,8%)**

### Direção

Lista A – 93

Lista B – 66

brancos – 5

**Vencedora: Lista A – 93 (58,5%)**

### Conselho Fiscal

Lista A – 83

Lista B – 76

brancos – 5

**Vencedora: Lista A – 83 (52,2%)**

### Mesa da Assembleia Geral

- **Presidente:** Pedro Oliveira (Universidade do Porto – ISP)
- **Primeiro Vogal:** Giovanni Loiola da Silva (Universidade de Lisboa – IST)
- **Segundo Vogal:** Soraia Pereira (Universidade de Lisboa – FCUL)

### Direção

- **Presidente:** Miguel de Carvalho (University of Edinburgh – SoM)
- **Vice-Presidente:** Paulo M. M. Rodrigues (BdP & NOVA School of Business and Economics)
- **Primeiro Vogal:** Regina Bispo (Universidade Nova de Lisboa – Departamento de Matemática)
- **Segundo Vogal:** Tiago Marques (University of St Andrews & Universidade de Lisboa)
- **Tesoureira:** Maria João Polidoro (Instituto Politécnico do Porto – E. S. de Tecnologia e Gestão)

### Conselho Fiscal

- **Presidente:** Vanda Lourenço (Universidade Nova de Lisboa – Departamento de Matemática)
- **Primeiro Vogal:** Conceição Ribeiro (Universidade do Algarve – I. Superior de Engenharia)
- **Segundo Vogal:** Paula Pereira (Instituto Politécnico de Setúbal – Escola Superior de Tecnologia)

## • Tomada de Posse dos Novos Órgãos Sociais da SPE



### A Direção da SPE para o triénio 2021 - 2023



### Tomada de posse com o auxílio das modernas tecnologias

## • XXV Congresso da Sociedade Portuguesa de Estatística

O XXV Congresso da SPE decorrerá em modelo virtual, de 13-16 de Outubro de 2021.



### Comissão Científica

- Miguel de Carvalho (**Presidente**) | Universidade de Edimburgo
- Fátima Ferreira | Universidade de Trás-os-Montes e Alto Douro
- João Andrade e Silva | Universidade de Lisboa
- Luís Meira-Machado | Universidade do Minho
- Marco Costa | Universidade de Aveiro
- Maria Eduarda Silva | Universidade do Porto
- Marília Antunes | Universidade de Lisboa
- Paula Brito | Universidade do Porto
- Regina Bispo | Universidade Nova de Lisboa
- Rosário Oliveira | Universidade de Lisboa
- Russel Alpizar-Jara | Universidade de Évora

### Comissão Organizadora

- Russell Alpizar-Jara (**Presidente**) | Universidade de Évora
- Dulce Gomes | Universidade de Évora
- Lígia Henriques-Rodrigues | Universidade de Évora
- Patrícia Filipe | ISCTE – Instituto Universitário de Lisboa

### Sessões Plenárias

- Anthony Davison | Escola Politécnica Federal de Lausanne
- António Pacheco | Instituto Superior Técnico, Universidade de Lisboa
- Genevera Allen | Universidade de Rice, Houston
- Maurizio Sanarico | Grupo SDG, Milão

Mais informações em <http://www.spe2021.uevora.pt>

FR

## • Sobre a nova página da SPE



A nova página da SPE merece “uma visita”, também pelas novas secções e propostas nela incluídas.

Para além dos já bem conhecidos módulos com Glossário, Publicações, Notícias...

Agora também pode contribuir para o progresso e prestígio da SPE de uma forma dinâmica e mais ativa...

Possui um novo contacto mail:

*spe-web@spestatistica.pt*



E ainda:

### Prémios & Desafios

A Sociedade Portuguesa de Estatística promove varios prémios como incentivo à atividade de estudo em Probabilidades e Estatística entre os jovens e estudantes.

Todos Anúncios de Prémios

Siga-nos na nossa rede social:



<https://twitter.com/spestatistica>

## Polígrafo Estatístico da SPE

A SPE vai começar a combater de forma mais ativa as afrontas diárias aos estatísticos através do *hashtag* do Twitter: **#PoligrafoEstatistico**.

A pandemia destaca o papel fundamental da Estatística na sociedade, mas é altura de recordar que a produção e comunicação de estatísticas envolve uma enorme responsabilidade e sentido ético.

Ações:

1. Subscriver e divulgar a conta *Twitter da SPE* (<https://twitter.com/home>) e o *hashtag* **#PoligrafoEstatistico** entre colegas, amigos e familiares por forma a dinamizarmos de forma massiva a nossa estratégia de comunicação.
2. Denunciar qualquer situação de aparente mau uso da(s) Estatística(s) usando o *hashtag* **#PoligrafoEstatistico**.

## • Prémio SPE 2020

O Prémio SPE, é promovido pela Sociedade Portuguesa de Estatística e pretende estimular a atividade de estudo e investigação científica em Probabilidades e Estatística entre os jovens.

Júri do Prémio SPE 2020:

- Prof. Giovanni Silva (Presidente), do IST / Universidade de Lisboa.
- Prof. Luís Machado, da Universidade do Minho.
- Prof. Pedro Oliveira, do ICBAS / Universidade do Porto.

O Prémio SPE 2020 foi atribuído a Ivo Miguel Sousa Ferreira, estudante de Doutoramento o Doutoramento em Estatística e Investigação Operacional na Faculdade de Ciências da Universidade de Lisboa, sob a supervisão das Professoras Cristina Rocha e Ana Maria Abreu.

No final desta edição do Boletim SPE damos o devido destaque desta notícia.

FR



## • Competição Internacional de Posters ISLP 2020-2021

Terminou a fase nacional da Competição Internacional de *Posters* dinamizada pelo ISLP (International Statistical Literacy Project).

Nesta competição, os estudantes das escolas e universidades de todo o mundo foram convidados a criar um *poster* que refletisse ou ilustrasse o uso de estatísticas na resolução de problemas em qualquer área do conhecimento.

Tendo em conta as edições anteriores e o contexto de pandemia em que esta decorreu, é de assinalar o grau de participação registado:

<b>Grupo</b>	<b>N.º de Trabalhos</b>	<b>N.º de Alunos</b>
3.º Ciclo	6	20
Ensino Secundário	28	124
Ensino Universitário	2	6
<b>TOTAL</b>	<b>36</b>	<b>150</b>

Com base nos temas “Ambiente”, “Biologia” e “Desenvolvimento sustentável”, os participantes apresentaram trabalhos bastante diversificados, dos quais o júri elegeu os seguintes:

Posição	Poster/Título	Estabelecimento Ensino	Professor	Alunos
<b>3.º Ciclo do Ensino Básico</b>				
1.º lugar	<a href="#">Impacto das atitudes diárias para a sustentabilidade do planeta</a>	Escola Básica e Secundária Artur Gonçalves	Alice Martins	Sara Raquel Silvério Manso Tomás Gabriel Garcia Diogo Miguel dos Santos Rodrigues
2.º lugar	<a href="#">Causas de morte em Portugal</a>	Escola Básica e Secundária Artur Gonçalves	Alice Martins	Margarida Freitas Mota Diego Mascarenhas N. Ferreira Joana de Oliveira Samagaio
3.º lugar	<a href="#">Precipitação em Portugal</a>	Escola Básica e Secundária Artur Gonçalves	Alice Martins	Maria Gameiro Gomes Rita Sequeira Rodrigues Rita dos Reis Nunes
<b>Ensino Secundário</b>				
1.º lugar	<a href="#">Relação entre o conhecimento sobre as alterações climáticas e a adoção de comportamentos para a sua mitigação</a>	Escola Secundária António Damásio	Hélder Manuel Martins	Manuel Correia João Silva Maria Alice Tiago Pereira Rafael Joaquim
2.º lugar	<a href="#">Impacto do consumo português nas populações de bacalhau da Noruega</a>	Escola Secundária António Damásio	Hélder Manuel Martins	Guilherme Abreu Miguel Colaço Alicia Cardoso Leonardo Rendeiro André Bartolomeu
3.º lugar	<a href="#">Sustentabilidade e preservação do ambiente</a>	Escola Secundária António Damásio	Hélder Manuel Martins	João Mendes Mafalda Vilela Sofia Guerreiro Joana Santa Ana Rita Fonseca
<b>Ensino Universitário</b>				
1.º lugar	<a href="#">Economia Circular</a>	Faculdade de Ciências e Tecnologia Universidade de Coimbra	Maria Emília de Mesquita Nogueira	Carolina Rodrigues Guímaro Matilde Neves Pessoa Vaz
2.º lugar	<a href="#">Exploring the Role of Climate Change Knowledge in Adapting to a More Sustainable Diet</a>	Universidade dos Açores	Maria Gabriela Meirelles	Ângela Ferreira Sara Braga Tatiana Cardoso Nuno Ouro

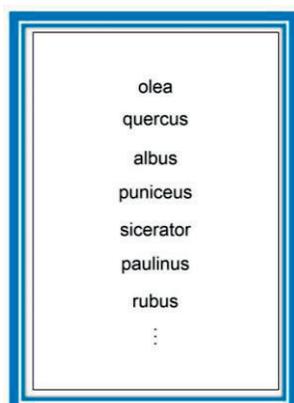
## Enigmística de mefqa

A I G A A

χ □

No Boletim SPE outono de 2020 (p. 7):

*"40 anos SPE: passado, presente e futuro"*



José Tiago de **Oliveira**  
Ivette de **Carvalho** Gomes  
João **Branco**  
Fernando **Rosado**,  
Carlo **Braumann**  
Carlos Daniel **Paulino**,  
Maria Eduarda **Silva**

# Episódios na História da Estatística

## Centenário de C. R. Rao – Sócio Honorário da SPE

Fernando Rosado, [fernando.rosado@ciencias.ulisboa.pt](mailto:fernando.rosado@ciencias.ulisboa.pt)

No passado dia 10 de setembro de 2020, Calyampudi Radhakrishna (C. R.) Rao completou 100 anos.

O Professor C. R. Rao é uma lenda viva, desde há longa data, decano da Estatística Mundial. O Professor C. R. Rao é um cientista galardoado pelas academias científicas de todo o mundo. O seu currículo único é o registo de uma vida de professor e investigador que, seguramente, a todos os níveis é um exemplo de enorme valor, um património, que pelos diferentes níveis da sociedade se reconhece como “um Memorial da Ciência Estatística”.



O Professor C. R. Rao, nascido na Índia, fez um percurso científico e académico que envolveu, por todo o mundo, as mais diversas Universidades e Associações Científicas.

Há dez anos, a Royal Statistical Society atribuiu-lhe a “Guy Medal in Gold”, “For his fundamental contributions to statistical theory and methodology, including unbiased estimation, variance reduction by sufficiency, efficiency of estimation, information geometry, as well as the application of matrix theory in linear statistical inference”. Esta Medalha de Ouro, é atribuída trienalmente pela Royal Statistical Society. A consulta da lista das dezenas de premiados desde 1892 é uma leitura histórica que inclui os nomes, as bases, os grandes dinamizadores, os faróis da Ciência Estatística Moderna. É um exercício que talvez seja interessante no sentido de avaliarmos a nossa relação com cada um desses pilares científicos; bem como medir o que cada um deles “significa para a ciência atual”. Para esse potencial consultor, especialmente o mais jovem, a lei do tempo, implacável, fará diluir e colocar na penumbra algum ou alguns dos nomes que para outro “anterior na linha do tempo”, seguramente eram identificados como um farol radiante e de primeira linha.

O “Fisher de hoje é o Fisher de há vinte ou trinta anos?” É como, assim à maneira de Camões: Quantos e quais da lista nomeada, são “aqueles que por obras valerosas se vão da lei da morte libertando”?

O Professor C. R. Rao foi Presidente de cinco Sociedades Científicas: Indian Econometric Society (1971-1976); International Biometric Society (1973-1975); Institute of Mathematical Statistics (1976-1977); International Statistical Institute (1977-1979) e o Forum for Interdisciplinary Mathematics (1982-1984).

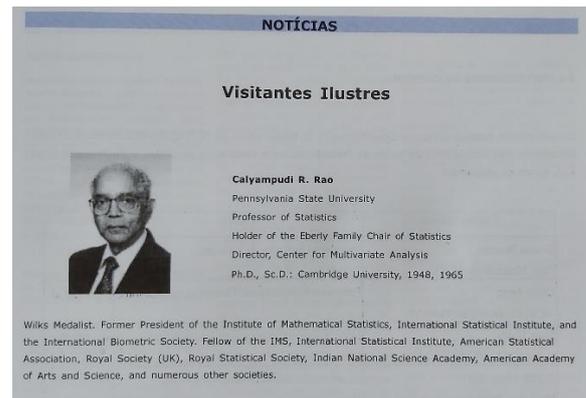
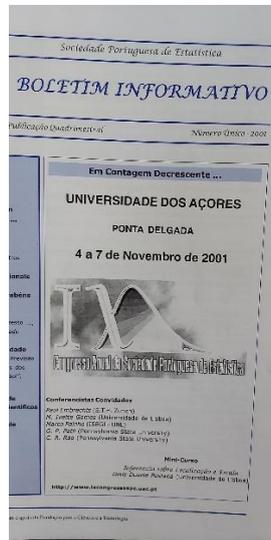
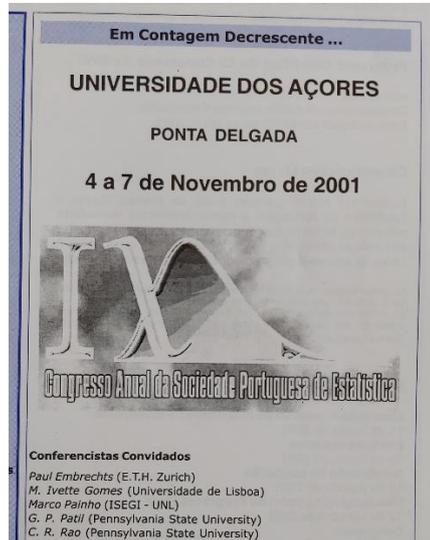
Em 1948, orientado por Ronald Fisher, o Professor C. R. Rao obteve o Doutoramento, na Universidade de Cambridge, com a tese: *Statistical Problems of Biological Classifications*.

Tão ilustre cientista num momento tão especial, tem recebido as mais diversas distinções e manifestações de apreço pela comunidade.

Em 2007, o Professor C. R. Rao recebeu o grau de Doutor Honoris Causa pela Universidade Nova de Lisboa.

O Professor C. R. Rao foi Orador Convidado no IX Congresso Anual da Sociedade Portuguesa de Estatística realizado, de 4 a 7 de novembro, nos Açores em 2001.

Passados vinte anos e com o devido enquadramento é interessante relatar este “episódio” na história da Estatística. Em primeiro lugar pela Honra de entre nós termos recebido tão ilustre convidado. Assim, foi relatado pelo Boletim Informativo da SPE:



Por todo o mundo e com as mais diversas sessões de homenagem e de prémios temos notícias que confirmam a importância deste homem, cientista e estatístico, que tive a oportunidade de, mais de perto, conhecer no referido Congresso da SPE.

