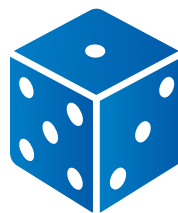


Boletim



SPE

Sociedade Portuguesa
de Estatística

Publicação semestral

outono de 2025



INE – 90 anos de rigor e inovação ao serviço da Sociedade

Preâmbulo	
<i>Instituto Nacional de Estatística, I.P.</i>	17
A Inovação no INE – 20 exemplos dos últimos 20 anos	
<i>Instituto Nacional de Estatística, I.P.</i>	19
Not Just Another Black Box: AI in the Age of Trusted Statistics	
<i>Sónia Quaresma</i>	48
Revisão do Regulamento das Estatísticas Europeias Desafios e oportunidades para os sistemas estatístico	
<i>Instituto Nacional de Estatística, I.P.</i>	53
Manual do Processo de Produção Estatística do Instituto Nacional de Estatística	
<i>Instituto Nacional de Estatística, I.P.</i>	56
ALEA – Um marco na promoção da literacia estatística: uma visão em 3 partes	
O ALEA e a produção de Estatísticas Oficiais	
<i>Pedro Campos</i>	61
O ALEA na Educação para a Matemática	
<i>Emília Oliveira</i>	61
O ALEA e a Educação em Portugal	
<i>João Gonçalves e Franck de Almeida</i>	62
Capacitação em Ciência de Dados para o Sistema Estatístico Nacional e Administração Pública	
<i>Pedro Campos e Jorge Magalhães</i>	64

Editorial	2
Mensagem do Presidente	4
Notícias	5
<i>Enigmística</i>	16
INE – 90 anos de rigor e inovação ao serviço da Sociedade	17
Ciência Estatística	67
Edições SPE - Minicursos	78
Boletim através do Tema Central	79

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

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

ISSN: 1646-5903

Depósito Legal: 249102/06

Tiragem: Edição digital

Execução Gráfica e Impressão: Gráfica Sobreireense

Editor: Fernando Rosado, fernando.rosado@fc.ul.pt

Sociedade Portuguesa de Estatística desde 1980



SPE

Sociedade Portuguesa
de Estatística

<https://www.spestatistica.pt>

Sociedade Portuguesa de Estatística desde 1980

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†: grátis!

† Até um ano após terminar a licenciatura.

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

Editorial

... sobre o valor curricular e não só; na história dos Congressos SPE...

Obviamente, que é muito difícil – tarefa que o passar do tempo ainda mais agrava – saber em quantas reuniões científicas participamos ao longo da nossa vida curricular. De todas elas e em cada uma, um objetivo se regista. Por esta ou por aquela razão; este ou aquele encontro internacional toma maior relevo na nossa memória. Mas se, na pesquisa, associarmos a ligação à comunidade científica nacional, decerto os Congressos da Sociedade Portuguesa de Estatística – porque, para os jovens investigadores, também são reuniões de iniciação – eles entram num campo memorial que agrega um pouco, arrisco mesmo, muito, da história da Estatística. Na minha longa carreira, disponho de memórias em congressos que me deixam muito orgulhoso. Passados 27 Congressos SPE estou incluído no grupo daqueles que em todos eles disseram: Presente! Estamos, pois, no subgrupo das reuniões científicas que apelidamos de Congressos SPE. Claro, também entre estes, surge aquele que, por este ou por aquele detalhe, o torna “diferente” – porque foi nele “a minha primeira comunicação oral” ou porque “no convívio científico” descobri “a minha área de eleição”.

Primorosamente organizado, terminou o XXVII Congresso SPE. A Comissão Organizadora Local, merece o maior reconhecimento. A seguir, neste Boletim, registamos “a história” dessa excelente organização. Pois, que mais pode o Editor dizer do ido XXVII Congresso SPE quando, com tão bela vivência, que uniu Sêniores e Jovens, em especial nas sessões de Prémios – Carreiras, SPE e Estatísticos Júnior – sentiu a presença do carinho, a generosidade e, permitam-me a vaidade que a humildade tolera, viveu palavras e momentos únicos nas mais variadas emoções – desde o reconhecimento científico até ao Amor Familiar. Foi tudo tão intenso que para o resumir, seguramente a palavra não basta. Mas, com a maior força, desejo enviar a minha maior Gratidão, a toda a comunidade, onde incluo o generoso júri do Prémio Carreira.



Por fim e não o menor, saliento, uma vez mais, a importância do convívio e a partilha científica entre as gerações, que testemunhei; e desta vez a grande participação de muitos jovens estatísticos investigadores do todo nacional e que muito bem manifestaram a vontade de viver e fazer crescer a SPE.

... sobre o valor da Estatística nas Estatísticas...

Esta edição do Boletim SPE foca um Tema Central de especial relevo.

O *Boletim outono de 2019*, como sabemos, teve o privilégio de iniciar uma colaboração periódica e mais regular por parte do *Instituto Nacional de Estatística – INE*.

A secção *SPE e a Comunidade* ficou, assim, mais rica.

O *Boletim primavera de 2020*, em data aniversária, com enorme prazer recebeu o INE e os seus convidados que nas nossas páginas transmitiram informação, conhecimento e história.

O *Boletim outono de 2025* também se sente honrado ao fazer parte integrante das comemorações que neste ano decorrem no *Instituto Nacional de Estatística*.

A SPE e o INE desde sempre estiveram muito próximos na partilha dos mais diversos acontecimentos estatísticos nacionais e internacionais. São inúmeras as ações em que a SPE teve o apoio do INE, fundamentais para a sua concretização. Em 2005, por ocasião da edição do livro *Memorial da SPE*, na celebração dos 25 anos SPE, tivemos a participação de dois antigos Presidentes, Manuel José Vilares e Paulo Gomes que contribuíram, a páginas 125 e 163, com textos sobre o Sistema Estatístico Nacional. Como testemunho pessoal, porque tive a felicidade de viver intensamente esse momento enorme de afirmação da Estatística e dos estatísticos portugueses no panorama da comunidade internacional; permito-me incluir no Memorial dos 90 anos do INE que aqui perpetuamos: a realização em Lisboa, do *ISI 2007*, do *International Statistical Institute*, levado a cabo por uma equipa científica liderada pela

Prof. Ivette Gomes é uma executiva comandada pelo Prof. Paulo Gomes. Foi um marco nacional e internacional onde a SPE e o INE inscreveram Portugal na lista de referência do *ISI*; com um número record de inscrições nessa magna reunião dos estatísticos de todo o mundo.

O Boletim SPE outono 2025, teve a preciosa colaboração do colega Carlos Marcelo, como coeditor e também ele um “obreiro SPE” de longa data.

Na pessoa do Prof. António Rua, Presidente do Instituto Nacional de Estatística, reconhecidamente, o Boletim SPE felicita pela data aniversária e por toda a colaboração que o INE, sócio institucional da SPE, tem prestada a esta edição SPE.

... sobre o valor da Estatística como Ciência “autónoma”...

Ou, por outras palavras, Estatística *versus* Matemática!

Os Sêniores na Estatística de hoje, iniciaram carreiras de estudo e investigação num Portugal onde a Estatística como Ciência não existia. A Estatística apenas referia “números” produzidos, por exemplo do INE; e para os quais o poder de então nada desejava que se “soubessem” analisar. Há 50 anos, os Sêniores de hoje estudaram Matemática onde se iniciou uma Matemática Aplicada – um grande avanço na época. Muito dinamizados pela facilidade computacional entretanto surgida foram dados os primeiros passos para uma Ciência autónoma que construiu um percurso científico a partir de bases teóricas de eminentes cientistas, principalmente da escola inglesa, que geraram a palavra Estatística para o nome próprio dessa ciência que desde então se foi e cada vez mais, autonomizando da Matemática. Em muitos domínios ainda com uma grande intersecção com a Matemática existem, no entanto, várias áreas da Estatística onde o “gene matemática” já não permite que sejam identificadas com a sua antepassada Matemática.

No presente e já com alguma intensidade, sentimos o valor do debate da referida autonomia. E de tal modo ele é importante que está na própria Matemática alguma dificuldade em aceitar a Ciência Estatística como um dos seus descendentes, por si gerados – desde a Geometria à Economia (reconhecida pela fundação Nobel) ou à Ciência dos Dados, a tão famosa *Data Science*, que está na moda.

Para melhor construir, estudar e aprofundar modelos e métodos de apoio à decisão (a todos os níveis) é fundamental que se reconheça o valor do trabalho do Estatístico, isto é, o verdadeiro trabalhador da Estatística que, sendo Matemática, é progenitora de um ramo científico autónomo.

E, esta tarefa tem sido e continua a ser uma missão estatutária da SPE (Art. 1º nº 2).

Que cada sócio seja um arauto desse desiderato!

O Tema Central do *Boletim primavera de 2026* será:

BioEconomia no Crescimento Estocástico – Uma Contribuição Portuguesa



Mensagem do Presidente

Caras e caros colegas,

Em nome da Sociedade Portuguesa de Estatística, deixo um agradecimento muito sincero a toda a comunidade que tornou possível o XXVII Congresso da SPE, realizado em Faro, no Campus da Penha da Universidade do Algarve, de 22 a 25 de outubro de 2025.

Um reconhecimento especial à Comissão Organizadora Local, pelo trabalho exemplar, sempre discreto quando necessário e plenamente presente nos momentos decisivos. Agradeço também à Comissão Científica, pela seleção criteriosa que garantiu a qualidade e a diversidade do programa. Estendo ainda este agradecimento aos responsáveis pelas sessões temáticas e a todas as pessoas que apresentaram comunicações, moderaram sessões ou participaram ativamente nos debates.

Aos nossos convidados plenários, Ana Luísa Papoila, Cláudia Neves, Przemyslaw Biecek e Renato Assunção, deixo um agradecimento caloroso pela generosidade científica e pela inspiração que trouxeram ao encontro. Destaco, em particular, o minicurso *The Hitchhiker's Guide to Responsible Machine Learning* e a publicação em língua portuguesa associada, preparada com a colaboração de Nuno Sepúlveda e com o apoio imprescindível do Instituto Nacional de Estatística (INE) na edição e impressão. Este contributo reforça a missão da SPE de disseminar conhecimento e valorizar a formação contínua.

Uma palavra de reconhecimento aos patrocinadores e entidades parceiras pelo apoio que nos permitiu melhorar condições e ampliar o impacto do congresso.

Foi também um momento feliz o reconhecimento público do trabalho de duas personalidades que muito deram à Estatística em Portugal e à SPE, na sessão do Prémio Carreira SPE: Professor Fernando Rosado e Professora Maria Eugénia Martins. A ambos, a nossa gratidão e aplauso.

Num ano particularmente exigente para a nossa comunidade, o elevado número de participantes e a energia que se sentiu em Faro confirmam a vitalidade da SPE e a importância destes momentos de encontro, colaboração e partilha.