Em momentos privados que, como anfitrião e Presidente da SPE, tive a feliz oportunidade de viver, permitiu-nos, apaixonadamente, partilhar vivências familiares. Foram momentos intensos que constroem excelentes memórias. Na dicotomia que, sempre, as nossas vidas contêm, entre “o eu público” e “o eu privado”, recordo a sua reserva, a sua modéstia e o recato que sempre mantinha no seu relacionamento com a sociedade e que nos foi transmitido pela sua esposa; ao mesmo tempo que tudo isto também significa, consolida e confirma a total “dedicação à sua causa primeira” – A Estatística.

Nessa época de terríveis momentos, vividos perante o “11 de setembro” e, obviamente, perante a relação direta com “os Açores e a base das Lajes”, muitos colegas, logo nos primeiros momentos, muito justamente, admitiam e aconselhavam uma alteração / adiamento do IX Congresso SPE. A decisão, aparentemente difícil tornou-se, na realidade muito simples. Um enorme apoio “de continuidade”, o primeiro e seguramente o mais forte, recebi do próprio Professor C. R. Rao que poucos dias depois do “11 de setembro” e por sua iniciativa, me transmitiu: “manterei a minha agenda em relação a Portugal”. A SPE foi honrada com a sua aceitação em ser sócio honorário; convite que lhe foi feito no Congresso dos Açores e confirmado pela sua eleição na Assembleia Geral da SPE em 2002. Assim, por ocasião do centenário do seu nascimento, a SPE de algum modo também partilha a festividade que todo o mundo celebra.

Ao lado, junto uma recordação pessoal que, revista tantos anos depois, com muita emoção partilho com os leitores do Boletim SPE primavera 2021; com os devidos respetos e homenagem por tão ilustre centenário.



## INE disponibiliza acesso a Microdados à Academia

Instituto Nacional de Estatística

A comunidade académica é reconhecida pelo INE como um importante utilizador de estatísticas oficiais. Esta apresenta necessidades específicas de informação relacionadas com o desenvolvimento de trabalhos de investigação, nomeadamente para a elaboração de artigos, projetos e dissertações.

Para responder a estas necessidades, o INE oferece, há já duas décadas, um serviço gratuito de acesso a microdados (dados individuais anonimizados). O acesso é concedido a investigadores credenciados, sob condições previamente definidas e contratualizadas através de um protocolo que o Instituto Nacional de Estatística estabeleceu com a Fundação para a Ciência e Tecnologia (FCT) e a Direção-Geral de Estatísticas da Educação e Ciência (DGEEC).

O Protocolo tem por objeto estabelecer as formas de cooperação entre as partes, nomeadamente no que respeita às condições de acesso, para fins científicos, a dados estatísticos individuais anonimizados, constantes de bases residentes no INE, com vista a implementar o disposto nos n.ºs 7 e 8 do artigo 6.º da Lei do Sistema Estatístico Nacional (Lei n.º 22/2008, de 13 de Maio).

### Processo

A lista de bases de microdados e os formulários de pedido de acreditação estão disponíveis no [Portal do INE](#) e no site da Direção-Geral das Estatísticas da Educação e Ciência ([DGEEC](#)). A documentação preenchida deve ser remetida à DGEEC, que avalia o projeto e a condição de investigador, concede a acreditação e remete o processo ao INE.

Após acreditação, o INE disponibiliza o acesso às bases de microdados de forma rápida e segura através de uma *Cloud* específica para investigadores. Trata-se da plataforma Cloud2FA.INE.PT, baseada num sistema de dupla autenticação de modo a garantir o acesso apenas ao investigador principal, com projeto acreditado pela DGEEC.



O acesso aos dados abrange os investigadores de universidades ou de outras instituições de ensino superior legalmente reconhecidas, bem como de organizações, instituições ou departamentos de investigação científica reconhecidos pela FCT e pela DGEEC.

### Bases de dados

A lista de base de microdados ao dispor dos investigadores acreditados é extensa, muito completa e variada nas temáticas que abrange, possibilitando trabalhos de investigação em diversas áreas,

nomeadamente: demografia, mercado de trabalho, empresas, saúde, cultura, agricultura, inovação e desenvolvimento.

O INE tem como objetivo aumentar o número e a abrangência das bases de microdados, permitindo o acesso a dados recolhidos e produzidos por si e pelas outras entidades do Sistema Estatístico Nacional, mas também a bases de microdados obtidas por via administrativa, no quadro da Administração Pública que, quando integrados entre si, permitem o estudo de importantes fenómenos da nossa sociedade, ampliando o conhecimento e desenvolvimento futuros.

Algumas destas Bases de microdados, designadamente as recebidas através do protocolo estabelecido pelo INE com a Autoridade Tributária, só estão disponíveis para acesso em ambiente de *Safe Center*. A sua utilização para acesso a microdados obedece a regras próprias e está sujeita a marcação prévia.

O INE dispõe de quatro *Safe Centers*, localizados no edifício sede, em Lisboa, nas Delegações Regionais de Coimbra e do Porto bem como um remote *safe center* na Direção Regional de Estatística da Madeira (projeto que concluiu a medida #33 – “INE Remote Safe Center” do Programa Simplex 2018).

Alguns indicadores sobre os microdados disponibilizados pelo INE para a Academia:

- Existem disponíveis 57 bases de microdados organizadas em 17 temas estatísticos;
- Foram já efetuados cerca de 900 pedidos de acesso;
- As bases foram já acedidas por 1900 investigadores.

A lista de bases de dados disponíveis, organizada por temas, encontra-se disponível no [Portal do INE](#).



## O Papel da Estatística no Combate à Pandemia COVID-19

João A. Branco, *jbranco1802@gmail.com*

*Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa*

### Anunciando a pandemia

Não há palavras que cheguem para descrever completamente a monumental agitação que uma minúscula partícula viral foi capaz de incutir na população humana do nosso planeta. Aparecido não se sabe porquê e também não se sabe bem onde, o vírus SARS-Cov-2, aproveitando a enorme capacidade de movimentação que as sociedades têm vindo a proporcionar aos potenciais viajantes nas últimas décadas, viajou vertiginosamente para todos os cantos do mundo tentando cumprir a sua inexorável missão de infectar todos os que conseguisse alcançar, provocando a doença Covid-19 que já levou à morte mais de dois milhões de pessoas no curto período de um ano.

As consequências não se fizeram esperar. Uma crise sanitária, económica e social de dimensões gigantescas instalou-se e os países depressa começaram a organizar-se reforçando os serviços de saúde, recorrendo a todos os meios disponíveis e também a novas capacidades, impondo aos seus habitantes novos comportamentos e novas maneiras de viver com vista a travar a propagação do vírus. Porém, logo se percebeu que a saída da crise e o eventual regresso à normalidade perdida só seriam possíveis com a disponibilização de vacinas que viessem proteger as pessoas da ameaçadora infecção. A comunidade científica respondeu com uma torrente de contribuições, incluindo a produção das almeçadas vacinas que já estão a ser administradas às populações.

O objectivo destas singelas notas é o de reflectir sobre a contribuição que a estatística pode dar no sentido de ajudar à compreensão do fenómeno da pandemia e concorrer para a sua erradicação e ao mesmo tempo observar o impacto que a Covid-19 tem sobre a estatística.

### Dados Covid-19 e a Estatística

Os dados relativos à pandemia Covid-19 constituem a matéria-prima que fornece informação útil que pode ajudar a compreender os vários aspectos deste fenómeno tão complexo. Por um lado é bom saber que o volume e fluxo dos dados é muito grande, e que estes estão disponíveis em repositórios nacionais e internacionais. Isto é, existe abundância de informação e esta está acessível para quem a queira usar nas suas investigações. Por outro lado não é bom saber que a qualidade desses dados nem sempre é boa. Este facto é frequentemente denunciado por vários daqueles que os pertentem analisar, não por falta de patriotismo, mas porque sabem que uma fraca fiabilidade dos dados pode conduzir a conclusões enganadoras com consequências imprevisíveis. As razões que justificam esta situação são muito compreensíveis e resultam da concorrência de diversos factores, entre os quais o grande número de intervenientes na sua recolha e organização e na urgência da sua divulgação. Os dados mais profusamente divulgados nos meios de comunicação social são principalmente os números respeitantes a infectados confirmados, óbitos, recuperados, activos, testes efectuados, internamentos em enfermaria e cuidados intensivos e vacinas administradas. O número de testes efectuados não tem sido devidamente valorizado e é fonte de divergência entre países (que teste? e, teste ou pessoa testada, o que conta?), mas o seu conhecimento, conjuntamente com o número de confirmados, é essencial para entender a real situação da pandemia e perceber como a doença se vai propagando. Não testar é como fechar uma janela para não ver uma parte da realidade. A análise estatística descritiva destes dados é extremamente valiosa

para a compreensão do desenrolar da pandemia e da sua essência. Os resultados destas análises são a componente forte das notícias diárias sobre a Covid-19.

Ao contrário do que acontece em muitas doenças em que o foco está em saber como tratar o indivíduo doente, no caso da Covid-19 e de outras doenças infecciosas, o interesse está centrado no que acontece em grupos da população. E quem fala de grupos, pensa em variabilidade, cujo estudo é do domínio da estatística. Este tipo de doenças levanta problemas que podem ser estudados recorrendo a métodos estatísticos tradicionais sendo que alguns desses problemas requerem métodos mais específicos. A causa da doença, com que frequência surge, como se transmite, se está associada a certos factores particulares e como se trata a doença são alguns desses problemas tipicamente estudados em epidemiologia, uma vasta área das ciências médicas que se ocupa dos fenómenos da doença e saúde das populações. Associadas a estas questões, típicas dos estudos observacionais, surgem os conceitos de prevalência, incidência, letalidade, mortalidade, riscos e factores de risco, testes de diagnóstico e outros que encontramos igualmente nos livros de estatística aplicada à medicina ou à biologia. Ao percorrer estes estudos e problemas é manifesta a convicção que a epidemiologia não dispensa a estatística, com a qual até partilha o modo indutivo de actuar. No que toca ao tratamento da doença, a pesquisa assenta em estudos de natureza experimental dominados pela realização de ensaios clínicos. A intervenção da estatística é imprescindível na condução de ensaios clínicos, sendo indispensável a presença de um estatístico em todas as fases do ensaio. Uma outra componente com forte tradição em epidemiologia são os modelos matemáticos usados para explicar o desenvolvimento da doença. Muitos destes modelos são de cariz determinístico mas alguns são de natureza estocástica. Sempre que haja lugar à estimação dos seus parâmetros voltamos ao domínio da estatística para obtermos estimativas e a respectiva precisão. Percebe-se, por tudo isto, que a estatística é uma arma indispensável para o combate à covid-19. Apresentam-se a seguir duas situações concretas que mostram que a utilização da estatística é uma tarefa delicada que é preciso executar com rigor e sempre com olho nos dados.

### **Um ensaio problemático**

Em 30 de Dezembro de 2020 a agência reguladora de medicamentos e produtos de saúde britânica (*MHRA*) aprovou o uso da vacina Oxford/AstraZeneca, produzida pela universidade de Oxford em parceria com o consórcio farmacêutico anglo-sueco AstraZeneca, para todo o Reino Unido. Numa altura em que a doença alastrava ferozmente produzindo os números de casos e de mortes mais elevados da Europa, a notícia da aprovação da vacina fez rejubilar o país, orgulhoso dos seus cientistas pela arma eficaz e segura de combate ao vírus produzida, e pelo natural despertar da esperança em salvar vidas e em retomar liberdades ameaçadas. Em 4 de Janeiro, cinco dias depois de aprovada, a vacina começou a ser inoculada na população em duas doses consecutivas com um intervalo entre tomas variando de 4 a 12 semanas. O monumental trabalho envolvido na execução deste projecto foi executado em menos de um ano, uma duração inacreditável, um tempo recorde, se nos lembrarmos de que no passado o tempo necessário para conseguir disponibilizar uma vacina ou um medicamento, geralmente muito longo, poderia chegar, em certos casos, a 10 ou mesmo até 15 anos. Os estritos padrões de segurança e eficácia da vacina foram avaliados com base em quatro ensaios clínicos de tipo aleatorizados e controlados envolvendo 23848 participantes com 18 ou mais anos de idade, oriundos de 3 países em 3 continentes (Reino Unido, Brasil e África do Sul) e provenientes de várias etnias. A complexidade do trabalho realizado bem como o rigor e cuidado com que foi executado pode ser apreciado no artigo disponibilizado em

<https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2932661-1> ,

onde se refere que o resultado final da investigação garante os níveis de eficácia e segurança exigidos pelas agências reguladoras, além de que a vacina apresenta vantagens relativas à conservação, transporte e distribuição quando comparada com outras vacinas criadas para combater o mesmo vírus. Contudo, apesar da qualidade dos resultados e da almejada conquista em termos de realização, muitos cientistas mostraram reservas quanto a vários aspectos do empreendimento. Uma análise desses aspectos foi publicada, como comentário, na revista *Lancet*, a mesma onde se encontra o artigo anterior, e pode ser vista em

<https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2932623-4>.

De todas as críticas apontadas destaca-se aqui aquela que, é aliás parte da razão destas notas, mereceu destaque e alerta para alguma preocupação, tanto da parte dos autores das duas referências mencionadas como da parte das agências reguladoras às quais a vacina foi submetida para avaliação. O que os autores confessam no seu texto é o seguinte:

*While the data presented here show that ChAdOx1 nCov-19 is efficacious against symptomatic disease, with most cases accruing in adults younger than 55 years of age so far, an important public health consideration is the morbidity and mortality of the disease in an older adult population and thus the potential efficacy in this age group. We have reported immunogenicity data showing similar immune responses following vaccination with two doses of ChAdOx1 nCov-19 in older adults, including those older than 70 years of age, when compared with those younger than 55 years. As older age groups were recruited later than younger age groups, there has been less time for cases to accrue and as a result, efficacy data in these cohorts are currently limited by the small number of cases, but additional data will be available in future analyses.*

Isto é, a representatividade do grupo de participantes com 55 anos de idade ou mais acabou por ficar enfraquecida pelo simples facto do recrutamento dos mais velhos se ter deixado para mais tarde em relação aos mais novos e por não se ter esperado pela acumulação dos resultados no grupo dos mais velhos, empobrecendo a informação esperada neste grupo. Simples facto digo eu, porque a reacção que este aparente detalhe causou foi deveras extraordinária. Com o argumento de que o número de participantes e número de casos no grupo etário dos mais idosos era reduzido para a avaliação da eficácia e segurança criou-se descrédito em relação à vacina e uma reacção negativa da parte da comunidade não se fez esperar. Enquanto países como o Reino Unido, Índia, México e Argentina decidiram não restringir a aplicação da vacina, em pelo menos 10 países europeus optou-se por não recomendar a vacina no grupo etário dos idosos. A decisão veio perturbar os planos de vacinação nacionais e teve certamente custos elevados, difíceis de avaliar. Esta inquietação inicial viria a ser aclarada posteriormente, com o acumular de novos dados para a faixa etária de interesse e depois da agência europeia do medicamento (*EMA*) e da organização mundial de saúde (*WHO*) terem recomendado o uso da vacina sem aquelas restrições. Espera-se que a submissão ao regulador americano venha a concretizar-se. O que se pretende realçar com esta narrativa é: i) destacar o significado e o valor do conceito de amostra num trabalho estatístico e as consequências que podem advir dos resultados de estudos e investigações se não respeitarmos escrupulosamente aquilo que a estatística nos exige, trabalhar com boas amostras, ii) sublinhar a atitude dos autores deste importante projecto pelo facto de, possivelmente motivados pela urgência em fornecer uma tábua de salvação à comunidade e eventualmente preocupados por terem efectuado um deficiente planeamento, não esconderam debilidades do seu trabalho. Na verdade o ensaio não provou que a vacina AstraZeneca não é segura ou não é eficaz, o que concluiu foi que num grupo etário não há informação suficiente para avaliar com rigor essas propriedades, iii) o grande envolvimento da sociedade em geral o que terá despoletado um interesse sério relativamente ao uso da estatística, pelo menos no presente contexto.

## **O gráfico do vigário**

Como é do conhecimento geral os resultados de uma análise estatística podem ser enganadores. Há muitas razões para que isso aconteça: dados deficientes, hipóteses impróprias, métodos inadequados, amostras não representativas e outras razões. Uma análise muito popular que pode conduzir a resultados enganadores é a simples interpretação da representação gráfica dos dados. A visualização da representação gráfica dos dados é um procedimento geralmente simples, rápido e muito poderoso no que respeita a descobrir relações, efectuar comparações e revelar aspectos subjacentes aos dados que a sua observação directa na fase bruta ou mesmo depois de organizados em tabelas raramente permite. Contudo estas aparentes vantagens podem ser ensombradas pelo facto da representação gráfica potenciar uma capacidade de produzir gráficos enganadores e ao mesmo tempo persuasivos e facilitadores de más decisões. Um dos gráficos mais populares que nos é apresentado nos primeiros contactos com a representação gráfica de dados é o gráfico de barras ou diagrama de barras usado para representar dados que de alguma forma se possam agrupar em classes. O gráfico de barras aparece todos os dias em grande abundância na imprensa escrita e nos outros meios de comunicação. A informação que transmite é facilmente captada e o facto de ser fácil alterar a sua estrutura (quando possível, escolhendo a posição

das barras, a sua ordem e cores, e modificando a própria escala) leva a que, por ignorância ou propositadamente, o gráfico de barras se transforme num gráfico transmitindo informação que não corresponde à informação contida nos dados.

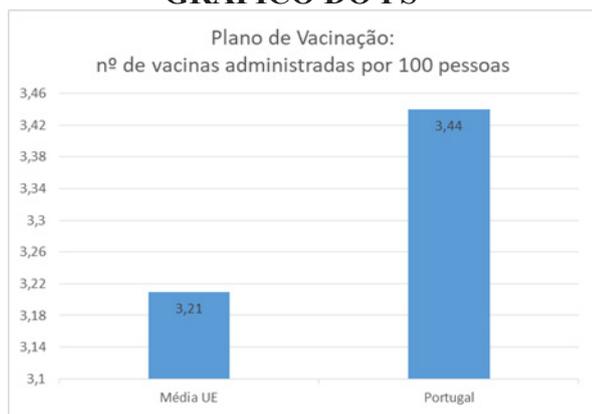
O diagrama de barras tem sido abundantemente usado para descrever o andamento da pandemia Covid-19 em vários países e regiões. Só que muitos dos diagramas apresentados são deliberadamente adulterados para que os leitores absorvam a ideia, diferente da realidade contida nos dados, que o autor convenientemente pretende transmitir. E a ideia errónea, voa imediatamente através das redes sociais e outros meios chegando ao mundo inteiro, engordando o volume diário das chamadas *fake news*. Os exemplos são imensos. Alguns, relativos a situações da Argentina, Rússia e Geórgia podem ser vistos em:

<https://towardsdatascience.com/stopping-covid-19-with-misleading-graphs-6812a61a57c9>

Para terminar este comentário apresenta-se um diagrama de barras relativo à candente situação da vacinação em Portugal, divulgado nas redes sociais e que deu origem ao artigo “*As vacinas em gráficos do PS: uma crónica de propaganda visual*” de Manuel Carvalho no jornal Público (2021/02/05), disponível no seguinte endereço electrónico

<https://www.publico.pt/2021/02/05/sociedade/noticia/vacinas-graficos-ps-cronica-propaganda-visual-1949517>.

### GRÁFICO DO PS



### GRÁFICO CORRECTO



Parece óbvio que o gráfico da esquerda (gráfico do PS) foi manipulado de modo a fazer passar a ideia que os autores pretendem fazer passar. A verdade é que a eliminação da origem das ordenadas e a escolha de uma escala conveniente fez com que Portugal apareça como vacinando 3 vezes mais do que a média europeia, quando de facto a verdadeira diferença entre o que Portugal vacina e a média europeia é cerca de 7%. Os autores do "gráfico do PS" caíram no logro de o divulgar nas redes sociais pois logo a seguir tiveram que o retirar em virtude da chuva de críticas e denúncias que surgiram uma vez descoberta a batota. As redes sociais não têm só a face negra, há também a face iluminada que funcionou para desmascarar o truque e difundir o acontecimento. O episódio também revela que a literacia estatística pode ser rara mas existe e o caldinho das redes sociais pode até ser um bom tónico para o seu crescimento. Muito grave e preocupante é também saber que o software Excel, um dos mais utilizados mundialmente, responde, automaticamente, precisamente com o mesmo erro, isto é, com o mesmo gráfico. Já o diagrama de barras que se apresenta do lado direito (gráfico correcto), livre de manipulações, mostra de forma atractiva a informação que os números realmente contêm.

Esta história, recorrente, faz recordar outra história semelhante intitulada “Ilusão de Óptica” publicada no jornal Expresso (nº1547, 2002/06/21) em que o jornalista critica a ousadia de se publicar um gráfico de barras para divulgar a vitória numa sondagem em que Jamila Madeira, muito conhecida pela sua actividade na área política, saía vencedora em relação ao seu oponente Filipe Costa na disputa preparatória para a eleição do secretário geral da Juventude Socialista. A barra de Jamila Madeira, que arrecadou mais 7 votos (2%) do que Filipe Costa, aparecia no gráfico com se tivesse obtido o dobro dos votos deste último.

Os casos que conduzem a resultados enganadores obtidos com base no mau uso deliberado dos métodos da estatística lembram histórias de contos do vigário. Todos ouviram falar dos truques maliciosos do vigário para ludibriar incautos, mas o vigário sabe, e é aí que reside a sua força e a sua esperança, que há sempre os que não resistem à tentação e ao encanto da proposta e acabam por acreditar na ilusão que lhes é oferecida. O gráfico do vigário é, nos dias de hoje, bem mais eficaz e perigoso, não porque seja mais ardiloso, mas porque num só ataque pode atingir simultaneamente múltiplas potenciais vítimas. A análise de um gráfico não dispensa a atenção que é imprescindível dar aos dados que são afinal a parte onde se encontra a verdade. Fazer estatística é desvendar essa verdade e para isso o estatístico trabalha pacientemente, qual mineiro procurando o precioso mineral. Em todos os casos a verdade é essencial.

### **Comentário Final**

Durante um ano verificou-se que a estatística foi abundantemente invocada e utilizada como arma de combate à actual crise pandémica. Desta exposição tão intensa resultou que a estatística revelou-se mais indispensável e o seu valor tornou-se mais conhecido e mais apreciado pela sociedade em geral. A Covid-19 veio afinal promover a estatística, uma das armas usada para a combater, elevando a credibilidade e respeito não só da estatística como também da ciência no seu todo. Mas veio também pedir mais trabalho: produção de dados, mais eficaz, e de dados de melhor qualidade. Nunca o mundo precisou tanto de dados credíveis. Os institutos nacionais de estatística, temporariamente impedidos, pela crise, do seu funcionamento normal, têm vindo a ser pressionados para acelerar a produção de dados não só relativos à pandemia como aos outros sectores que estão a sofrer as consequências da perturbação global a que têm estado sujeitos. A covid-19 veio ainda realçar a necessidade de uma maior literacia estatística e a incentivar a luta contra o uso indevido da estatística. Esta é uma das preocupações constantes de muitas instituições, em particular do Instituto Internacional de Estatística (*ISI*) e da sua Associação Internacional para as Estatísticas Oficiais (*IAOS*). O impacto da Covid-19 sobre a acção da estatística e a contribuição decisiva da estatística para a compreensão da Covid-19 acontecem porque os problemas da Covid-19 são de natureza estatística e o seu estudo tem de passar inevitavelmente pelo uso dos métodos da estatística.



## De passagem pela mortalidade...

Clara Cordeiro, [ccordei@ualg.pt](mailto:ccordei@ualg.pt)

*Faculdade de Ciências e Tecnologia da Universidade do Algarve  
CEAUL, Faculdade de Ciências, Universidade de Lisboa*

Os estudos de mortalidade têm sido utilizados em ciências atuariais, nas apólices de seguros de vida e nos planos de reforma, em emergências humanitárias, quando ocorre uma guerra, e em situações de saúde pública, quando se enfrenta uma doença desconhecida ou uma nova pandemia. Uma fonte valiosa de informação sobre mortalidade é a *Human Mortality Database*<sup>1</sup>(HMD), uma base de dados internacional que contém dados sobre mortalidade e população, para mais de 40 países.

Numa situação pandémica, tal como a COVID-19, a vigilância da mortalidade é essencial para monitorizar o impacto global da gravidade da doença na saúde pública, em termos de gravidade e mortalidade da doença (Setel et al., 2020). De acordo com os indicadores de mortalidade, esta pandemia tem ceifado inúmeras vidas por todo o mundo e Portugal não foi exceção (Fig.1).

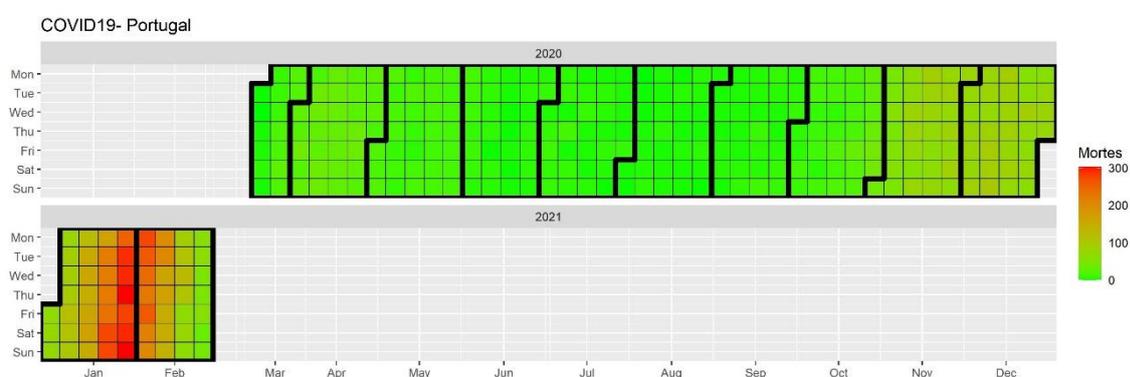


Figura 1: Número de óbitos diários de 17/03/2020 a 28/02/2021<sup>2</sup>.

Outro indicador útil é o excesso de mortalidade e consiste na diferença entre o número de mortes observado e o número esperado de mortes em condições "normais" (Checchi e Roberts, 2008). A *World Health Organization* define

*“Excess mortality, excess mortality rate as Mortality above what would be expected based on the non-crisis mortality rate in the population of interest. Excess mortality is thus mortality that is attributable to the crisis conditions. It can be expressed as a rate (the difference between observed and non-crisis mortality rates), or as a total number of excess deaths”* (Checchi e Roberts, 2008).

<sup>1</sup> <https://www.mortality.org/>

<sup>2</sup> Fonte: [https://raw.githubusercontent.com/CEAUL/Dados\\_COVID-19\\_PT/master/data/covid19pt\\_DSSG\\_Long.csv](https://raw.githubusercontent.com/CEAUL/Dados_COVID-19_PT/master/data/covid19pt_DSSG_Long.csv)

O número esperado de mortes é calculado sobre um determinado período, sendo o período de 5 anos o mais utilizado. A Fig.2 apresenta o excesso de mortalidade para o período de 17/03/2020 a 28/02/2021, onde o número esperado de mortes foi calculado com base nos 5 anos anteriores.

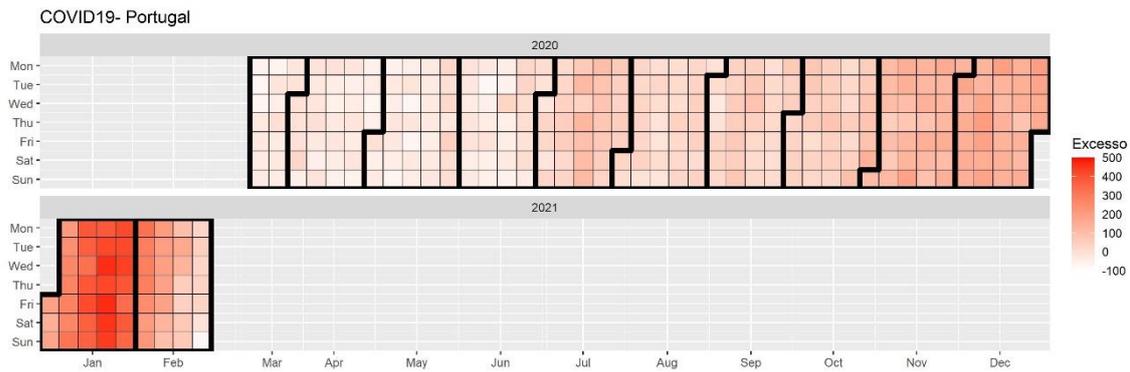


Figura 2: Excesso de mortalidade de 17/03/2020 a 28/02/2021<sup>3</sup>.

O excesso de mortalidade tem sido utilizado em estudos epidemiológicos e de saúde pública, como um indicador útil do impacto de ondas de calor (Matte et al., 2016), gripe (Nunes et al., 2011) e, atualmente, COVID-19 (Nogueira et al., 2020), entre outros. Os dados sobre o excesso de mortalidade têm sido disponibilizados pelas publicações *The Economist*<sup>4</sup>, *The New York Times*<sup>5</sup> e *The Financial Times*<sup>6</sup>. Na Europa, a atividade de mortalidade é registada pelo projeto EuroMOMO<sup>7</sup>, um sistema de monitorização fácil de usar, que mede o excesso de mortalidade semanal.

Em algumas situações, a falta de dados fiáveis pode ser um problema, especialmente nos países em desenvolvimento, devido a atrasos no registo ou contagem incorreta de mortes (Checchi e Roberts, 2008). ANACONDA<sup>8</sup> (*Analysis of National Causes of Death for Action*) é uma ferramenta que efetua um diagnóstico completo da qualidade dos dados relativos à mortalidade, verificando a existência de potenciais erros e inconsistências (Mikkelsen et al., 2020).