Olhando agora em frente: recordo-vos com satisfação que o XXVIII Congresso da SPE decorrerá em 2027, na cidade de Leiria. A equipa local, do Instituto Politécnico de Leiria já se encontra em preparação e, oportunamente, partilharemos mais informações. Contamos convosco para dar continuidade a este espírito de comunidade e excelência.

Como ficou claro na Assembleia Geral realizada durante o congresso, a SPE tem pela frente desafios significativos. Independentemente das decisões que venham a ser tomadas numa próxima reunião da Assembleia Geral, é essencial reforçar a visibilidade e a afirmação da Estatística no espaço científico e social, promover a literacia estatística e continuar a apoiar a nossa comunidade académica e profissional. São temas que permanecerão no centro da nossa ação e que exigem o contributo e o envolvimento de todos.

Com os melhores cumprimentos,

Luís Meira Machado

Presidente da Sociedade Portuguesa de Estatística (SPE)

Notícias

• SPE 25 – XXVII Congresso da Sociedade Portuguesa de Estatística

Conhecimento, partilha e inspiração – tudo aconteceu em Faro

O Congresso da Sociedade Portuguesa de Estatística (SPE) decorreu no Campus da Penha da Universidade do Algarve, em Faro, e reuniu, ao longo de quatro dias, investigadores, docentes e profissionais de diversas áreas da estatística, ciência de dados e inteligência artificial. O evento constituiu um espaço privilegiado de debate, partilha e aprendizagem, onde o conhecimento científico se cruzou com a inovação e a colaboração.

Dia 1

O primeiro dia abriu com entusiasmo. Após o acolhimento dos participantes, o minicurso *The Hitchhiker's Guide to Responsible Machine Learning*, orientado por Przemyslaw Biecek, deu o tom científico e dinâmico do congresso.



As atividades oficiais do congresso arrancaram formalmente com a sessão de abertura, às 16h, onde estiveram presentes a Pró-Reitora Ana Conceição (em representação do Reitor), Luís Machado (Presidente da SPE), Conceição Ribeiro (em representação do diretor do ISE, Paulo Santos, e presidente da Comissão Organizadora), Carlos Guerrero (diretor da FCT) e Clara Cordeiro (presidente da Comissão Organizadora).



Seguiu-se a Sessão Plenária *Statistics and Artificial Intelligence in Medicine*, apresentada por **Ana Luísa Papoila**, sublinhando a importância crescente da estatística no apoio à decisão clínica e à investigação biomédica.



O final da tarde ficou marcado pelas primeiras comunicações orais, que abordaram temas como a anonimização de microdados, as séries temporais, a ciência de dados e a aplicação da estatística ao ambiente e à agricultura. O dia encerrou com a receção de boas-vindas no Museu Municipal de Faro, um momento de convívio e *networking* entre os participantes.



Dia 2

O segundo dia começou cedo, com sessões temáticas que cruzaram fronteiras e continentes. A sessão **“Estatística em Português – Estatística com África”**, organizada por **Giovanni Silva** e **Rita Gaio**, destacou a cooperação científica entre Portugal e os países africanos de língua portuguesa (PALOP). A sessão temática **“Paving the Way: Making Careers in Statistics Education Research Visible and Viable”**, organizada por **Bruno de Sousa**, contou com a participação de **Iddo Gal**, **Dani Ben-Zvi** e **Katie Makar**, que abordaram a importância da investigação na educação estatística e a visibilidade das carreiras académicas neste domínio. Outras sessões exploraram séries temporais, estatística multivariada e a importância de tornar a investigação estatística mais visível e valorizada.

Durante a manhã decorreu a primeira sessão de posters, que contou com uma ampla e dinâmica participação. Os trabalhos apresentados abrangeram temas variados, desde a modelação estatística em saúde pública até à aplicação de métodos de aprendizagem automática, evidenciando a diversidade e qualidade da investigação desenvolvida pela comunidade estatística nacional e internacional.



As sessões plenárias de **Przemyslaw Biecek** e **Renato Assunção** aprofundaram a reflexão sobre a análise de modelos explicativos e sobre a eficácia da ciência de dados.



A tarde terminou com o tradicional passeio do congresso: um passeio a pé pela cidade velha de Faro, e um passeio de barco pela Ria Formosa, uma oportunidade descontraída de continuar a conversa fora das salas.



Dia 3

O terceiro dia do congresso manteve o ritmo intenso e inspirador dos anteriores. A manhã foi marcada por várias sessões paralelas, com destaque para a sessão temática da Comissão Especializada de Educação da SPE, centrada na importância da literacia estatística e na ligação entre a estatística e a cidadania.

Durante a manhã, teve ainda lugar a segunda sessão de posters, que reuniu um conjunto diversificado de trabalhos de investigação, refletindo a vitalidade e o carácter multidisciplinar da comunidade estatística.

Um dos momentos mais aguardados do dia foi a Palestra Plenária de **Cláudia Neves**, intitulada “**One Way to Estimate an Out-of-Sample Quantile of an Unknown Distribution Through Extreme Value Theory**”.



A parte da tarde começou com a tradicional fotografia de grupo, uma amostra representativa do congresso e, como sempre, um momento de bela distribuição de sorrisos, com ligeira assimetria à direita e algumas observações em falta.

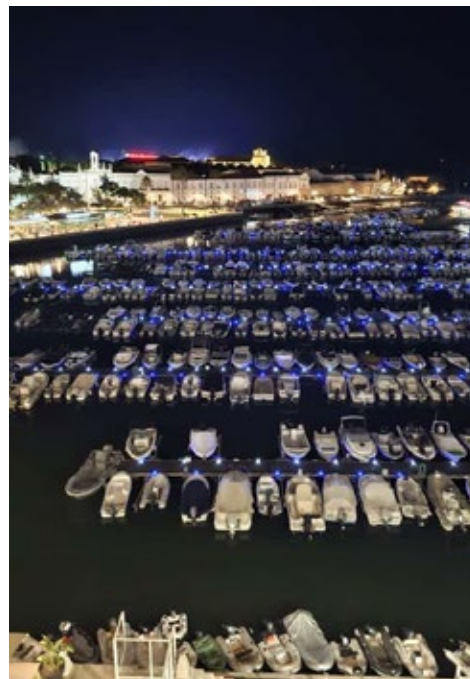


De seguida, a tarde prosseguiu com novas sessões temáticas e comunicações orais que abordaram desde métodos de reamostragem em estatística extrema até aplicações inovadoras da inteligência artificial em contextos industriais e de saúde.

O ponto alto do dia foi, contudo, a Cerimónia de Atribuição do Prémio Carreira da SPE, um momento particularmente emotivo que distinguiu duas personalidades de referência da comunidade estatística portuguesa – os professores **Fernando Rosado** e **Maria Eugénia Graça Martins** - pelo seu contributo excecional para o desenvolvimento da estatística e para o fortalecimento da Sociedade.



O dia encerrou com a Assembleia Geral da SPE, onde foram debatidos temas estratégicos para o futuro da Sociedade, e culminou com o Jantar do Congresso, realizado no Hotel EVA Senses, em Faro.



Este momento de convívio e celebração reuniu participantes, oradores e membros da Direção num ambiente de franca amizade e partilha, durante o qual foram anunciados os organizadores do próximo Congresso da SPE, que decorrerá no Instituto Politécnico de Leiria.

Dia 4

O último dia do congresso iniciou-se com um conjunto de sessões dedicadas à Bioestatística, Epidemiologia, Séries Temporais e Ciência de Dados, refletindo a diversidade e vitalidade da investigação estatística em Portugal.

Seguiu-se a sessão dedicada à apresentação dos trabalhos distinguidos com os Prémios SPE 2024 e 2025 - atribuídos a **Ana Martins** e **Lídia André**, respetivamente - que reconheceram a excelência de jovens investigadores, sócios da SPE, nas áreas da Estatística e da Ciência de Dados.

Nesta ocasião, foram entregues os respetivos diplomas de distinção e feita igualmente menção aos vencedores do Prémio Estatístico Júnior.

A sessão de Encerramento marcou o fim de quatro dias de ciência, partilha e amizade. Ficou a promessa de reencontro no próximo congresso da SPE, em 2027, e o sentimento comum de pertença a uma comunidade unida pelo rigor científico e pela paixão pela Estatística.



E assim, o XXVII SPE25 que muito nos honrou, encerra por agora a atividade de partilha científica, convívio e celebração entre os cerca de 200 participantes.

XXVII CONGRESSO

SPE25

SOCIEDADE PORTUGUESA DE ESTATÍSTICA

A página do Congresso [SPE2025](#) mantém-se ativa, em particular, para informações complementares sobre a Edição das Atas.

Agradecemos a todos os participantes e oradores que fizeram do **XXVII Congresso da SPE** um verdadeiro encontro de ciência e amizade.

Até breve e que a inspiração nunca caia fora do intervalo de confiança!

A Comissão Organizadora do SPE 2025



Prémio Carreira SPE

O Prémio Carreira SPE (Cf. Regulamento) visa reconhecer a atividade de estatísticos portugueses com papel de relevância no desenvolvimento científico, pedagógico e de divulgação da Estatística em Portugal. Este prémio de distinção de carreira é atribuído a uma ou mais personalidades a quem seja reconhecido um papel relevante, com nível de excelência, no desenvolvimento e divulgação da Estatística em Portugal e com obra científica de elevado mérito. Em 2025 a SPE atribuiu Prémio Carreira a Eugénia Graça Martins e Prémio Carreira a Fernando Rosado.

Na secção Ciência Estatística, desta edição *Boletim SPE outono de 2025* apresentamos o desenvolvimento desta notícia.

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 Prémio SPE 2024

- Luís Meira Machado – Presidente da SPE
- Carlos Tenreiro (UC)
- Irene Oliveira (UTAD)
- Gonçalo Jacinto (UEvora)

O **Prémio SPE 2024** foi atribuído a Ana Martins (Universidade de Aveiro), pelo trabalho:

“On the theory of spatio-temporal models for time series of counts”.

No final desta edição é apresentado o desenvolvimento desta notícia.

Júri Prémio SPE 2025

- Luís Meira Machado – Presidente da SPE
- Isabel Pereira (UA)
- Helena Mouriño (ULisboa)
- Marta Ferreira (UMinho)

O **Prémio SPE 2025** foi atribuído a Lídia André (Lancaster University), pelo trabalho:

“Neural Bayes Inference for Complex Bivariate Extremal Dependence Models”.

O júri, em 2025, decidiu ainda atribuir uma menção honrosa ao trabalho de **Daniela Silva**, intitulado *“A Zero-Inflated Spatio-Temporal Approach for Joint Modeling of Fishery-Dependent and Fishery-Independent Data to Understand Fish Distribution”*, destacando a sua elevada qualidade científica.