Ao longo dos anos, várias metodologias foram desenvolvidas para a modelação e previsão da mortalidade (Booth e Tickle, 2008). A metodologia da mortalidade Lee-Carter (Lee e Carter, 1992), e as suas variantes, têm sido um dos métodos mais utilizados. A metodologia de Lee-Carter tem várias vantagens: inclui um modelo estocástico simples, com apenas um parâmetro variável no tempo, tem um desempenho relativamente bom, quando as tendências passadas são lineares, e pode prever um padrão de mortalidade em função da idade (Booth e Tickle, 2008). Os modelos de séries temporais são, também, comumente utilizados, tais como a modelação ARIMA, univariada e multivariada. A introdução de modelos de séries temporais tem a vantagem de obter uma distribuição de probabilidade para a previsão e, também permite, a determinação de intervalos de previsão (Booth e Tickle, 2008). O artigo de Janssen et al. (2018) apresenta uma análise dos avanços efetuados na previsão da mortalidade e de possíveis desafios para a investigação futura. No entanto, a utilização de outros modelos de séries temporais, podem também juntar-se aos desafios.

A análise dos dados de mortalidade é fundamental para as autoridades de saúde pública, de cada país, e para os decisores políticos tomarem decisões, ou avaliarem a eficácia das estratégias de prevenção e

<sup>3</sup> Fonte: [https://raw.githubusercontent.com/CEAUL/Dados\\_COVID-19\\_PT/master/data/covid19pt\\_DSSG\\_Long.csv](https://raw.githubusercontent.com/CEAUL/Dados_COVID-19_PT/master/data/covid19pt_DSSG_Long.csv)

<sup>4</sup> <https://github.com/TheEconomist/covid-19-excess-deaths-tracker>

<sup>5</sup> <https://github.com/nytimes/covid-19-data/tree/master/excess-deaths>

<sup>6</sup> <https://github.com/Financial-Times/coronavirus-excess-mortality-data>

<sup>7</sup> <https://www.euromomo.eu/>

<sup>8</sup> <https://crvsgateway.info/anaconda>

resposta adotadas, tais como o confinamento. Os seus resultados têm implicações de comunicação rápida para os decisores políticos, clínicos e a população.

## Referências

- Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of actuarial science*, 3(1-2):3-43.
- Checchi, F. and Roberts, L. (2005). Interpreting and using mortality data in humanitarian emergencies. *Humanitarian Practice Network*, 52.
- Janssen, F. (2018). Advances in mortality forecasting: introduction. *Genus*, 74(21). doi.org/10.1186/s41118-018-0045-7.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659-671.
- Matte, T. D., Lane, K., and Ito, K. (2016). Excess mortality attributable to extreme heat in new york city, 1997-2013. *Health security*, 14(2):64-70.
- Mikkelsen, L., Moesgaard, K., Hegnauer, M., and Lopez, A. D. (2020). ANACONDA: a new tool to improve mortality and cause of death data. *BMC medicine*, 18(1):1-13.
- Nogueira, P. J., de Araújo Nobre, M., Nicola, P. J., Furtado, C., and Carneiro, A. V. (2020). Excess mortality estimation during the covid-19 pandemic: preliminary data from portugal. *Acta Médica Portuguesa*, 33(13).
- Nunes, B., Viboud, C., Machado, A., Ringholz, C., Rebelo-de Andrade, H., Nogueira, P., and Miller, M. (2011). Excess mortality associated with influenza epidemics in portugal, 1980 to 2004. *PloS one*, 6(6):e20661.
- Setel, P., AbouZahr, C., Atuheire, E. B., Bratschi, M., Cercone, E., Chinganya, O., Clapham, B., Clark, S. J., Congdon, C., de Savigny, D., et al. (2020). Mortality surveillance during the covid-19 pandemic. *Bulletin of the World Health Organization*, 98(6):374.



# Algumas considerações sobre a COVID-19

Isabel Natário, [icn@fct.unl.pt](mailto:icn@fct.unl.pt)

*Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa; CMA*

## Introdução

Em Dezembro de 2019 surgiu em Wuhan, República da China, um surto de um síndrome respiratório agudo severo provocado por um novo coronavírus, o SARS-CoV-2, resultando na doença COVID-19. A disseminação deste novo agente patogénico foi muito rápida e, mesmo isolando a cidade onde surgiu e outras vizinhas, tal não o impediu de alastrar a todos os continentes logo durante o 1º trimestre do ano de 2020 (<https://www.worldometers.info/coronavirus/worldwide-graphs/#countries-cases>).

No início de Abril já se contabilizavam mais de 1,3 milhões de casos em todo mundo e mais de 70000 mortes, a sua maioria entre a população envelhecida da Europa, onde o excesso de mortalidade por todas as causas subia acima do esperado num grande número de países [1]. Sentiam-se enormes pressões sobre os sistemas de saúde e de apoio social e sobre os seus trabalhadores. Era reportada a falta de capacidade de testagem, a falta de equipamento de proteção com consequente aumento de infeções entre o pessoal da saúde e cuidador, falta de equipamento de tratamento como ventiladores, multiplicavam-se os surtos em lares e casas de repouso. Era imperativa uma resposta rápida. As diretivas eram testagem e vigilância, com o seguimento dos contatos, medidas comunitárias, como o distanciamento social, reforço dos sistemas de saúde e informação às populações. Em muitos países estas foram reforçadas com medidas mais drásticas de confinamentos quase totais das populações.

Sendo um agente patogénico novo, era essencialmente desconhecido. No entanto, as suas consequências eram tão devastadoras que uma resposta pronta era imperiosa, tendo essa pressa causado a exposição pública de várias contradições sobre medidas e procedimentos a adotar. A comunidade científica viu-se na contingência de contribuir para essa falta de informação, tão necessária à tomada de decisões, e fê-lo de forma exemplar: mais de 110000 artigos usando o termo “COVID-19” estão indexados no motor de busca da PubMed (<https://pubmed.ncbi.nlm.nih.gov/?term=covid+19>) à data da escrita deste texto. A comunidade editorial também, mantendo uma política de publicação aberta sobre todos os textos de investigação referentes à COVID-19 e correspondentes dados de suporte. Tal contribuiu para a transparência e partilha do conhecimento, permitindo evitar informação contraditória, granjeando e reforçando a confiança nos achados científicos [2].

Neste trabalho a matemática e estatística têm uma função muito importante. Ao nível individual, a estabelecer fatores de risco e de transmissão para esta doença de etiologia desconhecida. Ao nível global a caracterizar a dinâmica dos surtos e a disseminação da doença, investigar a efetividade de medidas de contenção da transmissão da doença, desde a higiene respiratória e distância social, a vacinação, o fecho de escolas e instalações desportivas, do comércio e restauração, o trabalho a partir de casa até ao confinamento forçado às habitações [3], e também avaliar os efeitos da pandemia ao nível de saúde pública, da sociedade e da economia, com via à sua gestão [4].

## Transmissibilidade e fatores de risco da COVID-19

Esta doença é provocada por um agente patogénico muito eficiente e silencioso, com um grande potencial de dispersão, associado possivelmente a uma grande transmissão assintomática ou pré-sintomática [5]. O contágio parece ser feito essencialmente pelas gotículas de respiração entre pessoas que contactam proximamente [6]. Assim, devem-se adotar medidas de prevenção, como lavar

corretamente as mãos, etiqueta respiratória, desinfecção de superfícies, isolamento e distância social (mais de 2 m), não havendo ainda tratamento conhecido para esta doença.

Em Portugal foi conduzido um estudo de literacia dos Portugueses em saúde, no contexto pandémico da COVID-19, através de um inquérito a doentes de um Centro de Saúde, escolhidos de forma semi-aleatória entre indivíduos com consulta marcada no centro de saúde num dado período de tempo, tendo concluído que somos uma população relativamente bem informada sobre que medidas de prevenção que se devem tomar [7].

Verifica-se que à maioria dos casos sintomáticos documentados correspondem apenas sintomas ligeiros. Contudo uma parte importante desenvolve pneumonia, alguns com falha respiratória aguda e uma pequena porção tem um desfecho fatal. A severidade dos sintomas verificou agravar-se com a idade, presença de comorbilidades como diabetes, condições cardíacas e cancro, género, etnicidade, obesidade [4, 5, 7]. Advoga-se que esta pandemia é na verdade uma sindemia, em que a doença COVID-19 interage com outras doenças endémicas, com elas influenciando os eventos de doença correspondentes, e que deve ser tratada como tal em termos das políticas de saúde pública associadas [8].

## **Modelação epidemiológica**

Num cenário de epidemia, interessa medir os eventos de doença coletivamente em relação a uma população em risco, numa abordagem epidemiológica, de forma a determinar as suas causas e compreender a sua dinâmica face a intervenções.

A Organização Mundial de Saúde, o European Center for Disease Prevention and Control e os organismos nacionais que com ele se articulam, o seu congénere americano Centers for Disease Prevention and Control e muitos outros organismos têm grupos de trabalho que exploram as características de cada surto que vai surgindo, presentemente o de COVID-19. O uso de modelos epidemiológicos matemáticos e estatísticos, permite a caracterização da dinâmica dos surtos e a investigação da eficiência de diferentes estratégias de ação [5].

Central a esta modelação estão os modelos epidémicos determinísticos de disseminação da doença, primeiro propostos por Kermack e McKendrick (1927). Estes modelos dividem a população em grupos (compartimentos) representando estados discretos da doença (Suscetíveis, Infetados e Recuperados - SIR - por exemplo) e ainda em características demográficas, biológicas e comportamentais. Resolvem equações diferenciais que representam um sistema analítico da epidemia a tempo contínuo. Dentro de cada compartimento os seus elementos são considerados homogéneos e considerados apenas de forma agregada. Dependem de parâmetros chave como sendo a taxa (ou força) de infeção e a taxa de recuperação, que devem ser estimados para cada epidemia.

Com base nestes modelos determinísticos Reed e Frost e Greenwood (1928, 1931) desenvolveram modelos estocásticos, permitindo a inclusão de incerteza na transmissão da doença. Os valores esperados das suas soluções estocásticas correspondem às soluções dos modelos determinísticos correspondentes e permitem fazer previsões em epidemias.

Estes modelos permitem então conhecer o estado da epidemia em cada instante. Sendo simplificações da realidade, diferem naturalmente do real estado da epidemia, mas ainda assim permitem obter um conhecimento aproximado do que acontecerá quando se tomam certas medidas de contenção e controlo. Podem ser complexificados de forma a levar em consideração heterogeneidade de contactos humanos, a sua distribuição geográfica, etc..

Um conceito bastante importante associado às epidemias, é o chamado número básico de reprodução,  $R_0$ , que representa o número de infeções diretas que ocorrem como resultado da introdução de um único infetado na população dos suscetíveis. Ou seja, é o número de casos que se espera que ocorram em média numa população homogénea como resultado da infeção de um único infetado, quando a população é suscetível no início da epidemia, quando há imunidade nula na população.

Se  $R_0$  for maior que 1 a infeção espalha-se de forma exponencial; Se for menor que 1 espalha-se devagar e eventualmente extingue-se. Quanto maior  $R_0$  mais rápida é a progressão da epidemia.

Este número é estimado dos dados recolhidos (para um novo vírus, quanto mais cedo houver dados disponíveis o mais próximo é provável de estar o verdadeiro valor de  $R_0$ , assumindo dados de boa qualidade). Depende do modelo usado e dos dados.  $R_0$  é proporcional à transmissibilidade da doença, taxa de contato de um suscetível com um infetado e inversamente proporcional à taxa de recuperação

destes modelos. Assim, é maior para maiores populações com mais suscetíveis; é maior quanto maior for a infetividade do vírus; é menor quanto maior for a taxa de remoção dos infetados [9].

A COVID-19 é provocada por um agente patogénico muito eficiente e silencioso, com um grande potencial de dispersão, à qual corresponde um tempo de duplicação de casos grande (estimado em cerca de 3 dias), um número médio de infeções provocadas por um infetado quando introduzido numa população de suscetíveis  $R_0$  grande e possivelmente uma transmissão assintomática ou pré-sintomática grande [5].

É apontado por vários estudos que as estimativas do  $R_0$  para SARS-CoV-2 variam muito. Em Janeiro de 2020, a World Health Organization avançou com uma estimativa preliminar do  $R_0$  de 1.4 - 2.5 [10] e corrigiu depois para 2 - 4 [11]. Um outro estudo, baseado numa meta-análise de 29 estudos apenas da República da China, resultou nas estimativas de 3.32 IC(2.81-3.82) [12]. Baseado no surto que aconteceu no barco de cruzeiro Diamond Princess, que aí ficou confinado, estimou-se uma mediana para o  $R_0$  de 2.28 IC(2.06-2.52) [13]. Outro trabalho, analisou 12 estudos que estimavam o  $R_0$  entre 1.4 - 6.5, com média de 3.28 e mediana de 2.79 [14]. Baseado ainda num outro conjunto de 9 estudos da República da China e da Coreia do Sul e também num outro conjunto um pouco mais alargado de 16 estudos, o  $R_0$  estimou-se como variando entre 0.4 - 4.6, sendo a média respetivamente de 2.63 IC(1.85-3.41) e 2.65 IC(1.97-3.09) [5].

Um outro conceito muito importante para as epidemias é a chamada imunidade de grupo, que é adquirida quando uma grande percentagem da população ganha imunidade. Esta será adquirida preferencialmente através de vacinação, já que se for conseguida pela infeção natural de grande parte dos suscetíveis tal poderá estar associado a um aumento indesejável da taxa de morte. Além disso, no caso da COVID-19, a imunidade parece não ficar garantida pela contração da doença [15].

O número  $R_0$  pode prever a proporção de pessoas que devem ser vacinadas para se adquirir imunidade de grupo. Esta proporção deve ser maior que  $1-1/R_0$  [16]. Tomando o intervalo de valores eventualmente mais conservador apontado pela World Health Organization, que genericamente parece englobar as outras estimativas apresentadas, a proporção a vacinar deverá situar-se entre 50% e 75% da população. Intimamente relacionado com o  $R_0$  está o chamado número de reprodução efetivo,  $R_t$ , representando o número de pessoas na população que podem ser infetadas por um indivíduo, num qualquer instante de tempo específico  $t$ . Representa como a transmissibilidade do vírus varia ao longo do desenvolvimento da epidemia. Varia com o número de infetados, número de suscetíveis e política de contactos. Em qualquer instante  $t$ ,  $R_t=R_0(1-P_t)$ ,  $P_t$  é a proporção de pessoas imunes nesse instante [16].

## **Estratégia para lidar com a pandemia**

Tipicamente a estratégia de combate tem sido a de implementar medidas de contenção que controlem o número médio de casos gerados por um infectado típico no instante  $t$ , o número de reprodução  $R_t$  [3, 9]. As características deste vírus que o tornam muito eficaz a disseminar-se, com grande  $R_0$ , com eventuais contágios pré- e assintomáticos e relativamente longo período de contágio, justificam que se tenham vindo a considerar intervenções sociais e economicamente muito perturbadoras associadas a confinamentos globais demorados [5].