Prémio Estatístico Júnior

A Sociedade Portuguesa de Estatística promove, todos os anos, o Prémio Estatístico Júnior (PEJ) com o intuito de estimular e desenvolver o interesse dos estudantes dos Ensinos Básico e Secundário pelas áreas de Probabilidades e Estatística.

Este Prémio é financiado pela FCT através do projeto DOI: [10.54499/UIDB/00006/2020](https://doi.org/10.54499/UIDB/00006/2020)

Júri dos Prémios Estatístico Júnior (2024 e 2025)

- Lígia Rodrigues
- Arminda Manuela Gonçalves
- Rui Martins

No final desta edição é apresentada a lista dos Premiados Estatístico Júnior.

Prémio Jornalismo de Dados SPE

O **Júri do Prémio Jornalismo de Dados** é constituído por:

- Luís Machado
- Luísa Loura
- André Sendin (Presidente da Escola Superior de Comunicação Social)

O leitor pode consultar em [Prémio Jornalismo de Dados SPE](#) os vencedores já divulgados.

FR

• Modern Trends in Time Series



A 3.^a edição do **oneMAST** na Universidade de Aveiro é dedicada a um dos temas mais atuais e desafiantes: **Séries Temporais**. O programa inclui:

- ◆ **Sessão de Contribuições (Contributed Session)** – destinada a apresentações orais em tópicos avançados de Estatística por participantes selecionados
- ◆ **Sessão de Jovens Investigadores (Young Researchers Session)** – no formato de **pitch oral rápido** + **poster em formato *graphical abstract***, para a partilha da investigação realizada por alunos de mestrado, alunos de doutoramento, bolsiros e jovens investigadores com interesse em método avançados de Estatística.
- ◆ **3 Sessões Plenárias** com convidados especialistas:
 - Patrícia de Zea Bermudez (FCUL)
 - João Gama (FEUP)
 - David Stoffer (University of Pittsburgh)
- ◆ **Mesa Redonda:** *The Future of Time Series in a Data-Driven World*.

Convite à participação: São bem-vindas submissões (**título + resumo até 250 palavras**) no tema central desta edição ou em qualquer área da **Estatística**.

🔗 Programa atualizado, registo e submissão de trabalhos em
<https://sites.google.com/view/one-daymeeting-cidma>

Datas importantes

- Submissão de resumos: **até 14 de janeiro**
- Notificação de aceitação: **até 21 de janeiro**
- Registo no evento: **até 26 de janeiro**
- Envio do *graphical abstract*: **até 4 de fevereiro**

Não perca a oportunidade de partilhar conhecimento, trocar ideias e explorar o futuro das séries temporais.

Junte-se a nós no **oneMAST 2026**, dia 6 de fevereiro, na Universidade de Aveiro!

A comissão organizadora:

Adelaide Freitas
Magda Monteiro
Luís Silva

• IV Dia Mundial da Estatística

Numa iniciativa da Organização das Nações Unidas - ONU, neste ano, em 20 de outubro, comemorou-se o quarto [Dia Mundial da Estatística 2025](#) enquanto também decorre o octogésimo aniversário da fundação da ONU.



O desenho gráfico do símbolo, pretende mostrar a importância da confiança nos dados no mundo atual, que é o tema central do Dia Mundial da Estatística deste ano. O logótipo reflete isso com o uso de um “check” que se mistura com um gráfico de barras cujos dados fluem para o globo terrestre. O envolvimento destes elementos num círculo completo também retrata a segurança. Tal como nos logótipos de 2015 e 2020, o esquema de cores azul continua a ser utilizado como ligação ao facto de o azul ser uma das cores oficiais das Nações Unidas, além de transmitir universalidade e profissionalismo.

Mensagem do Secretário-Geral, António Guterres, para 2025:

UNITED NATIONS



NATIONS UNIES

THE SECRETARY-GENERAL

--

MESSAGE FOR WORLD STATISTICS DAY

20 October 2025

World Statistics Day is an occasion marked once every five years to spotlight the vital role of statistics in helping to address the challenges of our time.

The Sustainable Development Goals have helped transform statistical systems and increase the availability of data. In times of crisis, such as the COVID-19 pandemic, statistical systems proved their resilience, delivering the data that guided life-saving decisions as well as social and economic recovery.

As we confront increasingly inter-connected global challenges – from climate change to rising inequalities – timely, accurate, disaggregated, and independent data has never been more essential.

This year’s observance reaffirms our commitment to the Fundamental Principles of Official Statistics: impartiality, professionalism, and commitment to scientific principles and professional ethics. Citizens everywhere deserve trustworthy data to inform public policy and strengthen accountability.

Let us use this day to champion the power of data in advancing sustainable development for all.

Como um meio acrescido de oportunidades para refletir sobre o papel vital das estatísticas e, acima de tudo, da Estatística o tema escolhido para este ano foi *Estatística e Dados de Qualidade para todos* como refere a página oficial: “Dados de qualidade devem estar acessíveis para todos e não apenas para os especialistas ou instituições e assim melhorar e tornar mais poderosas as decisões tomadas em todo o mundo”.

Durante o dia decorreu uma webinar – 24 horas.



III Dia Mundial



II Dia Mundial



I Dia Mundial da Estatística

(Cf. Boletim SPE outono de 2010, 2015 e 2020)

FR



• A SPE em Sessão Convidada no Congresso Anual da Royal Statistical Society

Entre os dias 1 e 4 de setembro deste ano, realizou-se em Edimburgo, Escócia, o congresso anual da Royal Statistical Society, que contou com a participação da Sociedade Portuguesa de Estatística (SPE) numa sessão convidada intitulada “Advances in Statistical Methods for Diverse Applications”.

A sessão, foi organizada pelo presidente da SPE, Professor Luís Machado, da Universidade do Minho, e pela Professora Vanda Lourenço, da Universidade NOVA de Lisboa.

Na sessão participaram também os seguintes oradores convidados:

A Professora Paula Brito, da Universidade do Porto e o Professor Nuno Sepúlveda, da Universidade Técnica de Varsóvia e colaborador do CEAUL.



A “delegação SPE” no Congresso da RSS

A Professora Paula Brito, apresentou o trabalho “*Statistical models for distributional data*”. O Professor Nuno Sepúlveda, falou sobre a sua investigação em “*Prevalence estimation of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome among long COVID patients using summary symptoms’ data*”. A Professora Vanda Lourenço, da Universidade NOVA de Lisboa, apresentou o trabalho “*On the robustness of random forests for genomic prediction and selection in breeding studies*”.

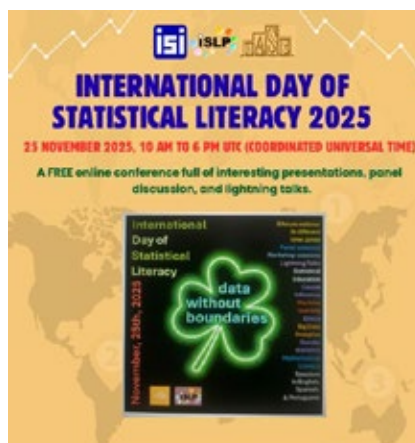
A moderação da sessão esteve a cargo do Professor Miguel de Carvalho, da Universidade de Edimburgo e ex-presidente da SPE.

A sessão foi muito bem acolhida pelo público, tendo a sala ficado completamente preenchida, apesar de ser de pequenas dimensões. A participação da SPE neste congresso constituiu uma excelente oportunidade de divulgação da investigação estatística desenvolvida em Portugal, reforçando a sua presença internacional. O sucesso desta iniciativa deixa o desejo de repetir esta colaboração em futuras edições do congresso.

Como curiosidade, durante uma das pausas para café do congresso, na mesa da Cambridge University Press, destacamos o livro “*Computational Bayesian Analysis: An Introduction*”, da autoria do falecido Professor Carlos Daniel Paulino, antigo presidente da SPE, da Professora Maria Antónia Turkman e do Professor Peter Müller, numa forma de recordar a contribuição destes nossos colegas à estatística em Portugal.

Vanda Lourenço

• Dia Internacional de Literacia Estatística 2025



Mais informações em: <https://tinyurl.com/IDSL-2025>

Pedro Campos

Enigmística de mefqa

A M o S T R A G E M

A A Ã
C C Ç
F
I I
L
O
S S

Enigmas 49 e 50

No Boletim SPE primavera de 2025 (p. 12):

Eu ☐ latino
Tu ☐ latinas
Ele ☐ latina
Nós ☐ latinamos
Vós ☐ latinais
Eles ☐ latinam

distribuição

Enigma 47: Quadrados latinos conjugados

Enigma 48: Distribuição de caudas pesadas

Preâmbulo

Instituto Nacional de Estatística, I.P.

No passado dia 23 do mês de maio, o Instituto Nacional de Estatística (INE) comemorou o seu 90º aniversário.

À semelhança do que aconteceu com a comemoração dos 85 anos do INE, a Sociedade Portuguesa de Estatística (SPE) convidou gentilmente o INE para que a comemoração dos 90 anos fosse a temática do seu Boletim semestral, na edição de outono, convite que o INE agradece com muito gosto.

Para a comemoração dos 90 anos, foi criado um logotipo alusivo a este momento da vida do INE:



É uma representação dos Valores e da Missão do INE, sendo o centro da identidade visual da instituição durante este ano de celebração. O slogan que o acompanha: **“90 anos de rigor e inovação ao serviço da Sociedade”** resume a trajetória do INE e reforça a mensagem que desejamos transmitir.

A atividade do INE bem como das entidades com delegação de competências, DREM e SREA pertencentes ao Sistema Estatístico Nacional (SEN), para além das entidades ao nível dos restantes países da União Europeia (UE) pertencentes ao Sistema Estatístico Europeu (SEE), seguem os princípios do Código de Conduta para as Estatísticas Europeias (CCEE); trata-se de uma garantia associada à produção e difusão de estatísticas de qualidade.

Para esta edição comemorativa dos 90 anos, o INE escolheu um conjunto de temas estruturantes da sua atividade, que se apresentam em seis contributos cujos resumos são apresentados de seguida:

1. Como referido no slogan associado a esta comemoração, o primeiro tema trata da inovação; ao longo do tempo, o INE tem procurado inovar sempre que possível, seja com utilização de novas ferramentas, metodologias inovadoras, ou mesmo reestruturação de processos de trabalho, que acrescentem ganhos em eficácia e eficiência, melhoria da qualidade dos resultados e redução de prazos de disponibilidade dos resultados. Neste contributo são apresentados 20 casos significativos que foram iniciados nos últimos 20 anos e que têm tido continuidade. No que se refere ao CCEE,

os casos apresentados estão relacionados com os princípios do Compromisso da Qualidade, Metodologia Sólida, Procedimentos Estatísticos Apropriados, Carga Não Excessiva Sobre Respondentes, e Eficácia Na Utilização De Recursos.