## **Dados**

Os dados epidemiológicos, a sua qualidade e sua interpretação expedita são cruciais para a definição de políticas de gestão, que dependem do conhecimento atual da evolução da pandemia. O  $R_0$  e o  $R_t$  são estimados a partir de dados recolhidos sobre os números de casos.

Mais do que nunca, esta pandemia aguçou a necessidade do uso e tratamento dos dados quase em tempo real, uma dificuldade acrescida já que estes dados estão sujeitos a várias condicionantes, tornando este trabalho muito difícil e ingrato [5]. A utilização deste tipo de dados é complexa e sujeita a várias fontes de enviesamento, desde a necessidade de considerar distribuições do tempo entre dois acontecimentos, que pode ter muita incerteza associada devido a padrões de reporte variáveis, censura, inconsistências na definição de caso que pode variar à medida que se conhece melhor a doença, impactos de eventuais intervenções, alterações na dinâmica de transmissão, taxa de reporte de casos variável, condicionada até pelo excesso de trabalho em situações de aumento abrupto de casos, etc.. Acresce que os dados de vigilância em contexto de pandemia acabam muitas vezes por ser recolhidos de forma apressada e pouco

cuidada, já que a prioridade dos profissionais de saúde vai para os próprios doentes, resultando em dados incompletos, atrasados e por vezes de pouca qualidade [3].

Os dados em “tempo real” encontram barreiras nas infraestruturas de cada país para registrar, centralizar e disponibilizar os dados de vigilância epidemiológica, não sendo homogêneo entre países. O atraso da notificação dos casos depende naturalmente de como está implementado o sistema de reporte. Em Portugal o sistema é gerido pelo SINAVE Sistema Nacional de Vigilância Epidemiológica, que engloba duas plataformas informáticas associadas aos médicos e aos laboratórios, o SINAVE-MED e o SINAVE-LAB, que recolhem, respetivamente, as notificações clínicas e inquéritos epidemiológicos e as notificações laboratoriais. Esta base de dados, que iremos espreitar de seguida, é disponibilizada aos investigadores sob pedido.

### **Base de dados COVID-19 do SINAVE**

Apresenta-se de seguida uma breve análise da base de dados disponibilizada pelo SINAVE, incluindo os casos do início de Março de 2020 ao fim de Dezembro de 2020. Esta base de dados agrega todos os casos notificados, pela data de notificação, tanto incluídos pelos laboratórios como pelos clínicos, sendo incorporada pelo SINAVE. Inclui 419910 casos incidentes, pessoas infetadas que foram confirmadas como tal através de um teste laboratorial. Contém informação individual em 50 variáveis, que se podem dividir em 3 grupos:

1. O identificador de pessoa, que é único. Cada pessoa nesta base de dados só consta 1 única vez, de acordo com informações prestadas pelo SINAVE;
2. Datas: data de notificação (data em que o sistema tem conhecimento do caso); data de colheita da amostra e data de início dos sintomas;
3. Dados demográficos e geográficos;
4. Existência de sintomas e sua apresentação;
5. Existência de comorbilidades e sua apresentação.

Esta base de dados não tem informação sobre o desfecho do caso – cura, hospitalização simples, hospitalização em cuidados intensivos ou morte. Apesar de essa informação existir e o seu acesso poder ser feito através dos Serviços Partilhados do Ministério da Saúde, não é possível articulá-la com esta base de dados por questões administrativas.

Esta base apresenta incompletudes resultantes, por exemplo, da notificação não ter sido feita pelo médico, mas sim pelo laboratório que não tem acesso a informações como comorbilidades, por exemplo. Entrando posteriormente para a notificação clínica, a informação vai sendo completada, mas não para todos os casos. Esta informação, em alturas de maior pressão, tende a vir cada vez menos detalhada. Muita gente insere dados neste sistema, sendo por isso esta uma base de dados com grande fluxo, com alguns erros, mesmo sendo sujeita a monitorização de controle de qualidade (SINAVE).

Na Tabela 1 apresenta-se a descrição das 3 variáveis relativas a datas na base de dados, incluindo informação sobre informação faltante. Verifica-se que o último trimestre do ano apresenta um aumento muito significativo de casos.

Na Tabela 2 apresenta-se a descrição das variáveis demográficas.

Verifica-se que faltam muitos dados relativamente à variável data de início de sintomas. Tal pode-se dever às pessoas não terem tido sintomas ou a esta data não ter sido registada.

Assim, restringiu-se a base de dados apenas aos casos em que a data de sintomas existia. Compara-se na Figura 1 o número de casos por data de notificação total (incluindo os casos que não têm data de início de sintomas associado) com o número de casos por data de notificação restringido apenas aos que têm data de início de sintomas associado. Verifica-se que sempre que há uma tendência de aumento do número de casos notificados, a variável data de início de sintomas surge mais incompleta. O cenário que se observa entre os meses de maio e setembro seria o expectável, já que haveria uma franja de casos notificados sem data de início de sintomas associada, simplesmente por essas pessoas não terem sintomas.

Variável	Histogramas das datas por mês de notificação	% de NAs
Data de notificação		0
Data de início dos sintomas		59.5
Data de colheita da amostra		10.3

Tabela 1: Descrição das variáveis datas de notificação, início de sintomas e colheita da amostra.



Figura 1: Número diário de casos por data de notificação total (mais claro) e restringido àqueles casos que têm associada uma data de início de sintomas (mais escuro).

Variável	Gráficos das datas por mês de notificação	% de NAs
Idade		10
Sexo		0.04
Concelho de morada		2.8

Tabela 2: Descrição das variáveis idade, sexo e número de casos por concelho de morada.

No primeiro painel da Figura 2 apresenta-se o número diário de casos, restringidos aos que têm data de início de sintomas associada, por data de notificação (mais escuro) ou por data de início de sintomas (mais claro). Apesar de o desfaseamento de datas ser pequeno quando há menos casos, verifica-se que a curva do número de casos por data de notificação tende a estar um pouco adiantada em relação à curva do número de casos por data de início de sintomas. No segundo painel da Figura 2 apresenta-se o valor da diferença na data de notificação, em dias, entre a data de notificação e a data de início de sintomas, assim como uma linha indicadora de tendência, com um comportamento essencialmente constante. Neste gráfico é muito notório que a diferença tende a ser crescente no decorrer da pandemia, passando progressivamente de uma situação inicial em que a notificação acontecia antes do início dos sintomas (possivelmente associado à pesquisa de contatos) para outra em que a data de início de sintomas é anterior à data de notificação. Esta última situação prende-se com os sintomas de doença surgirem antes da notificação, quando são os sintomas que levam à realização de um teste, mas também porque com o grande aumento de casos verificado o processo de notificações sofreu certamente maiores atrasos. Nota-se também que há muitas diferenças que parecem excessivamente grandes. No entanto verifica-se que a mediana da diferença é de 4 dias, a média de 8.72 dias, o 1º quartil é 2 dias e o 3º quartil 7 dias. Sendo esta diferença bastante variável (desvio padrão de 25.9 dias) e assimétrica, não será tão bem representada pela média, mais sensível a valores extremos, mas melhor pelos quartis.



Figura 2: Painel Superior: número diário de casos restringidos aos que têm data de início de sintomas associada, por data de notificação (mais escuro) ou por data de início de sintomas (mais claro). Painel Inferior: diferença entre as datas de notificação e de início de sintomas, em dias, e linha indicadora de tendência.

Apresenta-se na Figura 3 os diagramas em caixa de bigodes das diferenças, em dias, entre a data de notificação e a data de início de sintomas, por mês de notificação, permitindo que os bigodes avancem apenas até aos quantis 10% e 90% de todas as diferenças, para que os *outliers* fora deste faixa de valores não monopolizem o gráfico. Apresentam-se ainda, em linha tracejada, a mediana (4 dias) e os 1º e 3º quartis (2 e 7 dias, respetivamente) de todas as diferenças. Verifica-se que essencialmente esta diferença é positiva, indicando que a data de início de sintomas é anterior à data de notificação do sistema. Verifica-se ainda uma ligeira tendência decrescente desta diferença ao longo do tempo, com a maioria dos meses tendo uma mediana de diferenças coincidente com a mediana global.

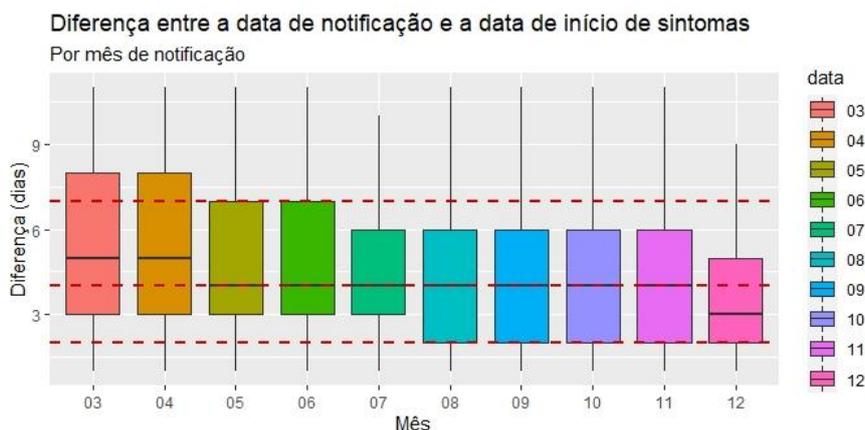


Figura 3: Tempo (em dias) entre a data de notificação e a da início de sintomas, por mês de notificação.

Na Figura 4 podem-se ver os diagramas em caixa de bigodes das diferenças, em dias, entre a data de notificação e a data de início de sintomas, por quinzena de notificação, permitindo que os bigodes avancem apenas até aos quantis 10% e 90% de todas as diferenças, para que os outliers fora deste faixa de valores não monopolizem o gráfico. Apresenta-se uma linha de tendência, confirmando-se a tendência decrescente, e verificando a existência de uma certa sazonalidade.

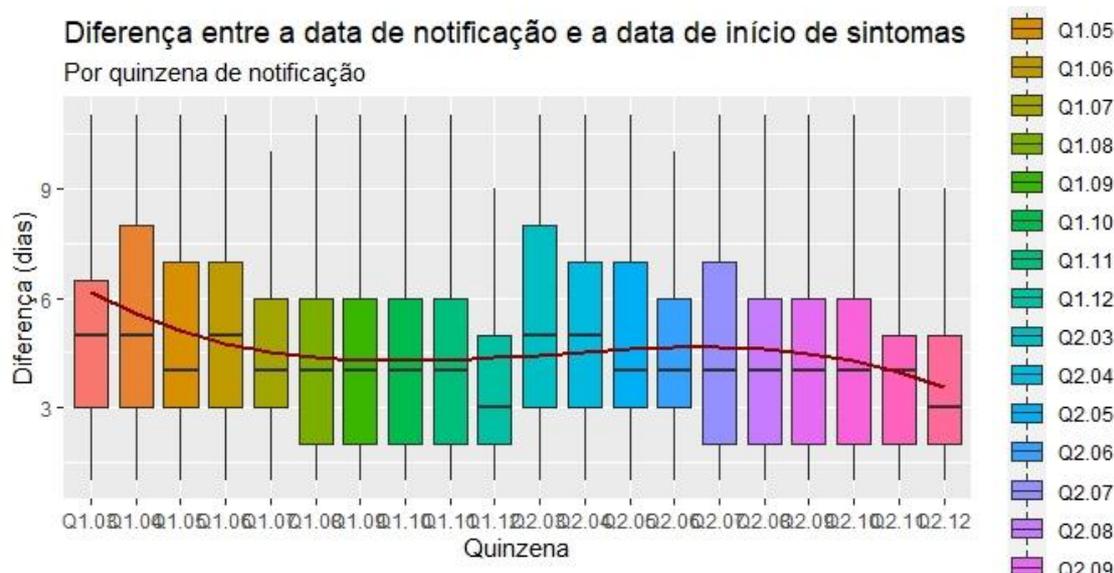


Figura 4: Tempo (em dias) entre a data de notificação e a do início de sintomas, por quinzena de notificação.

Para a base de dados restringida aos casos com data de sintomas, apresenta-se na Tabela 3 a descrição dos dados relativos a alguns dos sintomas mais comuns. Registam-se muitos dados faltantes, mas estão de acordo com o que se sabe sobre a doença.

Variável	Sim %	Não %	% de NAs
Tosse seca	49.7	21.3	28.9
Febre	36.6	31	32.3
Cefaleia	34.4	26.4	39.2
Fraqueza geral	19.3	30.8	49.8
Diarreia	9.0	91.0	0.0
Náuseas-vômitos	6.7	93.3	0.0
Pneumonia	2.5	26	71.3
Auscultação pulmonar anómala	2.3	97.7	0.0
Coma	0.1	99.9	0.0

Tabela 3: Descrição das variáveis relacionadas com os sintomas.

## Notas Finais

A pandemia de COVID-19 veio relembrar a fragilidade do ser humano e colocá-lo à prova a vários níveis. A sociedade viu-se forçada a adotar novos comportamentos e a prescindir de direitos básicos de liberdade para conseguir mitigar a transmissão da doença.

A comunidade científica foi chamada a dar respostas concretas e urgentes para a resolução da pandemia, desde o descortinar da etiologia desconhecida de um novo agente patogénico e da regulação da sua transmissão, à gestão dos efeitos da pandemia a nível de saúde pública, da economia e da sociedade, passando pela vacinação. E a comunidade científica correspondeu.

No entanto, o carácter de urgência da resposta, fez também progredir a ciência e o processo científico. A ânsia de publicação de resultados precipitou a validação de alguns trabalhos que, sem qualidade, foram

publicados. Tal faz repensar o método de revisão científica, havendo propostas de melhoria como, por exemplo, a proposta de ciência aberta [2]. Resultados contraditórios minam a confiança na ciência, mas a partilha de dados e resultados em pré-publicações, para que se possam reproduzir os resultados descritos, ajuda todos os investigadores em particular e a ciência em geral.

Sem dúvida que se salientam no combate à pandemia a matemática e a estatística. Nunca tantos números, gráficos e modelos foram apresentados e discutidos nos media! Talvez seja a primeira vez que o grande público de apercebe da sua utilidade. Em particular, destacam-se os modelos epidémicos, determinísticos e estocásticos e toda a sua teoria já antes existente, e o *software* que os acompanha livremente no R Epidemics Consortium (<https://www.repidemicsconsortium.org/>).

Para esta pandemia há ainda questões a responder. Qual é a efetiva duração da imunidade de quem ficou doente? E da imunidade após a vacina? Qual a efectividade da vacina para diferentes estirpes da doença? Quais os efeitos clínicos pós-infecção, quais os efeitos das medidas de contenção na saúde mental a longo prazo [4]? Quais as consequências para a saúde pública de as pessoas se terem afastado dos serviços de saúde por outras doenças não COVID-19 [17]? E há outras questões indiretas, muito relevantes, que também se levantam. Quais são as consequências sociais, económicas e ambientais desta pandemia? Já se percebeu que acentuou as desigualdades sociais e o direito à saúde dentro de cada país e entre países, que fez aumentar muito o desemprego e que possivelmente o mundo se prepara para viver uns tempos em recessão. Mas o pior de tudo é a incerteza. Mas nisso os estatísticos são peritos!

## Agradecimentos

A autora agradece à Divisão de Epidemiologia e Estatística da Direção de Serviços de Informação e Análise da Direção Geral da Saúde, os dados de saúde pública do Sistema Nacional de Vigilância Epidemiológica (SINAVE) relativos a doentes infetados pelo novo coronavírus SARS-CoV-2. Este trabalho foi parcialmente suportado por fundos nacionais através da Fundação para a Ciência e Tecnologia, Portugal, projeto UID/MAT/00297/2019.

## Referências Bibliográficas

- [1] European Centre for Disease Prevention and Control (ECDC). *Rapid Risk Assessment: Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – eighth update*. Abril 2020.
- [2] Schwab S., Held L. *Science after Covid-19 Faster, better, stronger?* Significance. Royal Statistical Society. Agosto 2020.
- [3] Nunes B., Caetano C., Antunes L., Dias C.. *Statistics in Times of Pandemics: the Role of Statistical and Epidemiological Methods During the COVID-19 Emergnecy (Invited Paper with Discussion)*. REVSTAT – Statistical Journal, 18: 553–564. Outubro 2020.
- [4] Lutchmansingh D.D. et al. *A Clinic Blueprint for Post-Coronavirus Disease 2019 RECOVERY - Learning From the Past, Looking to the Future*. Chest, 29: 949-958. Março 2021.
- [5] Overton C.E. *Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example*. Infectious Disease Modelling 5: 409-441. Julho 2020.
- [6] World Health Organization. COVID-19 questions and answers site. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted>. Accessed on March 2021.
- [7] Gomes da Silva J., Silva C.S., Alexandre B., Morgado P.. *Health Literacy of the Inland Population in Mitigation Phase 3.2 of the COVID-19 Pandemic in Portugal: A Descriptive Cross-Sectional Study*. Portuguese Journal of Public Health, 38: 51–61. Dezembro 2020.
- [8] Fronteira I. et al.. *The SARS-CoV-2 pandemic: A syndemic perspective*. One Health, 12: 100228. Fevereiro 2021
- [9] Aronson J.K., Brassey J., Mahtani .KR. “When will it be over?”: *An introduction to viral reproduction numbers, R0 and Re*. University of Oxford’s COVID-19 Evidence Service Team. Abril 2020. <https://www.cebm.net/covid-19/when-will-it-be-over-an-introduction-to-viral-reproduction-numbers-r0-and-re>
- [10] Viceconte G., Petrosillo N.. *COVID-19 R0: Magic number or conundrum?* Infectious Disease Reports, 12; 8216. Fevereiro 2020

- [11] World Health Organization. Update-28. COVID-19 – a global pandemic. What do we know about SARS-CoV-2 and COVID-19? Junho 2020.  
[https://www.who.int/docs/default-source/coronaviruse/risk-comms-updates/update-28-covid-19-what-we-know-may-2020.pdf?sfvrsn=ed6e286c\\_2](https://www.who.int/docs/default-source/coronaviruse/risk-comms-updates/update-28-covid-19-what-we-know-may-2020.pdf?sfvrsn=ed6e286c_2)
- [12] Alimohamadi Y., Taghdir M., Sepandi, M.. *Estimate of the Basic Reproduction Number for COVID-19: A Systematic Review and Meta-analysis*. Journal of Preventive Medicine & Public Health, 53: 151-157. Março 2020.
- [13] Zhanga S., Diaob M.Y., Yuc W., Peic L., Lind Z., Chena D.. *Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis*. International Journal of Infectious Diseases, 93: 201-204. Fevereiro 2020.
- [14] Liu Y., Gayle A.A., Wilder-Smith A., Rocklov J.. *The reproductive number of COVID-19 is higher compared to SARS coronavirus*. Journal of Travel Medicine, 27(2):taaa021. Fevereiro de 2020.
- [15] Bartsch Y.C. et al. *Discrete SARS-CoV-2 antibody titers track with functional humoral stability*. Nature Communications, 12: 1018. Fevereiro de 2021.
- [16] Becker N. G., Hall, R.. *Immunization Levels for Preventing Epidemics in a Community of Households Made up of Individuals of Various Types*. Mathematical Biosciences, 132: 205-216. Março 1995.
- [17] Santana R., Sousa J.S., Soares P., Lopes S., Boto P., Rocha J.V.. *The Demand for Hospital Emergency Services: Trends during the First Month of COVID-19 Response*. Portuguese Journal of Public Health, 38: 30-36. Abril 2020.



# CoronaSurveys na estimação de casos ativos de Covid-19

Raquel Menezes<sup>a</sup>, [rmenezes@math.uminho.pt](mailto:rmenezes@math.uminho.pt)

Carlos Baquero<sup>b</sup>, [cbm@di.uminho.pt](mailto:cbm@di.uminho.pt)

CoronaSurveys Team, <http://coronasurveys.org/team/>

<sup>a</sup> Centro de Matemática da Universidade do Minho

<sup>b</sup> Universidade do Minho & INEST TEC

## 1. Introdução

A epidemia de SARS-CoV-2, o vírus responsável pela doença Covid-19, causou um profundo impacto na sociedade face à rápida disseminação global do vírus em início de 2020. Na primeira vaga de propagação do vírus, a ainda incipiente capacidade de testagem aliada à forte taxa de crescimento do número de casos tornava muito difícil a contabilização do número total de casos de Covid-19. Adicionalmente, no primeiro semestre de 2020 havia ainda pouca informação sobre a duração da fase de contágio da doença. Conjugados estes fatores, tornava-se difícil estimar o *número de casos ativos em fase de contágio*, parâmetro importante para aferir o risco de contágio num dado local e momento. Neste artigo, procuramos mostrar como foi feita a estimação deste indicador, que se encontra disponível desde Julho de 2020, em conjunto com outros indicadores, na página do projeto *CoronaSurveys* (<https://coronasurveys.org>), onde é facultada informação sobre a evolução da pandemia nos diversos países do mundo.

## 2. Cumulativo de casos

Uma vez conhecida a sequência genética do vírus, no início de 2020, passou a ser possível a sua deteção por meio de RT-PCR<sup>1</sup>, a partir de amostras biológicas. As autoridades de saúde de vários países passaram a reportar os números de casos detetados diariamente, bem como a mortalidade atribuída a Covid-19. Uns dos primeiros agregadores de dados a nível mundial foram a Universidade de Johns Hopkins (<https://coronavirus.jhu.edu>) e o ECDC (*European Centre for Disease Prevention and Control*, <https://www.ecdc.europa.eu/en>). Estas entidades disponibilizaram também ficheiros em formato “.csv” ou “.xls” com atualizações diárias, tendo mais recentemente o ECDC passado a atualizações semanais. O número acumulado de casos é particularmente relevante para aferir a proporção da população que já teve contacto com a doença e poderá assim ser menos suscetível à mesma. A comparação do cumulativo de casos com o cumulativo de fatalidades permite também, como veremos de seguida, aferir o nível de deteção de casos em cada país. Em 13 Março de 2020, surge um artigo na secção de notícias na *Nature* (<https://www.nature.com/articles/d41586-020-00760-8>), onde se chama a atenção para o elevado nível de sub-deteção nos casos identificados por RT-PCR. Nesse mesmo dia, é feita em Espanha uma “sondagem” via Twitter que aponta para um número acumulado de casos de cerca de 30 vezes o número até então conhecido. Mais tarde esta técnica de amostragem, onde se pergunta a cada inquirido a proporção de casos que conhece no seu círculo social, veio a ser identificada como uma instância do método *network-scale-up*, Bernard HR. (1991), desenvolvido nos anos 90 para a contagem indireta de sub-populações difíceis de inquirir diretamente, Bernard HR. (2010). Foi esta a génese do projeto *CoronaSurveys*, que passou a disponibilizar em permanência inquéritos deste tipo em dezenas de países e línguas.

---

<sup>1</sup> Acrónimo de *reverse transcription-polymerase chain reaction*.

A necessidade de validar as estimativas obtidas por *network-scale-up* levou a equipa a considerar um outro estudo de Março de 2020, Russell TW. (2020), mais tarde publicado na BMC Medicine em Outubro de 2020, Russell TW. (2020b). A ideia é que, mesmo quando não podemos observar diretamente 100% dos casos, podemos observar seu efeito na taxa de mortalidade, *case fatality rate* (CFR), sendo que taxas de mortalidade superiores ao esperado são em circunstâncias normais indicador de sub-deteção de casos. O modelo toma como referência a taxa de mortalidade esperada de Covid-19 e que resulta da divisão do número cumulativo de mortes pelo número cumulativo de casos. Em Abril de 2020, foi publicado no *The Lancet* um artigo, Verity R. (2020), que estabelece o CFR como sendo de 1.38%, sendo este valor já conhecido em Março via *pre-print*. Observando a proporção deste CFR de referência com o valor do CFR observado em cada momento, torna-se possível, em cada região e momento, estimar um multiplicador que permite inferir quantos casos terão aproximadamente ocorrido até uma dada data, Baquero C. (2020) e Russell TW. (2020).

No caso de Espanha, ambos os estimadores do número acumulado de casos, com base em *survey network-scale-up* e com base no multiplicador CFR, vieram a concordar com os resultados de inquéritos serológicos, España G. (2020) (prevalência de IgG), como se pode ver na Figura 1, retirada do artigo, Baquero C. (2020).

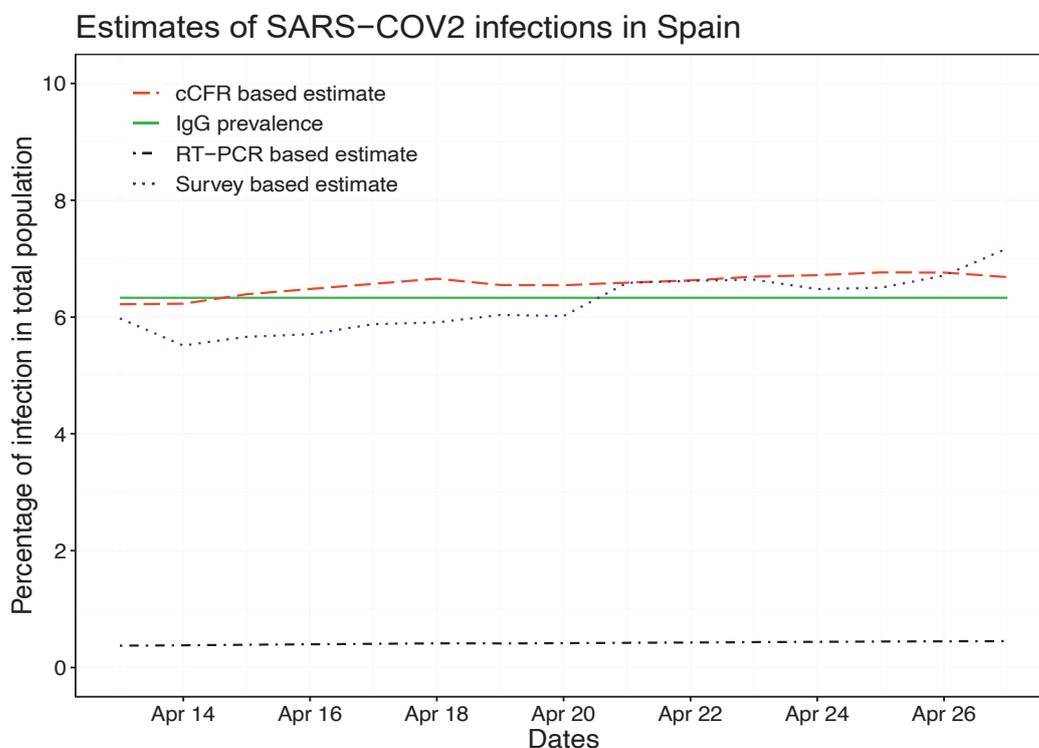


Figure 1. Estimativas de infeções SARS-CoV-2 em Espanha (retirado de Baquero C. (2020))

### 3. Casos diários - curva epidémica

O próximo passo no sentido da obtenção do número de casos ativos em fase de contágio passa pela aferição do número de casos identificados em cada dia, por outras palavras a obtenção da curva epidémica. Este passo é trivial quando aplicado aos dados oficiais de totais de casos confirmados por RT-PCR e reportados pelas autoridades de saúde. Sendo  $D = (d_1, \dots, d_n)$  a série de novos casos e  $C = (c_1, \dots, c_n)$  a respetiva cumulativa do total de casos, temos  $c_t = \sum_{i=1}^t d_i$ . Do mesmo modo, e no sentido inverso, temos  $d_i = c_i - c_{i-1}$  com  $c_0 = 0$ . Se a série  $D$  tiver sempre valores não negativos a correspondente série  $C$  nunca decresce. (É raro, mas por vezes há correções em alguns países onde se recorre a imputações negativas nos casos e na mortalidade.)

Como vimos na secção anterior, os números oficiais de casos podem sub-estimar largamente a real prevalência, tal como aferida em campanhas de serologia onde se mede a prevalência de anticorpos SARS-CoV-2 na população. Restam os dois estimadores desenvolvidos acima: *network-scale-up* e via

CFR. O primeiro, sendo baseado em inquéritos de resposta livre e com uma taxa de resposta diária muito variável, não se revela muito adequado à desagregação da soma e inferência da série D de casos diários. Resta-nos trabalhar com o estimador baseado no CFR e que podemos designar por  $\hat{C}$ .

O estimador  $\hat{C}$  resulta de uma função de transformação que recebe como parâmetro duas séries: a série de casos acumulados (dados oficiais confirmados por RT-PCR), acima designado por C e a série de óbitos cumulativos atribuídos à doença, a designar por  $C_o$ . Temos então que  $\hat{C} = f(C, C_o)$  em que a função  $f(\cdot)$ , para além de considerar os óbitos e os casos, entra também em linha de conta com a distribuição de tempo entre doença e eventual óbito, tal como definido em Russell TW. (2020). Para a nossa análise presente o fator importante é que sendo  $\hat{C}$  um estimador de um valor cumulativo, esta série pode, contudo, exibir decréscimos pontuais apesar de tendencialmente ser crescente.

Coloca-se agora o objetivo de derivar uma série de estimadores de casos diários  $\hat{D}$ , pelo que se considera  $\tilde{C} = (\tilde{c}_1, \dots, \tilde{c}_n)$  onde  $\tilde{c}_i = \text{Max}\{\hat{c}_1, \dots, \hat{c}_i\}$ , garantindo que a série cumulativa é sempre não-decrescente. Torna-se então possível definir  $\hat{D}$  tal que  $\hat{d}_i = \tilde{c}_i - \tilde{c}_{i-1}$  com  $\tilde{c}_0 = 0$ .