2. Desde 2016, o INE tem vindo a explorar fontes de *Big Data* e as técnicas associadas, nomeadamente na área da aprendizagem supervisionada. Com o advento dos *Large Language Models* (modelos de linguagem de grande escala capazes de compreender e gerar texto de forma semelhante à humana, aprendendo a partir de enormes volumes de dados), o cenário mudou profundamente. Torna-se, assim, necessário equacionar as oportunidades que estas novas técnicas de inteligência artificial podem trazer às estatísticas oficiais, bem como os desafios e riscos que inevitavelmente lhes estão associados.
3. A maioria da atividade do INE é desenvolvida para dar resposta às necessidades solicitadas ao nível do Sistema Estatístico Europeu, maioritariamente através de regulamentos próprios. Ao longo do tempo foi sendo necessário efetuar revisões ao Regulamento (CE) n.º 223/2009, regulamento das Estatísticas Europeias, de modo a adequá-lo às questões emergentes.
4. O macroprocesso chave da atividade do INE é a produção estatística. O Manual do Processo da Produção Estatística (MPPE) do INE é o documento de suporte para o desenvolvimento das diversas tarefas que se desenvolvem neste processo, as quais têm como referência o modelo internacional amplamente utilizado pelos institutos de estatística - o GSBPM – Generic Statistics Business Process Model da UNECE.
5. O projeto ALEA, Ação Local de Estatística Aplicada (www.alea.pt), fez 25 anos e foi criado pelo INE, a Escola Secundária Tomaz Pelayo e o Ministério da Educação (Direção-Geral dos Estabelecimentos Escolares), com o propósito de proporcionar instrumentos relacionados com o ensino, a aprendizagem e, por consequência, a compreensão e a utilização da Estatística, destinados essencialmente a docentes e alunos dos ensinos básico e secundário, mas dos quais também podem beneficiar outros públicos. Neste boletim damos conta de várias novidades, nomeadamente da disponibilização de conteúdos relacionados com a Educação para a Cidadania.
6. A capacitação em ciência de dados para o Sistema Estatístico nacional e Administração Pública é uma iniciativa desenvolvido com apoio do PRR, promovida pelo INE com vista a melhorar a transformação digital e científica da produção estatística, apoiando uma Administração mais atual e informada, ágil e preparada com os recursos necessários para os desafios da sociedade. Trata-se de uma iniciativa que no âmbito do CCEE, vem dar resposta nomeadamente ao Princípio de Metodologia Sólida.

O INE espera que estas informações tenham contribuído para um melhor conhecimento da sua atividade e melhor entendimento da sua missão.

A Inovação no INE

20 exemplos dos últimos 20 anos

Instituto Nacional de Estatística, I.P.

1. Introdução

Num ecossistema de dados em rápida mudança, é fundamental aumentar a capacidade de resposta do Instituto Nacional de Estatística (INE) à crescente procura por estatísticas mais relevantes, oportunas e detalhadas. A inovação assume um papel central, tornando-se transversal à organização, abrangendo processos, metodologias, produtos, serviços e a própria cultura de trabalho.

Esta abordagem alinha-se diretamente com os objetivos da Rede Europeia de Inovação (EIN), que visam dar resposta às necessidades de novas estatísticas, aumentar a eficiência na produção estatística, e adaptar-se a mudanças nas normas sociais e a choques inesperados – a necessidade de novos dados em tempo útil durante a pandemia COVID 19 é um exemplo claro dessa urgência. Em última análise, pretende-se melhorar a capacidade do INE reagir de forma rápida e flexível.

A crescente disponibilidade de dados e a velocidade de transmissão de informação traz igualmente outros desafios, nomeadamente a uma disseminação generalizada de desinformação. É uma oportunidade para que o INE reforce o seu papel como fonte de informação credível e robusta, fazendo das estatísticas oficiais um pilar da democracia.

As iniciativas desenvolvidas pelo INE nestes últimos anos, de entre as quais se destaca a Infraestrutura Nacional de Dados (IND), são a prova deste compromisso estratégico e multidimensional. A inovação, neste contexto, não se restringe a uma modernização tecnológica, mas sim à promoção de uma abordagem holística que integra novas abordagens, em diferentes dimensões. Os pontos seguintes aprofundam esta estratégia, apresentando alguns exemplos de como o INE tem concretizado este novo paradigma de inovação, tanto a nível institucional como prático.

Eurostat, 2023, Innovation Agenda Network, <https://cros.ec.europa.eu/landing-page/ess-innovation-agenda-network-ein-home-page>

2. A Infraestrutura Nacional de Dados: Um Objetivo Estratégico

A necessidade de dar resposta a um sistema em mudança, aliada ao aumento do volume e complexidade de novas fontes de dados – em especial dos dados administrativos - traz desafios importantes para o processo de produção de estatísticas oficiais.

O INE tem estado ativamente empenhado numa estratégia de inovação, desenvolvendo e reforçando não somente as suas capacidades de gestão e de análise de dados, mas também várias outras dimensões da sua atividade.

A Infraestrutura Nacional de Dados (IND) incorpora várias estratégias e iniciativas que o INE implementou nos últimos anos. Pode ser definida como um sistema desenvolvido pelo INE para integrar e utilizar dados administrativos e de outras fontes para a produção de estatísticas oficiais, proporcionando um único ponto de acesso para vários utilizadores e finalidades. O sistema garante a proteção e a integridade dos dados ao longo de toda a cadeia de produção, desde a recolha de dados até à validação e análise, permitindo tirar partido de toda a robusta cadeia produtiva implementada pelo INE.

A nível organizacional, o INE ajustou a sua organização interna, renovou o Manual do Processo de Produção Estatística e criou uma Unidade de Dados Administrativos dedicada. No plano tecnológico e metodológico, houve investimento significativo em infraestrutura de tecnologias de informação, ferramentas de análise de dados e formação de colaboradores para reforçar a exploração e análise de dados utilizando inteligência artificial e *machine learning*. Ao nível da difusão e comunicação, foram implementadas novas estratégias para adquirir dados diretamente das fontes, incluindo o envolvimento direto da gestão de topo e equipas multidisciplinares e tem apostado na disponibilização crescente de dados para fins de investigação. Também ao nível de recursos humanos, o INE tem feito investimentos significativos para reforçar as competências dos seus colaboradores no tratamento e análise de grandes conjuntos de dados administrativos, num espírito colaborativo e abrangente.

Estas iniciativas são relevantes por si só, mas são muito mais impactantes quando trabalhadas em conjunto; foram concebidas para criar um ecossistema de dados abrangente e integrado.

Nos últimos anos, são muitos e variados os exemplos de aspetos inovadores em que o INE tem estado envolvido. Estes estendem-se desde projetos estruturantes, como a aposta na integração de dados para o desenvolvimento do Recenseamento da População e Habitação com base em dados administrativos; a produção de novas estatísticas e projetos com base em dados de fontes externas, como sejam o Índice de Preços de Habitação ou a Carta de Equipamentos e Serviços de Interesse Geral; até iniciativas que visam o desenvolvimento da cultura organizacional.

Não obstante, a estratégia de inovação do INE abrange também projetos menos recentes que foram e continuam a ser profundamente transformadores, tanto ao nível da organização como do aproveitamento de dados administrativos. Para ilustrar o conceito abrangente de inovação para este organismo, foram selecionados, para este artigo, exemplos de diferentes áreas que o INE tem procurado, de forma contínua, antecipar e adaptar às transformações da sociedade e do ecossistema de dados.

Para ilustrar este conceito abrangente e demonstrar o percurso evolutivo do INE, foram selecionados 20 exemplos de inovação que o INE tem promovido ativamente nos últimos 20 anos – um período de profunda transformação no ecossistema estatístico, e que são descritos no capítulo seguinte. Estes projetos abrangem diferentes áreas de atuação e serão apresentados e organizados em torno de quatro pilares essenciais: reengenharia organizacional e de processos; cultura colaborativa; tecnologias e método; fontes externas e produtos; refletindo a diversidade e a natureza holística da estratégia de inovação do INE.

INE, 2023, “How to keep timeliness and frequency with monthly administrative data?”
https://unece.org/sites/default/files/2023-06/CES%202023%2031_Item%209.pdf

Francisco Lima, 2020, “The National Data Infrastructure at Statistics Portugal: Recent developments and international examples”,
https://cse.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=469169601&att_display=n&att_download=y

Francisco Lima; José Pinto Martins; Katja Neugebauer; Joana F. Pimentel; Sofia Rodrigues; Paulo Saraiva; Rita Sousa; Maria João Zilhão, 2025 (expected), “Data-Driven Learning in the EU – Administrative Data in Portugal

3. A inovação na prática: 20 exemplos dos últimos 20 anos

➤ O Modelo Centralizado do INE para Gestão de Dados: Eficiência e Reforço da Capacidade Analítica

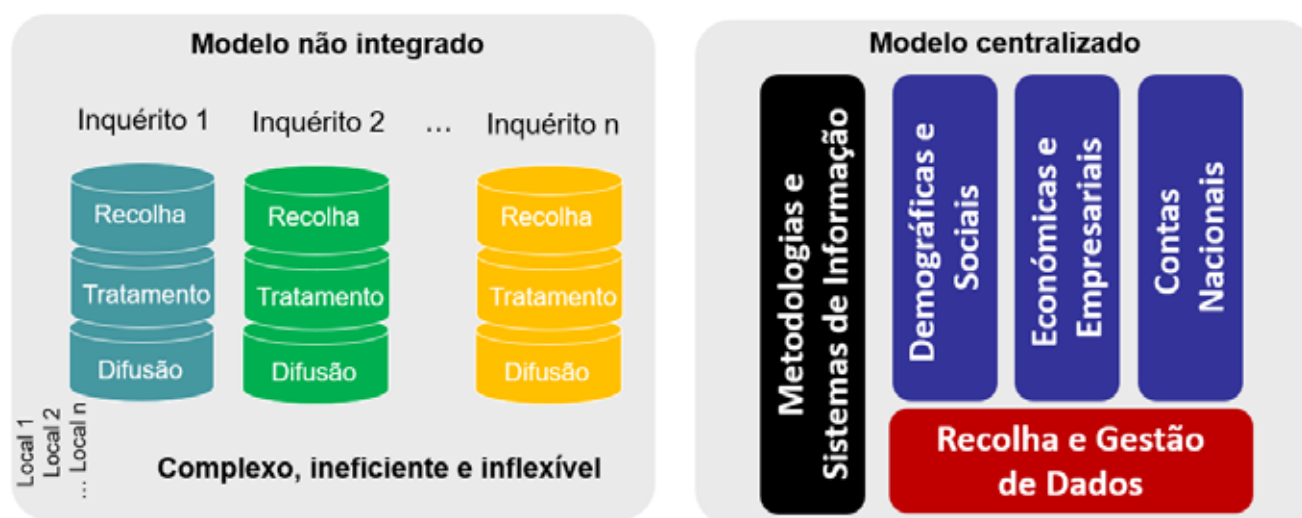
O Instituto Nacional de Estatística (INE) segue um modelo centralizado para a recolha e gestão de dados. Todos os microdados — sejam provenientes de inquéritos, censos, registos administrativos ou outras fontes — são geridos sob uma estrutura unificada, com supervisão central, recursos coordenados e estratégias partilhadas para a inovação.

Apesar desta forte coordenação central, as operações estão distribuídas por cinco locais no Continente, bem como nas Regiões Autónomas dos Açores e da Madeira. Estes serviços regionais colaboram estreitamente com o INE nas operações nacionais, utilizando as mesmas infraestruturas, processos e padrões.

A unidade orgânica responsável pela gestão de dados reporta diretamente ao Conselho Diretivo e desempenha um papel estratégico na organização. Compreende três áreas principais: inquéritos às famílias, inquéritos às empresas e dados administrativos. Esta unidade é responsável por cerca de 30% dos recursos humanos do INE e 40% do seu orçamento, sublinhando a sua importância operacional e estratégica. As suas atividades estão estreitamente alinhadas com várias fases do Manual do Processo de Produção Estatística (GSBPM), especificamente: Especificações, Recolha, Processamento e Análise.

Esta abordagem decorreu de um processo evolutivo fortemente alavancado na criação do Departamento de Recolha de Informação (DRI) em 2004, inserido na profunda reestruturação orgânica então efetuada no INE. Tal como em muitos outros institutos nacionais de estatística, a anterior arquitetura do INE assentava num sistema não integrado, baseada em numerosos processos paralelos, domínio por domínio, seguindo uma abordagem tradicional de “silo”. Esta forma de produção de estatísticas foi, então, considerada ineficiente e pouco flexível.

Após um processo de reflexão e reorganização em 2004, foi empreendido um projeto de reengenharia da arquitetura de produção, baseado numa abordagem integrada e orientada por processos, visando melhorar a sua eficiência e flexibilidade. Consequentemente, foi criado um departamento central de recolha de dados, as direções regionais foram extintas, e os departamentos de produção por domínio foram fundidos em três unidades: economia, social e contas nacionais. Os métodos e os sistemas de informação foram fundidos num só departamento.



Foi um desafio notável, tendo em conta a nova distribuição de recursos, funções e responsabilidades. Apesar de alguma resistência e outros constrangimentos, a transição foi realizada de modo que as operações estatísticas não fossem substancialmente afetadas.

Com a criação do DRI, numa primeira fase, procedeu-se à concentração da recolha num Departamento especializado, mantendo-se, no entanto, a infraestrutura tecnológica e aplicacional existente. Numa segunda fase, deu-se início à reengenharia da arquitetura do processo produtivo, tendo sido criado um sistema integrado, o SIGINQ – Sistema Global de Gestão de Inquéritos. Este sistema foi implementado em 2008 como suporte da recolha do INTRASTAT e foi sucessivamente alargado e atualizado para permitir a gestão de outros inquéritos, nomeadamente inquéritos agrícolas e inquéritos sociais (às famílias e indivíduos). Adicionalmente, foi evoluindo para, além da recolha, incluir outras fases do processo produtivo, nomeadamente o tratamento e análise dos dados.

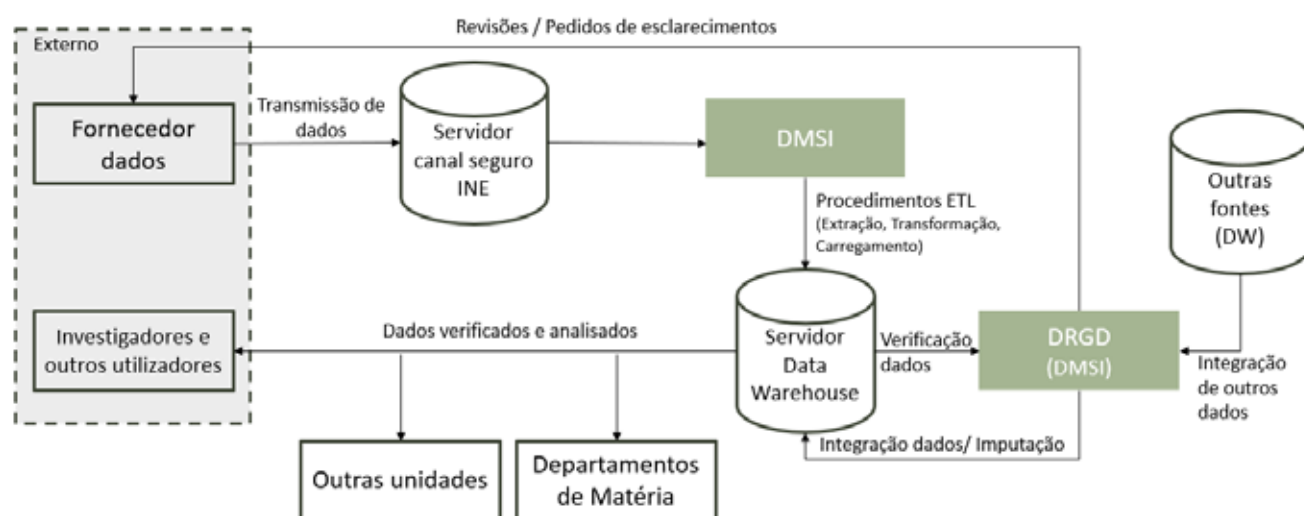
O trabalho desenvolvido levou ao abandono da anterior abordagem segmentada por operação estatística (e dos anteriores sistemas informáticos específicos por operação estatística) e à adoção de uma abordagem integrada, aproveitando sinergias existentes e permitindo uma significativa economia de

escala. Contribuiu também significativamente para a definição do MPPE (Manual do Processo de Produção Estatística) que reúne as normas internas referentes ao processo de produção estatística e está completamente alinhado com as normas internacionais definidas no GSBPM (GSBPM – *Generic Statistical Business Process Model*).

À luz das inovações previstas no âmbito da Infraestrutura Nacional de Dados, e de modo a aumentar a capacidade analítica e a flexibilidade na adaptação da informação às necessidades de tomada de decisão, no contexto da utilização de múltiplas fontes de dados, o INE implementou, em 2019, novos ajustamentos na sua organização interna, nomeadamente nas competências do Departamento de Metodologias e Sistemas de Informação e do Departamento de Recolha e Gestão de Dados (anteriormente designado por DRI), reforçando os seus papéis nas capacidades de gestão e análise de dados.

Em 2020, foi criada uma unidade dedicada ao tratamento e análise de dados administrativos – o Núcleo de Dados Administrativos, no âmbito do Departamento de Recolha e Gestão de Dados. Esta unidade, em colaboração com o Departamento de Metodologia e Sistemas de Informação, é responsável, por proceder a tratamentos de limpeza, classificação, melhoria e processamento central de dados administrativos, associando a respetiva metainformação, previamente à disponibilização dos conjuntos de dados aos diferentes utilizadores, quer para a produção de estatísticas oficiais quer para fins de investigação. São criadas bases integradas e consolidadas, que tenham em conta as diferentes necessidades dos vários utilizadores – também neste caso se pode falar da aplicação do princípio Only Once.

Processamento Dados Administrativos



Refere-se ainda que o "Manual do Processo de Produção Estatística" tem sido atualizado em conformidade, sendo que nas últimas versões, foi introduzido o projeto da Infraestrutura Nacional de Dados nos processos estatísticos através da sua referência em tarefas específicas e foi dada maior ênfase às atividades de apropriação, validação, tratamento, análise e integração de microdados provenientes de fontes alternativas (diferentes de inquéritos).

A consolidação desta abordagem centralizada, que se apoia na gestão eficiente de microdados e na utilização crescente de fontes administrativas, estabeleceu uma vantagem fundamental para o INE. Contudo, a evolução dos dados é constante. Este modelo organizacional, com o seu foco contínuo na integração e na melhoria dinâmicas dos processos, não se limita a reagir às transformações, mas sim a acompanhar e antecipar as tendências do ecossistema de dados, assegurando que a capacidade analítica do INE permanece dinâmica, flexível e apta a responder às futuras necessidades de informação para a tomada de decisão.

➤ Webinq: 20 anos de um sistema inovador

Criado em 2005, o WebInq surgiu com o objetivo de facilitar o preenchimento dos inquéritos, reduzir o esforço dos respondentes e criar processos mais ágeis e eficientes, contribuindo para aumentar a qualidade dos dados recolhidos. É um serviço online, dedicado à recolha eletrónica de dados, que simplifica a resposta a inquéritos oficiais por parte de empresas, indivíduos, famílias e outras entidades, promovendo um relacionamento mais próximo e eficiente com os respondentes. O WebInq tem sido essencial na desmaterialização dos questionários (em formato papel), com mais de 100 inquéritos dirigidos às empresas. Esta aposta resultou numa proporção de resposta online de 98,8% em 2024, para os dados empresariais.



Com o objetivo de otimizar a experiência dos utilizadores no WebInq, o INE tem vindo a introduzir novas funcionalidades. Um marco relevante foi a disponibilização, em 2013, da Transmissão Automática de Dados (TAD). Esta funcionalidade permite aos respondentes usar os dados dos seus próprios sistemas de informação e gerar um ficheiro XML para upload no WebInq ou para envio direto para o INE através de Webservice, sendo uma alternativa mais eficiente à introdução manual das respostas nos formulários eletrónicos.

Outra inovação crucial foi a introdução do Retorno de Informação ao Respondente (RIR) em 2014. Esta funcionalidade, que oferece às empresas relatórios com informação de natureza genérica e personalizada, disponibilizados no WebInq, foi reconhecida como uma prática inovadora num exercício de *Peer Review* ao Sistema Estatístico Nacional em 2022 – avaliação inter-pares ao nível europeu de avaliação ao Sistema Estatístico Nacional sobre a implementação do Código de Conduta para as Estatísticas Europeias.

Um exemplo da sua aplicação pode ser observado no ponto dedicado ao Inquérito Rápido e Excecional às Empresas (IREE).

No ano em que celebra o seu vigésimo aniversário, o WebInq renovou a sua imagem e algumas das suas funcionalidades. Tem um novo *layout*, cores mais contrastantes e está em conformidade com os critérios de acessibilidade e usabilidade definidos na legislação portuguesa. Foram adicionadas novas funcionalidades, como a autenticação para cidadãos estrangeiros através da Chave Móvel Digital e a

introdução da Autenticação da Autoridade Tributária, que reforça a segurança e acelera a adesão das empresas ao sistema. A funcionalidade de Retorno de Informação ao Respondente (RIR) foi expandida, disponibilizando relatórios a todos os utilizadores, principais e secundários. Uma nova funcionalidade de Indicadores do WebInq também foi disponibilizada, permitindo a consulta de dados como o número de aderentes, respostas recebidas e visitas.