#### 4. Casos ativos em fase de contágio

Até fins de Maio de 2020, os casos que tinham testado positivos só eram considerados isentos de risco de contágio após dois testes RT-PCR negativos separados de 24h. Em 27 de Maio, é feita nova recomendação em *WHO Clinical management of COVID-19*<sup>2</sup> onde para casos ligeiros de considera uma janela de risco de contágio até dez dias depois da ocorrência de sintomas. Já era sabido que “*viral RNA detection by PCR does not equate to infectiousness or viable virus*”, esta citação surge num estudo publicado em 23 de Maio de 2020<sup>3</sup>, onde se considera a ausência de vírus viável e replicável a partir de 11 dias. Sendo que, nas 48h anteriores ao aparecimento de sintomas, pode já haver transmissão óptimos por considerar uma janela de 12 dias em que há risco de contágio.

A sequência de “casos ativos em fase de contágio” é assim obtida ativando cada novo caso por um período de 12 dias e classificando os casos como expirados findo esse período, deixando de contribuir para o somatório diário de casos ativos. Não se fez um ajuste fino da data de arranque dos casos, sendo assumida alguma incerteza de posicionamento temporal pois os casos não são usualmente reportados na data de primeira ocorrência de sintomas.

#### 5. Dados para Portugal

A Figura 2. ilustra a situação de casos ativos em fase de contágio em Portugal até 20 de Fevereiro de 2021. A informação atualizada está disponível *online* em <https://coronasurveys.org/data/>, opção “*Coronasurveys Global Estimates*”, selecionando Portugal. A curva a verde mostra o resultado de considerar ativo durante 12 dias cada caso oficial RT-PCR, algo que como foi referido tende a subestimar o número total de casos. Este é, no entanto, um indicador importante do número de casos que estão em isolamento e que em geral não contribuem para a transmissão. A curva a amarelo é a que resulta do processo descrito neste artigo, sendo a curva a roxo a que resulta de dividir a mortalidade diária por 0.0138 (CFR base de 1.38%) e desfazar o valor 13 dias para o passado (separação média entre sintomas e morte<sup>4</sup>) para obter um outro estimador de casos.

<sup>2</sup> <https://www.who.int/publications/i/item/clinical-management-of-covid-19>

<sup>3</sup> [https://www.ams.edu.sg/view-pdf.aspx?file=media%5C5556\\_fi\\_331.pdf](https://www.ams.edu.sg/view-pdf.aspx?file=media%5C5556_fi_331.pdf)

<sup>4</sup> <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#table-2>

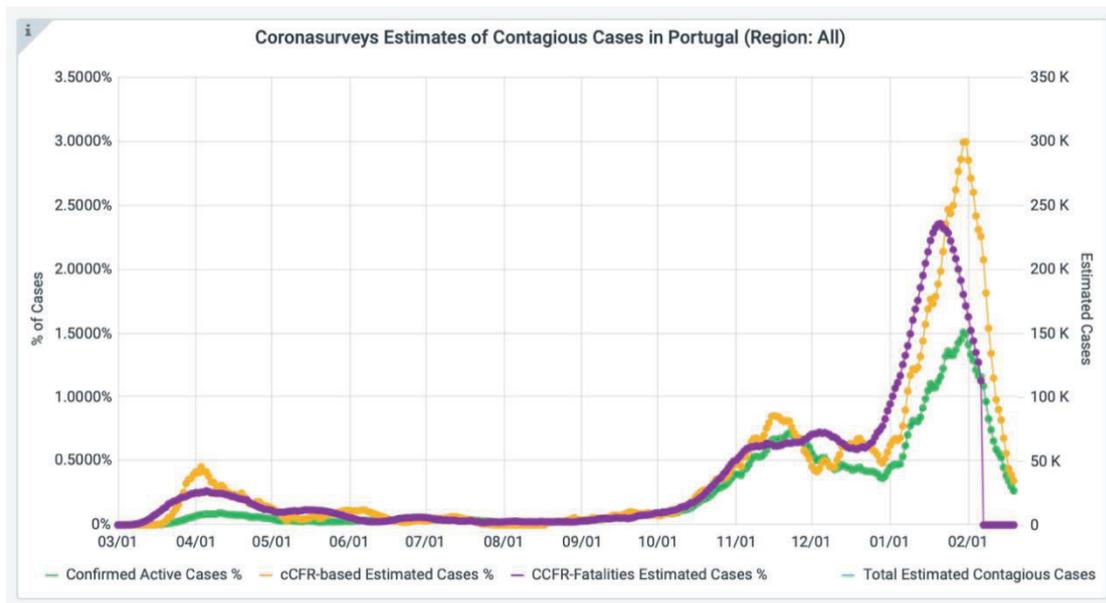


Figure 2. Estimativas de casos ativos em fase de contágio em Portugal até 20 de Fevereiro de 2021 (retirado de <https://coronasurveys.org>).

Como se pode observar no pico da primeira vaga, Abril 2020, houve um expressiva proporção de casos não detetados, evidenciada pela separação entre as linhas amarela e verde. Estimamos que, nessa altura, 0.5% da população sintomática esteve contagiosa (à qual acresce a população assintomática, possivelmente 50% mais). Já no pico de Janeiro de 2021, estimamos 3% da população em fase de contágio, sendo que apenas 1.5% foi detetada nos dados oficiais. Outro aspeto interessante é a possível ocorrência de um número significativo de casos não detetados no período de Natal, possivelmente devido a uma diminuição da testagem. Por fim, observamos que a 20 de Fevereiro de 2021 já quase não há separação entre o estimador e os números oficiais, possível indicador de uma situação bem mais controlada em termos de deteção de casos.

## 6. Trabalho futuro

Este estudo foi feito com base na estimativa do CFR, fundamentada pelos dados associados ao estudo de Wuhan (Verity R., 2020). É sabido, no entanto, que este valor varia de acordo com as regiões e o momento temporal. Fatores importantes serão a distribuição etária em cada país, e a proporção de gamas etárias infetadas em cada momento. Uma evolução possível passa pela utilização de um CFR ajustado em função da distribuição etária num dado país, e das proporções relativas de infeção detetada em cada grupo etário num dado momento do tempo.

O projeto *CoronaSurveys*, que conta com uma vasta equipa de internacional de investigadores, para além de estimativas globais de números totais de casos e de casos ativos, tem desenvolvido inquéritos regionais no sentido de melhor caracterizar comunidades específicas, nomeadamente na região de Madrid. No contexto de uma colaboração com a Facebook, University of Maryland e Carnegie Mellon University, que facultou o acesso a outra fonte de dados de inquéritos em grande escala e de abrangência global, está em curso o desenvolvimento de outros indicadores, como seja a taxa de contatos com não coabitantes, medida relevante para aferir o efeito de confinamentos, e o desenvolvimento de projeções de número de casos e mortalidade a uma e duas semanas.

## Referências

Baquero C. (2020). Carlos Baquero, Paolo Casari, Antonio Fernández Anta, Davide Frey, Augusto Garcia-Agundez, Chryssis Georgiou, Raquel Menezes, Nicolas Nicolaou, Oluwasegun Ojo, Paul Patras. Measuring icebergs: Using different methods to estimate the number of covid-19 cases in portugal and spain. *medRxiv*, 2020.

- Bernard HR. (2010). H Russell Bernard, Tim Hallett, Alexandrina Iovita, Eugene C Johnsen, Rob Lyerla, Christopher McCarty, Mary Mahy, Matthew J Salganik, Tetiana Saliuk, Otilia Scutelniciuc, et al. Counting hard-to-count populations: the network scale-up method for public health. *Sexually transmitted infections*, 86(Suppl 2):ii11–ii15, 2010.
- Bernard HR. (1991). H Russell Bernard, Eugene C Johnsen, Peter D Killworth, and Scott Robinson. Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Social science research*, 20(2):109–121, 1991.
- España G. (2020). Consumo y Bienestar Social Gobierno de España, Ministerio de Sanidad. Estudio nacional de sero-epidemiología de la infección por sars-cov-2 en españa, informe preliminar tras la primera ronda. Technical report, Gobierno de España, Ministerio de Sanidad, Consumo y Bienestar Social, 5 2020.
- Russell TW. (2020). Timothy W Russell, Joel Hellewell, Sam Abbott, Nick Golding, Hamish Gibbs, Christopher I Jarvis, Kevin van Zandvoort, Stefan Flasche, Rosalind M Eggo, W John Edmunds, Adam J Kucharski. Using a delay-adjusted case fatality ratio to estimate under-reporting. Available at the Centre for Mathematical Modelling of Infectious Diseases Repository, 2020. [https://cmmid.github.io/topics/covid19/global\\_cfr\\_estimates.html](https://cmmid.github.io/topics/covid19/global_cfr_estimates.html)
- Russell TW. (2020b). Russell, T.W., Golding, N., Hellewell, J. *et al.* Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med* **18**, 332 (2020). <https://doi.org/10.1186/s12916-020-01790-9>
- Verity R. (2020). Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Win skill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G T Walker, Han Fu, Amy Dighe, Jamie T Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 2020/04/11.



## **Estatística no CEAUL em tempos de COVID**

Tiago A. Marques, *tiago.marques@st-andrews.ac.uk*

*Centre for Research into Ecological and Environmental Modelling, University of St Andrews,  
Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa,  
Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa*

e

Lisete Sousa, *lmsousa@fc.ul.pt*

*DEIO, Faculdade de Ciências, Universidade de Lisboa,  
Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa*

Desde o início da Pandemia que no CEAUL – Centro de Estatística e Aplicações da Universidade de Lisboa, percebemos que seria importante tentar intervir no contexto da COVID19. Tal como em muitos outros centros de investigação um pouco por todo o mundo assistimos a um movimento conjunto, uma tentativa de contribuir para um “esforço de guerra” comum que nos pareceu fundamental abraçar. Afinal, era óbvio que se tratava de uma guerra global contra um inimigo, também ele global. Assim sendo, foi com naturalidade que criámos uma equipa de trabalho interna que foi reunindo informalmente, analisando as notícias, pensando nos números que eram reportados, e tão ou mais importantes, nos números que não eram reportados, tentando ajudar a perceber o que os números poderiam, ou não, querer dizer. Muitos de nós fizeram as suas próprias projeções pessoais, mais ou menos sofisticadas, mas, curiosamente, não nos atrevemos a torná-las públicas. Porquê? Porque não nos parecia útil. Não sabíamos o suficiente sobre o problema. Não bastava ajustar uma curva exponencial aos dados para saber o que aí vinha. Esse tanto nós sabíamos, mas ficámos a ver as previsões de muitos outros que pelos vistos não sabiam. Neste contexto, acabámos por nos envolver numa série de projetos relacionados com a COVID19. Desse trabalho, provavelmente, a conclusão mais óbvia é que numa epidemia a noção de variabilidade tem de ser considerada e tentar lidar com o problema de forma simplista é meio caminho andado para o desastre. Foi interessante, e ao mesmo tempo deprimente, ver como apareceram de repente especialistas em epidemiologia um pouco por todo o lado e, até confrangedor, ver alguns colegas da academia a darem o flanco às setas de forma tão infantil como foi feito. A elegância impede-nos de citar nomes ou exemplos concretos, mas foram frequentes as previsões disparatadas, por vezes feitas sob pressão jornalística. Algo que não é preciso numa pandemia é a existência de alegados especialistas ou peritos a ensaiarem análises ingénuas apenas para terem cobertura mediática. Torna-se fundamental um trabalho conjunto entre especialistas na matéria concreta, neste caso virologistas e epidemiologistas, em articulação, se necessário, com quem perceba os dados, os filtros que estes sofrem e a forma de os modelar. Pelo nosso lado, tentámos não sucumbir a essa tentação, e sendo que teria sido fácil apresentar coisas bem melhores do que foi frequente ver na comunicação social, achámos que no que diz respeito à modelação epidemiológica do fenómeno o nosso colega Baltazar Nunes liderava o grupo certo no local certo à hora certa. Tentámos, conscientes da nossa dimensão, contribuir para esse esforço comum em outras áreas.

Tentámos contribuir para os inquéritos serológicos nacionais – ferramentas que seriam fundamentais para perceber qual a proporção da população que já teria sido exposta ao vírus – mas não conseguimos convencer quem de direito que uma amostra aleatória seria fundamental. Essa talvez tenha sido, assumidamente, a nossa maior derrota neste processo. Como é que se pode pensar que uma amostra de

conveniência é uma boa base para fazer previsões sobre uma população? Ou que aumentar o tamanho dessa amostra poderá reduzir o enviesamento desse processo de amostragem de conveniência. Há ainda muito trabalho a fazer nesse sentido.

Foram ensaiados uma série de artigos na comunicação social, que pretenderam alertar a sociedade para conhecimentos fundamentais necessários neste contexto. Os meios de Comunicação Social digitais aderiram à nossa pretensão, mas infelizmente, não é claro se tal aconteceu por abraçarem a nossa causa ou porque anseiam por artigos contraditórios que promovam as audiências. Mas os fins justificam os meios e optámos por ganhar alguma visibilidade dessa forma.

Um dos projetos mais interessantes em que nos envolvemos, via FCUL – Faculdade de Ciências da Universidade de Lisboa, foi o COVIDTECT, em parceria com o Instituto Superior Técnico e as Águas de Portugal, que pretende relacionar cargas virais em águas residuais com o número de pessoas doentes na população. Há uma série de filtros entre as duas quantidades, mas parece cada vez mais evidente que as águas residuais poderão funcionar como um sistema de alerta precoce para a COVID19. Os resultados preliminares são encorajadores, e parece ser possível encontrar sinal nos dados que possam levar a que as águas residuais sejam uma forma de alerta precoce nesta e, eventualmente, noutras futuras pandemias. O problema, como tantas vezes acontece com dados reais, é a extrema variabilidade e a distinção entre sinal e ruído. Este projeto deu origem ao SARS Control, que pretende continuar a investigação e estender as análises não só à fração líquida das águas residuais como também à fração sólida das mesmas, além de perceber se as águas reutilizadas das ETAR são seguras. Sendo que em ambos os projetos o CEAUL tem um envolvimento secundário, parece-nos cada vez mais importante o papel fundamental que os estatísticos devem assumir em estudos interdisciplinares. Claro que para isso é preciso massa crítica e esse tem sido o principal problema do CEAUL nos últimos anos. A vontade existe, as condições e a massa crítica para dar resposta às solicitações teimam em tardar a concretizar-se.

Alguns conceitos estatísticos foram deturpados e abusados durante esta pandemia. Gonçalves et al. (2021) abordou um desses casos, em que ignorando a natureza de série temporal dos dados (número de novos casos diários), se confundiram conceitos abrindo a porta para erros nas previsões ao imaginar que um modelo puramente empírico poderia ser uma forma de prever o futuro. Mais uma vez, o cuidado a ter com o tratamento dos números foi evidenciado. Mas aqui surge também uma autocrítica à comunidade estatística em geral e a nós próprios em particular. Nesta epidemia, o lugar de copiloto foi muitas vezes ocupado confortavelmente pelos estatísticos. No entanto, está claramente identificada a necessidade de atuar em certas áreas da sociedade e ocupar o espaço existente. Caso isso não seja feito, corremos o risco de este espaço que devia ser ocupado por estatísticos, ser ocupado por outros eventualmente com menos competências, mas certamente com maior sentido de oportunidade.

Escrevemos hoje ainda sem saber onde vai terminar esta pandemia. Se há resultados encorajadores com a vacinação em larga escala da população mundial, há ainda muitas incógnitas, nomeadamente como irão as novas variantes do vírus reagir à vacina. Será esta eficaz? Qual será a duração do seu efeito? Estas, e outras perguntas, dependem de uma estatística forte e robusta para serem respondidas de forma cabal. Isto porque as respostas não são simples, nem tudo é branco ou preto. Existem muitos tons de cinzento, e é aí que a estatística pode e deve atuar. Da parte do CEAUL continuaremos a estar atentos e a contribuir sempre que acharmos que o podemos fazer com a qualidade que a importância da questão justifica. O exercer de uma cidadania plena depende hoje em dia de uma boa dose de Literacia Estatística e, além da investigação, é no sentido de aumentar essa literacia que, sempre que possível em articulação direta com a Sociedade Portuguesa de Estatística, atuamos em permanência.

## Referências

Gonçalves, L., Turkman, M. A. A., Geraldes, C., Marques, T. A. & Sousa, L. (2021). COVID-19: Nothing is normal in this pandemic. *Journal of Epidemiology and Global Health*. (DOI: 10.2991/jegh.k.210108.001)

- Marques, T. A. (2020). A estatística comanda a vida. Público Online, 20 outubro. Artigo temático sobre o Dia Mundial da Estatística 2020.
- Marques, T. A. (2020). Por que razão estamos sempre errados em relação ao futuro. Público Online, 22 abril.
- Marques, T. A. (2020). O caso da Vitamina D: Porque devemos ter muito cuidado com as notícias. Info FCUL, 3 abril 2020.
- Sousa, L., Marques, T. A. & Turkman, M. A. A. (2020). Não se deixe iludir pelas aparências. Não é tão simples como parece. Observador, 8 abril.



# Indicadores para a pandemia COVID-19 em Portugal

## *Uma leitura do contexto demográfico e da expressão territorial*

Instituto Nacional de Estatística

Com o intuito de acompanhar o impacto económico e social da COVID-19, o INE tem vindo a produzir novas estatísticas e a divulgar novos produtos (acessíveis a partir do portal do INE em “[Especial INE COVID-19](#)”), destacando-se a publicação quinzenal, iniciada a 9 de abril, com informação diária sobre o total de mortes ocorridas no território nacional, desde 1 de Março de 2020 até ao final da quinzena de referência. É analisada a heterogeneidade regional, tendo em conta os números absolutos de casos confirmados e de mortes, e indicadores relativos de acordo com a dimensão e densidade populacional por km<sup>2</sup> (municípios).

O impacto socioeconómico da pandemia é, também, diferenciado em todo o território, por exemplo, utilizando informação mensal sobre os levantamentos nas ATM e sobre indicadores mensais do mercado de habitação.

Neste âmbito específico, foram divulgados ao todo 18 Destaques específicos COVID-19 até à data, associados aos seguintes tópicos:

- “COVID-19: uma visão estatística integrando território e demografia” (9 de abril e 24 de abril);  
Nestes destaques é efetuado um estudo do impacto deste fenómeno recente, com base numa análise granular sobre os indicadores do COVID19 ao nível espacial (município), juntando informação sobre dinâmicas demográficas, tendo em conta indicadores de dimensão e densidade geográfica.
- “COVID-19: uma leitura territorial do contexto demográfico e do impacto socioeconómico” (8 de maio, 22 de maio, 5 de junho, 19 de junho e 3 de julho);  
Nestas divulgações, mantém-se a análise ao nível regional e de município do fenómeno, associado não só às dinâmicas demográficas, passando também a conter indicadores mensais na área socioeconómica. Passam a incluir ainda dados sobre mobilidade da população ao nível regional proporcionados pela iniciativa “Data for Good” do Facebook e que pode ser consultada no domínio do [Statslab do INE](#).
- “COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia” (11 de setembro, 25 de setembro, 9 de outubro, 23 de outubro, 4 de dezembro, 18 de dezembro, 15 de janeiro, 29 de janeiro e 12 de fevereiro);  
Continuação da divulgação de indicadores ao nível municipal e regional sobre dados da pandemia, fenómenos demográficos e associados ainda a diversos dados económicos.
- “COVID-19: o que distingue as 19 freguesias em estado de calamidade do resto da AML?” (17 de julho);  
A expressão da pandemia no território nacional continua a ser caracterizada por uma elevada heterogeneidade, situação que conduziu à declaração de medidas de política pública territorialmente diferenciadas. A manutenção do estado de calamidade num conjunto de 19 freguesias contíguas da Área Metropolitana de Lisboa motivou um olhar aprofundado sobre este território.
- “Mobilidade da população ao nível regional no contexto da pandemia COVID-19” (20 de novembro);  
Tirando partido dos dados da iniciativa “Data for Good” do Facebook, é possível obter informação sobre a mobilidade da população numa perspetiva regional. Estas análises

permitem obter informação sobre volume de deslocações entre os períodos de Estado de Emergência, períodos de desconfinamento, e também outros momentos como a circulação entre municípios em feriados.

Com esta iniciativa, o INE pretende disponibilizar nova informação para auxílio dos investigadores, decisores e para conhecimento da Sociedade em geral.

**Destaques, respetivos títulos e datas de divulgação** (com links para documentos)

- [COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia, 12 de fevereiro 2021](#)
- [COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia, 29 de janeiro 2021](#)
- [COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia, 15 de janeiro 2021](#)
- [COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia, 18 de dezembro 2020](#)
- [COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia, 4 de dezembro 2020](#)
- [Mobilidade da população ao nível regional no contexto da pandemia COVID-19, 20 de novembro 2020](#)
- [COVID-19: uma leitura do contexto demográfico e da expressão territorial da pandemia, 23 de outubro 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 9 de outubro 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 25 de setembro 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 11 de setembro 2020](#)
- [O que distingue as 19 freguesias em estado de calamidade do resto da AML?, 17 de julho 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 3 de julho 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 19 de junho 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 5 de junho 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 22 de maio 2020](#)
- [Uma leitura territorial do contexto demográfico e do impacto socioeconómico, 8 de maio 2020](#)
- [Uma visão estatística integrando território e demografia, 24 de abril 2020](#)
- [Uma visão estatística integrando território e demografia, 9 de abril 2020](#)



# Ciência Estatística

## • Artigos em Revistas

Gonçalves, L., Turkman, M. A. A., Geraldês, C., Marques, T. A. & Sousa, L. (2021). COVID-19: Nothing is normal in this pandemic. *Journal of Epidemiology and Global Health*. (DOI: 10.2991/jegh.k.210108.001)

## • Teses de Mestrado

**Título:** Estudo das atitudes em relação à Estatística de estudantes do 3º ciclo do ensino básico

**Autora:** Ana Julieta Oliveira Morais, [leta.morais@gmail.com](mailto:leta.morais@gmail.com)

**Orientadora:** Adelaide Freitas

(com as devidas desculpas; esta informação corrige a correspondente da página 81 do Boletim SPE outono 20)

**Título:** Modelação e predição do valor do tempo de vida do cliente (*Customer Lifetime Value*)

**Autora:** Mariana Mourão, [mc.mourao@campus.fct.unl.pt](mailto:mc.mourao@campus.fct.unl.pt)

**Orientadora:** Regina Bispo

**Título:** Modelos preditivos de abandono do consumo

**Autora:** Sara Machado <[sa.machado@campus.fct.unl.pt](mailto:sa.machado@campus.fct.unl.pt)>

**Orientadora:** Regina Bispo

**Título:** Um Novo Pacote no R para Visualizar Dados em Análise de Sobrevivência

**Autor:** Pedro Daniel Ponte Camacho, [2016711@student.uma.pt](mailto:2016711@student.uma.pt)

**Orientadora:** Ana Maria Abreu

**Título:** Bayesian Generalized Additive Models for Car Insurance Data

**Autor:** João Pedro Rodrigues Góis, [joapedrogois5@gmail.com](mailto:joapedrogois5@gmail.com)

**Orientador:** Giovanni Silv

**Título:** Copula-Based Joint Survival Modelling of Overall Survival and Disease Free Times: A Study of Breast Cancer Data

**Autora:** Beatriz dos Santos Lourenço, [bds.lourenco96@gmail.com](mailto:bds.lourenco96@gmail.com)

**Orientador:** Giovanni Silva

## • Livros

Na página da SPE, <https://www.spestatistica.pt/>, estão disponíveis as edições digitais de:

***Introdução à Probabilidade e à Estatística, Com complementos de Excel***

Ano: 2005

Autora: Maria Eugénia Graça Martins

Estado: Disponível para download

***Memorial da Sociedade Portuguesa de Estatística***

Ano: 2005

Editor: Fernando Rosado

Estado: Disponível para download



# SPE

Sociedade Portuguesa  
de Estatística

## Prémio SPE 2020

### A new lifetime distribution and its applications in recurrent events analysis

Ivo Sousa-Ferreira, [ivo.ferreira@staff.uma.pt](mailto:ivo.ferreira@staff.uma.pt)  
CEAUL e DEIO, Faculdade de Ciências da Universidade de Lisboa

Neste trabalho, são apresentadas duas contribuições para a análise estatística de dados de sobrevivência: uma nova distribuição do tempo de vida indexada por três parâmetros, denominada distribuição Chen-Poisson; e um modelo de taxa marginal para acontecimentos recorrentes baseado na distribuição proposta.

A distribuição Chen-Poisson surge em cenários de riscos competitivos ou complementares (CCR), onde apenas é possível observar o tempo de vida mínimo ou máximo entre todas as causas passíveis de desencadear o acontecimento de interesse, em vez de se observar o tempo associado a cada causa em particular. Neste contexto, muitas vezes não se sabe qual das causas é responsável pela ocorrência do acontecimento, nem se conhece o número total de causas existentes. Recentemente, Ramos *et al.* (2019) mostraram que quando se assume que o número de causas é uma variável aleatória com distribuição de Poisson truncada em zero (ZTP), ambas as distribuições do mínimo e do máximo podem ser unidas de uma forma simples, estendendo o espaço-parâmetro da distribuição ZTP ao conjunto dos números reais exceto o zero. Com base neste resultado, desenvolveu-se a distribuição composta Chen-Poisson, onde se assume que os tempos associados às causas seguem uma distribuição de Chen (2000). Posteriormente, foram estudadas várias propriedades da nova distribuição, incluindo a função de sobrevivência, a função de risco, a função quantil, a função geradora de momentos e a função de vida residual média. A distribuição Chen-Poisson é bastante flexível, pois a sua função de risco consegue acomodar uma forma monótona crescente, monótona decrescente, unimodal, *bathtub*, crescente-decrescente-crescente ou decrescente-crescente-decrescente-crescente. Para estimar os três parâmetros da distribuição considerou-se o método de estimação de máxima verosimilhança (MLE) na presença de um mecanismo de censura à direita. Através de um estudo de simulação verificou-se a eficácia do método MLE em cenários com uma percentagem de censura baixa ou nula.

Com o intuito de expandir a aplicabilidade da distribuição Chen-Poisson para modelar os tempos entre acontecimentos recorrentes (*gap times*), foi desenvolvido um modelo de taxa marginal baseado em Zhao e Zhou (2012). Seguindo a abordagem destes autores, o que se propôs foi considerar uma forma paramétrica para a função de taxa marginal derivada de um processo de Poisson não homogêneo, tendo por base a nova distribuição. Foi então deduzida a distribuição condicional dos *gap times* a partir de uma função de taxa marginal Chen-Poisson, de onde se obteve um modelo paramétrico para acontecimentos recorrentes.

Devido à flexibilidade da nova distribuição, bem como do modelo para acontecimentos recorrentes, os resultados obtidos na análise de dois conjuntos de dados reais disponíveis na literatura revelaram que os modelos propostos são uma melhor opção para estes conjuntos de dados comparativamente aos já existentes.

O trabalho laureado encontra-se publicado nas Notas e Comunicações do Centro de Estatística e Aplicações da Universidade de Lisboa (Sousa-Ferreira *et al.*, 2020) e um artigo correspondente a uma parte do mesmo foi submetido para publicação numa revista científica internacional.

#### Referências

- Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, 49(2), 155-161.
- Ramos, P. L., Dey, D. K., Louzada, F. e Lachos, V. H. (2019). An extended Poisson family of life distribution: a unified approach in competitive and complementary risks. *Journal of Applied Statistics*, 47(2), 306-322.
- Sousa-Ferreira, I., Abreu, A. M. e Rocha, C. (2020). A new lifetime distribution and its applications in recurrent events analysis. Notas e Comunicações CEAUL 02/20.
- Zhao, X. e Zhou, X. (2012). Modeling gap times between recurrent events by marginal rate function. *Computational Statistics & Data Analysis*, 56(2), 370-383.

Ivo Sousa-Ferreira, **galardoado com o Prémio SPE 2020**, é Licenciado em Gestão e Mestre em Matemática pela Universidade da Madeira. Atualmente está a realizar o Doutoramento em Estatística e Investigação Operacional na Faculdade de Ciências da Universidade de Lisboa, sob a supervisão da Professora Doutora Cristina Rocha e da Professora Doutora Ana Maria Abreu, com o apoio da Bolsa de Doutoramento DFA/BD/6459/2020 concedida pela Fundação para a Ciência e a Tecnologia.



# Índice

Editorial .....	2
Mensagem do Presidente .....	4
Notícias .....	6
<i>Enigmística</i> .....	12
<b><i>Episódios na História da Estatística</i></b> .....	<b>13</b>
<b><i>SPE e a Comunidade</i></b>	
Contributo do Instituto Nacional de Estatística .....	15
<b><i>Especial Covid: a Estatística ao serviço da sociedade</i></b>	
O Papel da Estatística no Combate à Pandemia COVID-19	
<i>João A. Branco</i> .....	17
De passagem pela mortalidade...	
<i>Clara Cordeiro</i> .....	22
<i>Algumas considerações sobre a COVID-19</i>	
<i>Isabel Natário</i> .....	25
CoronaSurveys na estimação de casos ativos de Covid-19	
<i>Raquel Menezes e Carlos Baquero</i> .....	35
Estatística no CEAUL em tempos de COVID	
<i>Tiago A. Marques e Lisete Sousa</i> .....	40
Indicadores para a pandemia COVID-19 em Portugal/ Uma leitura do contexto demográfico e da expressão territorial	
<i>Instituto Nacional de Estatística</i> .....	43
<b><i>Ciência Estatística</i></b>	
<i>Artigos Científicos</i> .....	45
<i>Teses de Mestrado</i> .....	45
<i>Prémio SPE 2020</i> .....	46