O WebInq tem vindo a ser apresentado e referenciado como um dos grandes impulsionadores do processo de modernização ocorrido no INE, em particular na recolha por autopreenchimento. Este sistema continua a ser apresentado em vários eventos internacionais e o INE tem sido convidado a receber visitas de estudo de outros países, demonstrando o interesse e o reconhecimento pela inovação do WebInq.

INE, 2025, Destaque “WebInq - 20 Anos a simplificar as respostas”,
https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=737815219&DESTAQUESmodo=2

➤ A Criação do Inquérito Rápido e Excecional às Empresas em Tempo de COVID

O impacto da pandemia COVID-19 estendeu-se a todos os setores da economia. Para identificar alguns dos seus principais efeitos na atividade empresarial, o Instituto Nacional de Estatística (INE) e o Banco de Portugal (BdP) lançaram o Inquérito Rápido e Excecional às Empresas (COVID-IREE).

Compreendendo a difícil situação que as empresas atravessavam, as duas instituições tomaram a iniciativa de lançar um questionário simples, com seguimento semanal, de modo a monitorizar a evolução económica num futuro próximo, assim que possível. O conhecimento da realidade empresarial

nessa altura era de particular importância para avaliar as consequências da pandemia atual (COVID-19) nos diversos setores de atividade.

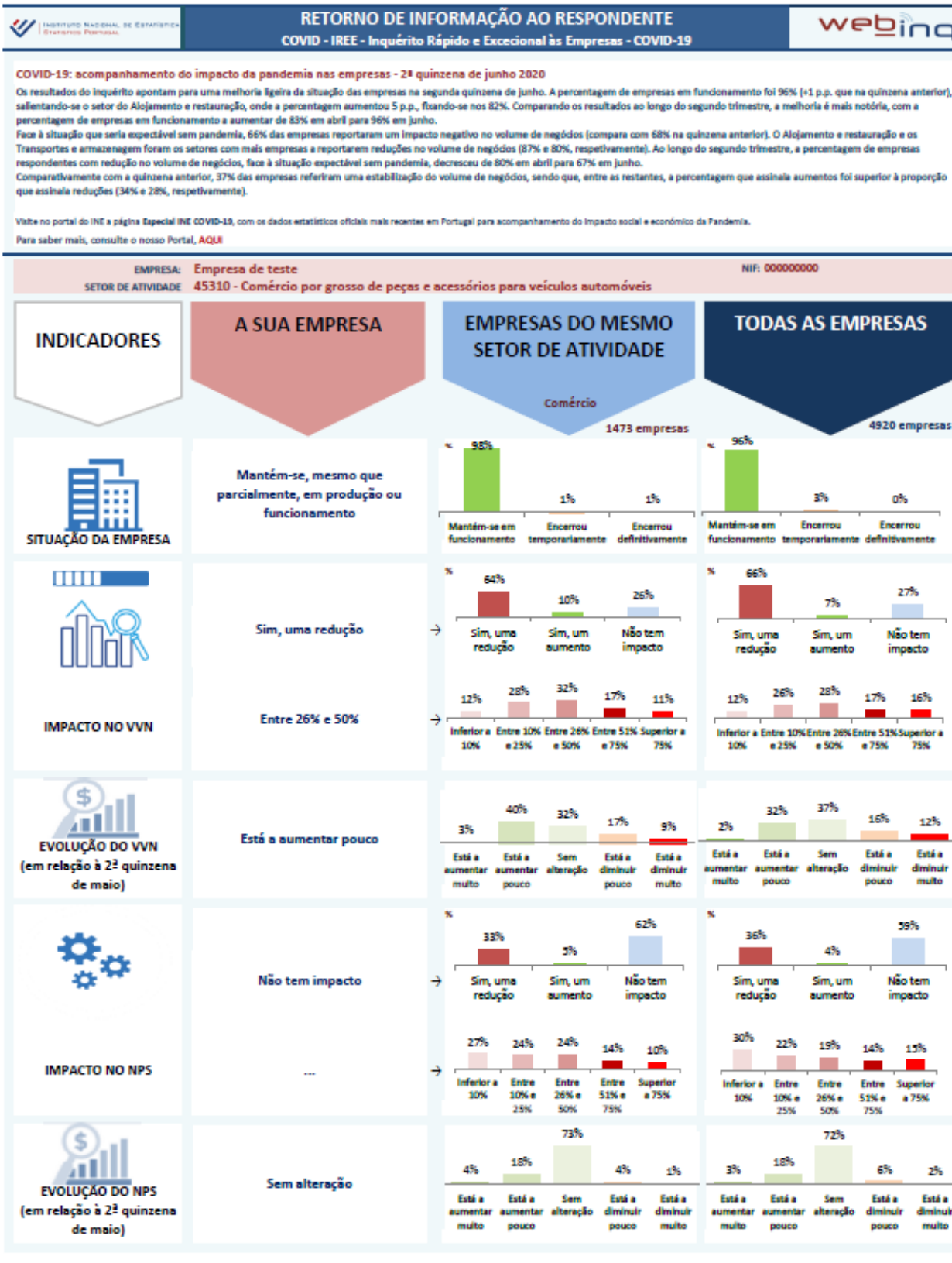
A flexibilidade e integração associadas ao Sistema Integrado de Gestão de Inquéritos, que é um componente essencial da infraestrutura de produção do INE, permitiu uma resposta rápida e flexível a diferentes tipos de necessidades.

Neste inquérito, foi possível satisfazer a necessidade identificada de conhecer a situação das empresas durante este período excecional em menos de cinco dias, desde a conceção do questionário ao início da recolha.

Apesar de o INE possuir um sistema de recolha de dados maduro e flexível, esta foi a primeira vez que tentou recolher dados de um grande número de unidades numa base semanal, o que gerou desafios significativos. O questionário foi cuidadosamente concebido para ter uma resposta simples e rápida, através de caixas de seleção, garantindo que o tempo de preenchimento nunca excedesse os cinco minutos. A amostra incluiu cerca de 8.000 empresas (micro a grandes), aproveitando a robustez e representatividade da amostra dos inquéritos mensais do Volume de Negócios. Foi crucial a comunicação com os inquiridos, onde se enfatizou que a participação era um apelo essencial para a definição de políticas públicas. Para assegurar as boas taxas de resposta no apertado prazo de cinco dias, foram afetados recursos humanos com dedicação exclusiva, e o sistema de lembretes foi reforçado, incluindo contactos mais personalizados. Além disso, a partir da segunda edição, o envio do relatório de feedback específico logo no dia 2 da recolha serviu também como um lembrete eficaz.

Houve 11 edições deste inquérito: quatro edições semanais (entre 6 de abril e 1 de maio de 2020), cinco edições quinzenais (entre a primeira quinzena de maio e a primeira quinzena de julho de 2020) e duas edições ad-hoc (uma em novembro de 2020 e outra em fevereiro de 2021). O inquérito manteve-se ativo enquanto se justificou e a sua frequência foi ajustada em função da evolução das restrições à atividade económica.

Desde 2013, o INE tem melhorado a relação com as empresas ao oferecer relatórios de feedback regulares e personalizados, visando aumentar a motivação e reforçar a utilidade da informação estatística. Para o COVID-IREE, apesar do tempo limitado, foi feito um esforço para criar um relatório de feedback personalizado que acompanhava a frequência do inquérito. Este relatório era enviado a cada empresa respondente na semana seguinte à recolha, contendo a sua resposta e o enquadramento na sua atividade económica. Além de cumprir o objetivo de dar algo em troca, o envio do relatório personalizado serviu como um lembrete eficaz e uma fonte tangível de motivação para as empresas continuarem a participar.



Este inquérito pode ser considerado uma prova do sucesso do sistema implementado no INE. Apesar do tempo muito limitado disponível, foi possível conduzir um inquérito de alta frequência, a um número significativo de unidades e foi bem aceite pelos inquiridos. O foco numa forte relação com os inquiridos que o INE tem vindo a desenvolver deu frutos, como se pode constatar nos bons resultados obtidos.

As principais dificuldades inerentes a um inquérito com estas características, nomeadamente ao nível da estabilização da primeira versão do inquérito, os prazos muito apertados e a necessidade de afetar recursos de forma quase imediata e exclusiva, foram possíveis de ultrapassar graças a um trabalho árduo e à importância dada a todos os detalhes desta operação estatística.

➤ Disponibilização de diferentes modos de recolha nos inquéritos às famílias

Os inquéritos às famílias, realizados em contexto de entrevista, foram durante muitos anos exclusivamente assegurados no modo de recolha presencial. Após alguns anos de estudo, conhecimento de experiências de outros países e vários testes, foi disponibilizado o modo de recolha telefónica. Com o progressivo alargamento deste modo a diversos inquéritos, o INE adquiriu uma solução de Centro de Contactos que foi sendo adaptada, com o apoio do DMSI, ao contexto das estatísticas oficiais. Essa customização resulta da necessidade de gestão das amostras e *case management* para promoção das taxas de resposta nos inquéritos às famílias, bem como da integração de procedimentos de acompanhamento e monitorização com vista a garantir a qualidade da recolha.

A recolha telefónica no INE faz 20 anos em novembro de 2025 e evoluiu de um contexto interno (primeiro com colaboradores da organização e depois em conjunto com uma equipa contratada para trabalhar fisicamente no INE em Lisboa) para uma abordagem inovadora ao nível da gestão de equipas de Entrevistadores. Em outubro de 2011, aproveitando a existência de uma equipa de Entrevistadores com experiência na realização de inquéritos em CAPI (*Computer Assisted Personal Interviewing*), o CATI (*Computer Assisted Telephone Interviewing*) passou a ser assegurado também por esta equipa. Esta nova abordagem designou-se por HomeCATI e consiste na realização de entrevistas telefónicas a partir de casa com infraestrutura do INE e ligação ao Sistema Integrado de Centro de Contactos e ao SIGINQ, garantindo os procedimentos e monitorização da qualidade já implementados, a par do formato de entrevista em guião idêntico nos dois modos que promove a devida harmonização na realização das entrevistas.

A estratégia de recolha em modos mistos foi em 2014 alargada ao CAWI (*Computer Assisted Web Interviewing*), sendo esta uma aposta do INE para permitir aos respondentes optarem pelo modo de resposta preferencial, a par da otimização das funcionalidades de transição de alojamentos entre modos, de forma a promover as taxas de resposta e a gestão das equipas de Entrevistadores.

Para além das iniciativas de inovação referidas, importa assinalar que o referido sistema integrado de gestão de inquéritos às famílias em modos mistos, se revelou fundamental durante a pandemia do COVID-19. Permitiu ao INE assegurar a continuidade da recolha de todos os inquéritos às famílias previstos no Plano de Atividades durante a fase de suspensão da recolha presencial.

➤ Sistema de Gestão Integrado

Nos últimos cinco anos, o INE fez progressos significativos na formalização do seu Sistema de Gestão Integrado (SGI), tal como reconhecido no exercício de *Peer Review* efetuado em 2022.

O SGI integra o Sistema de Gestão da Qualidade e o Sistema de Gestão de Segurança da Informação, e constitui uma infraestrutura organizacional orientada para o desempenho institucional sustentável. Tem como objetivo ter um Sistema transversal que assegure a conformidade técnica, a robustez metodológica e a confiança institucional no processo de produção e divulgação das estatísticas oficiais do INE, num quadro de melhoria contínua e de resposta adaptativa às exigências emergentes.

Este sistema tem como referenciais:

- ISO 9001 para o Sistema de Gestão da Qualidade.
- ISO/IEC 27001 para o Sistema de Gestão de Segurança da Informação e a extensão ISO/IEC 27701 para a Gestão de Informação Privada.

Estas normas definem os requisitos e controlos para estabelecer, implementar, manter e melhorar de forma contínua o SGI. O SGI implementa os princípios orientadores destas normas a todos os processos, em integração com a estrutura de gestão global.

Do ponto de vista operacional, a estrutura do SGI baseia-se no Generic Statistical Business Process Model (GSBPM), desenvolvido pela UNECE. Este modelo, de referência internacional, foi implementado no Manual do Processo de Produção Estatística (MPPE) do INE. Cobre de forma sistemática as oito fases do ciclo de vida da produção estatística — identificação de necessidades, especificação, desenvolvimento, recolha, processamento, análise, divulgação e avaliação — permitindo a rastreabilidade do processo, a padronização documental e uma avaliação sistemática da qualidade e da eficiência operacional.

O SGI está sustentado por um quadro de certificações externas, que reconhecem a maturidade e eficácia dos sistemas de gestão implementados. Desde 2019, o INE mantém a certificação da sua componente de Gestão da Segurança da Informação, ao abrigo da ISO/IEC 27001, no contexto do ESS IT Security Framework, inicialmente aplicada às estatísticas de comércio internacional intra-EU, e posteriormente alargada ao domínio extra-EU. Em 2024, foi formalmente instituído o Sistema de Gestão Integrado (SGI) do INE, consolidando num único sistema os domínios da gestão da qualidade e da segurança da informação. Esta integração permitiu alinhar de forma transversal os processos da organização com os princípios da melhoria contínua, gestão por processos e controlo do risco, reforçando a coerência e eficácia da arquitetura de gestão. No mesmo ano, o SGI foi objeto de certificação externa, com o INE a obter, além da manutenção da certificação da ISO/IEC 27001 (segurança da informação), a certificação do Sistema de Gestão da Qualidade segundo a norma ISO 9001:2015, abrangendo a totalidade do processo de produção e divulgação de estatísticas oficiais do INE. Este reconhecimento formal reforça o posicionamento do INE no Sistema Estatístico Europeu, evidenciando a adoção consistente de boas práticas organizacionais e a credibilidade das suas estatísticas junto das partes interessadas.

De salientar que, no domínio específico da segurança da informação, o INE adotou uma abordagem estruturada, baseada na gestão de risco e orientada pelos princípios da confidencialidade, integridade e disponibilidade da informação. Este domínio integra de forma articulada a Segurança da Informação, a Cibersegurança e a Proteção da Privacidade, abrangendo igualmente a gestão da informação pessoal em conformidade com os requisitos da norma ISO/IEC 27701:2019, extensão da ISO/IEC 27001 para Sistemas de Gestão da Informação de Privacidade. O sistema abrange controlos técnicos, organizacionais e procedimentais, apoiado por políticas e procedimentos, auditorias regulares e mecanismos de resposta a incidentes. Esta abordagem permite ao INE operar com elevados padrões de resiliência digital, especialmente críticos num contexto de crescente integração de fontes externas e utilização intensiva de dados administrativos.

Importa salientar que a metodologia de gestão de risco, inicialmente aplicada ao domínio da segurança da informação, foi progressivamente incorporada no Sistema de Gestão Integrado (SGI), sendo atualmente transversal a outras componentes, como por exemplo a gestão da qualidade e a proteção de dados pessoais. Esta abordagem holística contribui para uma gestão mais integrada e proativa, potenciando a mitigação de riscos operacionais e o cumprimento dos requisitos legais e normativos aplicáveis.

Neste quadro, o SGI do INE funciona não apenas como um sistema de conformidade, mas como um mecanismo dinâmico de inovação organizacional, orientado para a antecipação de necessidades, reforço da confiança dos utilizadores e incorporação de novas abordagens metodológicas e tecnológicas. A sua operacionalização sustenta-se num compromisso explícito com os princípios do Código de Conduta para as Estatísticas Europeias, assegurando a independência profissional, a imparcialidade e a transparência, enquanto sustenta a adoção de abordagens metodológicas inovadoras na produção estatística e de instrumentos de governança integrados.

[INE, \[SGI - Sistema de Gestão Integrado\]](#)

➤ Colaboração entre o INE e a academia

A colaboração com a academia é reconhecida como uma das componentes estratégicas que o INE tem adotado para responder eficazmente aos novos desafios impostos pelo ecossistema de dados atual. Esta abordagem foi, aliás, identificada como uma prática inovadora pela equipa do *Peer Review* efetuado em 2022, que destacou o recurso a investigadores residentes no INE.

O recrutamento de investigadores especializados, dotados de competências especializadas, contribui significativamente para o processamento e a análise de bases de dados complexas, em estreita colaboração com os técnicos do INE. Exemplos específicos destacam as colaborações em diversos projetos, tais como o Inquérito às Condições de Vida e Rendimento (ICOR) e os Censos Administrativos. A mentoria emerge como um aspeto crítico, facilitando o desenvolvimento de competências e a construção de capacidade no INE. Este papel de mentoria estende-se para além do apoio individual, abrangendo contribuições centradas em projetos, garantindo a transferência contínua de conhecimento.

O INE aposta igualmente na construção de parcerias contínuas com a academia. Um exemplo é a submissão, em 2022, do projeto '*Inequality, individual behaviour, and policy reforms: evidence from administrative tax records*' à Fundação para a Ciência e a Tecnologia (FCT), que promove a investigação sobre o uso de dados administrativos. Adicionalmente, o INE é membro da rede EMOS (*European Master in Official Statistics*) em colaboração com a Faculdade de Economia da Universidade do Porto. Esta parceria visa avançar na educação de pós-graduação em estatística oficial e *data science* a nível europeu, facilitando o intercâmbio de conhecimento e a formação de estatísticos altamente qualificados.

As iniciativas de formação desempenham um papel crucial no desenvolvimento de competências internas. O INE tem investido significativamente em cursos de formação especializada oferecidos em colaboração com universidades de prestígio, destacando-se programas em 'Inteligência Artificial para Empresas' e 'Macroeconometria'. Mais recentemente, em 2023, o INE iniciou o programa de Formação em Data Science para o Sistema Estatístico Nacional e a Administração Pública, com um volume superior a 6,5 mil horas de formação, abrangendo competências críticas como *Machine Learning*, *Deep Learning* e *Python* para Estatística Oficial.

O acesso à microinformação oficial para investigação é também um pilar fundamental na promoção de um ecossistema de dados abrangente e interligado. O INE, reconhecendo as necessidades da comunidade académica, estabeleceu procedimentos rigorosos para facilitar o acesso de investigadores à informação estatística para projetos e teses. Atualmente, o INE disponibiliza 72 microbases de dados para fins científicos, utilizando diferentes modos de acesso que incluem ficheiros totalmente anonimizados, tabelas pré-preparadas ou o ambiente seguro do *Safe Centre*.

Em suma, esta secção demonstra a abordagem holística do INE à modernização, onde a colaboração com a academia, as estratégias de recrutamento, os programas de mentoria, as iniciativas de formação e o incentivo ao acesso a dados se conjugam.

2025; [Collaboration with academia - Statistics Portugal | Eurostat CROS](#)

2024; Rodrigues, S. e al, [Collaboration with academia in Statistics Portugal](#)

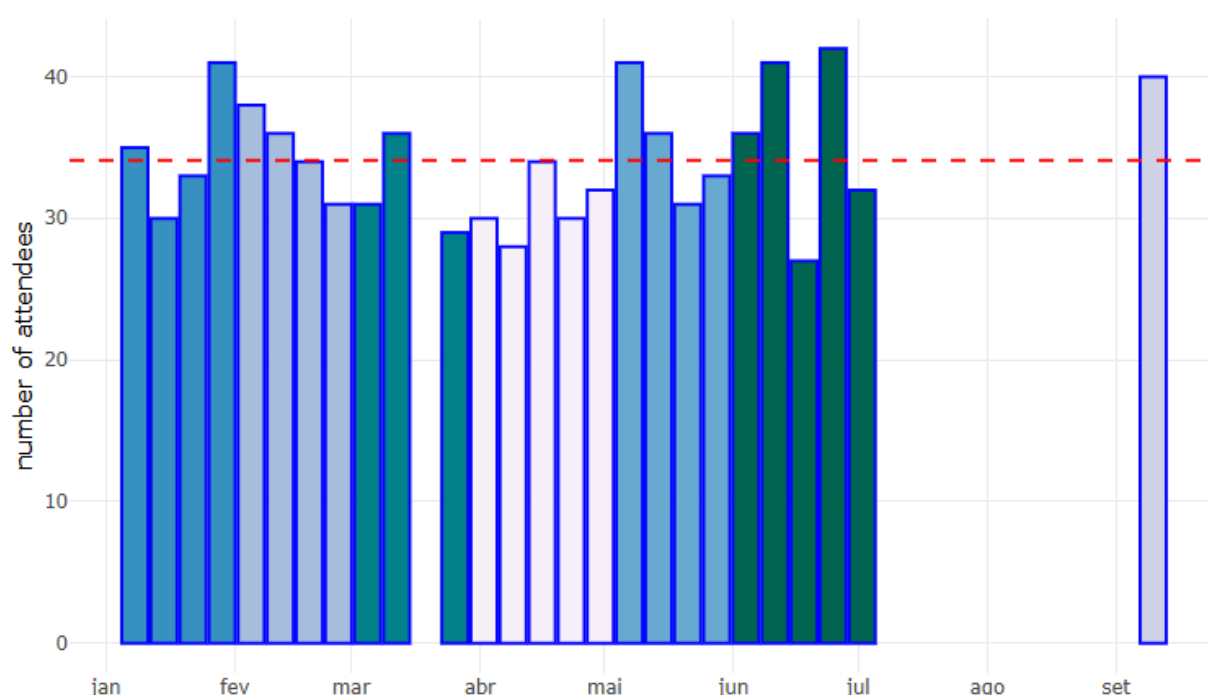
2022; [Relatório do Peer Review](#)

➤ A Iniciativa Meia-hoRa

A inovação pode manifestar-se também em pequenas iniciativas internas, mas com grande potencial de impacto na cultura da organização. A iniciativa Meia-hoRa é também um exemplo claro de como o INE promove a inovação de base, focada no desenvolvimento e partilha de competências.

Como *software* de código aberto, o R não é apenas o *software* estatístico mais popular e amplamente usado no mundo, é também o ambiente ideal para partilhar e colaborar em estatísticas oficiais. A utilização do ambiente R no INE é uma prática estabelecida, não só em projetos de investigação, mas também na produção de estatísticas oficiais. Este ambiente fornece, entre outras coisas, funções para preparar, limpar e analisar dados, metodologia, visualização e criação de aplicações.

Criada em 2023, esta iniciativa visa encorajar a partilha de experiências na utilização do R no contexto de trabalho tanto no INE como com outros parceiros. Para este fim, são feitas curtas apresentações com exemplos concretos em que o R é a ferramenta de trabalho utilizada. A iniciativa decorre semanalmente, através do Teams, e permite que os colaboradores, mesmo os que trabalham remotamente, estejam apenas "a um clique de distância". No ano de 2025, até à data, um total de 133 utilizadores diferentes participou nas sessões, com uma média de 33,9 pessoas por sessão, o que evidencia o sucesso e a relevância desta iniciativa. O sucesso da iniciativa Meia-hoRa só é possível porque existem curadores que definem o tema mensal, moderadores que gerem cada uma das sessões e, acima de tudo, apresentadores que estão dispostos a partilhar um pouco do que fazem.



Esta iniciativa não visa apenas tornar o INE mais inclusivo, permitindo que os seus colaboradores partilhem o trabalho que está a ser feito na organização, mas também, quando possível, com outras Organizações Nacionais de Estatística.

INE, 2023, Meia-hoRa, <https://hour.ine.pt/>

➤ Introdução de abordagens inovadoras na produção mensal do IPC

O INE produz e publica o Índice de Preços no Consumidor (IPC), o indicador estatístico que tem por finalidade medir a evolução mensal dos preços de um cabaz de bens e serviços, representativos da despesa (monetária de consumo final) das famílias residentes em Portugal. Ao longo dos anos têm sido introduzidas várias iniciativas de âmbito inovador, podendo ser destacadas a recolha através de *web scraping* e a integração de dados administrativos.

Desde 2013 que o INE, a par de outros institutos nacionais de estatística que integram o Sistema Estatístico Europeu, está fortemente empenhado na utilização de novas fontes de dados para a produção de estatísticas oficiais na UE tal como se encontra sublinhado no Scheveningen Memorandum.

Neste sentido, e após cerca de 3 anos de exploração das melhores abordagens a adotar, em 2016, entrou em produção a recolha de preços através de *webscraping* que é uma técnica que recorre a soluções informáticas que extraem, de forma automática e segura, os dados de websites, reduzindo o risco de erros de digitação, diminuindo os prazos de recolha e publicação, redução da carga estatística, aumento potencial das amostras, bem como diminuição dos custos associados à recolha tradicional.

O progressivo alargamento da recolha através de *webscraping* a novos sites, mesmo apenas na fase exploratória, permitiu que a produção mensal do IPC fosse mantida, a par da recolha manual, internet e integração de dados administrativos, durante a suspensão da recolha presencial na pandemia do COVID-19.

Esta inovação ao nível da recolha de preços tem vindo a ser alargada de uma forma faseada na produção mensal do IPC, em paralelo com ganhos ao nível da monitorização das variações que se refletem na recolha no terreno, nomeadamente nos produtos da classe alimentar. Acresce que o potencial da recolha através de *webscraping* permite acompanhar a evolução constante do perfil das famílias no que se refere ao consumo, permitindo acompanhar não só a evolução dos produtos adquiridos (p.e. introduzir novos produtos na amostra) mas também as novas formas de venda de produtos, nomeadamente as compras online ou os preços dinâmicos.

Atualmente, os preços de algumas grandes cadeias de artigos de mobiliário, de telemóveis, de vestuário e de calçado são obtidos por *webscraping*, permitindo adotar procedimentos mais robustos de substituição de artigos quando necessário, através do histórico de preços mantido de forma integrada. Dada a diversidade de sites, a exploração de novas abordagens tem vindo a ser assegurada, prevendo-se o alargamento de novos produtos em substituição da recolha tradicional.

Adicionalmente, são usadas fontes de dados administrativos no apuramento dos preços dos combustíveis, dos medicamentos e das rendas de habitação (através dos dados da AT).

➤ **Aplicação móvel com georreferenciação de edifícios com mapas offline – Censos 2021**

Foi desenvolvida uma aplicação denominada eRecenseador para *smartphone* e *tablet* de uso exclusivo dos recenseadores e que permitia a realização e gestão do trabalho no campo, incluindo um módulo de georreferenciação de edifícios. Destacam-se as seguintes funcionalidades: resposta ao questionário de edifício; inserção de registos de edifícios e alojamentos; atribuição do par código/*password* ao alojamento; apoio ao respondente; inserção de eventos de calendário; notificações da cadeia hierárquica; download e gestão de partilha de mapas (*Mobile Map Packages*); e Sincronização.

A aplicação eRecenseador continha ainda um módulo geográfico para localização e gestão dos registos de edifícios. Para otimizar a sua utilização e poupar o consumo de dados, este módulo trabalhava em *offline*, sobre um ficheiro do tipo MMPK que contém uma base de dados geográfica completa da área de trabalho de cada recenseador. Para produzir os 10.401 ficheiros MMPK, um por cada secção do CENSOS 2021, foi desenvolvido pelo INE um processo automático de criação de MMPK e executado num servidor virtual dedicado com 8 CPU e 32 Giga de memória.

➤ **Mailing: Envio automático de mails**

O INE desenvolveu uma solução denominada MAILING para envio assíncrono de mensagens de correio eletrónico (*e-mails*) e controlo da sua entrega e receção. Esta solução é baseada em metainformação, com várias hipóteses de parametrização do envio, sendo possível personalizar os atributos de cada mensagem: título, corpo e anexos. Esta solução foi usada no Censos 2021 para envios de grande quantidade (cerca de 300 mil mails), demonstrando a sua eficácia sem causar impacto na infraestrutura de mail do INE, e continua a ser utilizada regularmente.

➤ **A experiência de classificação automática nos Censos 2021 utilizando algoritmos de Deep-Learning**

Para os Censos de 2021, o INE tinha como objetivo classificar mais de 10 milhões de expressões relativas a profissões, atividades económicas e designações de cursos superiores, com qualidade e num curto espaço de tempo. Adicionalmente, pretendia-se disponibilizar à sociedade uma ferramenta que permitisse a utilização do processo desenvolvido pelo INE.

A estratégia adotada resultou na criação, em 2019, de uma equipa multidisciplinar e interdepartamental, ETDS – Equipa Técnica “Data Science”, com a finalidade de definir a Metodologia de processamento e tratamento de dados dos Censos 2021, em particular a codificação das referidas questões abertas (escolaridade, profissão e atividade económica).

A codificação dos Censos 2021 foi assegurada em quatro etapas:

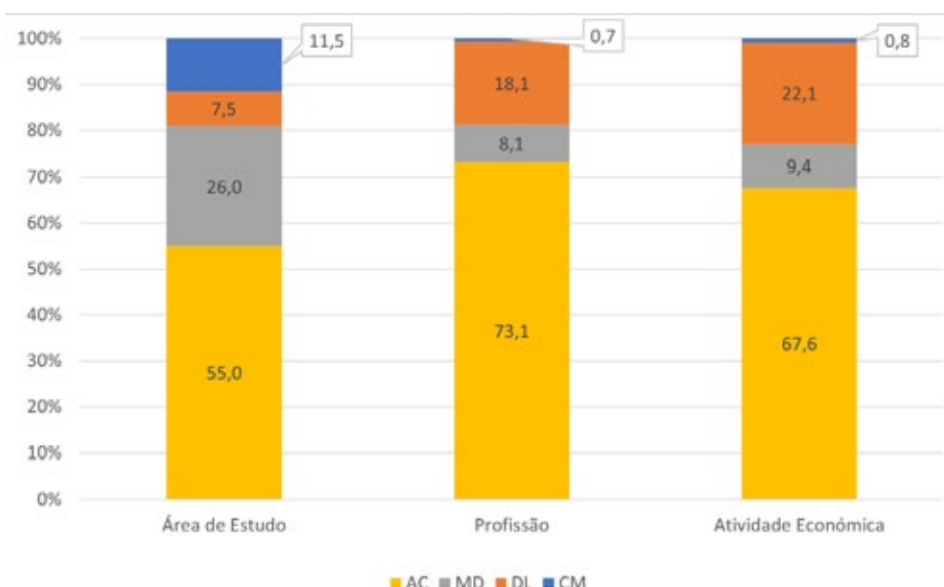
1. Autocomplete (AC);
2. Matching direto Classificadores/Dicionários (MD);
3. Algoritmos de Deep Learning (DL);
4. Codificação manual (CM) (que foi residual).

1. Autocomplete – A primeira etapa foi disponibilizada no questionário web e permitiu a seleção da profissão, atividade económica e curso superior pelo respondente; através do preenchimento de uma palavra ou expressão, foi possível apresentar resultados possíveis e mais prováveis, definidos a partir de uma classificação prévia de frequências. A classificação de frequência teve como base o histórico de codificação nos inquéritos às famílias, tendo este processo sido, ainda, otimizado com a incorporação de palavras-chave, para além das listas de descritivos. O autocomplete foi suportado por uma API desenvolvida pelo INE para uso no processo de codificação automática. Esta API está disponível na internet e possibilita a qualquer utilizador que desenvolva um formulário WEB evocá-la e, através dela, obter uma codificação harmonizada com as nomenclaturas do INE, em vez de desenvolver e usar as suas próprias codificações e agregações.

2. Matching – A segunda etapa correspondeu ao matching direto entre as expressões preenchidas e os classificadores/dicionários que tinham vindo a ser utilizados nos inquéritos às famílias. Foram construídos com base nas listas de expressões codificadas corretamente e ampliados com base no histórico de codificação de mais de 700 mil entrevistas dos Censos anteriores, bem como em técnicas de tratamento de texto, nomeadamente de string distance, que permitiram aproximar as expressões dos classificadores com as respostas dos cidadãos sobre a sua profissão, atividade económica do local onde trabalham e área de estudo ou designação do curso superior.

3. Deep Learning – Por fim, e de uma forma completamente inovadora, foram exploradas soluções de Inteligência Artificial para potenciar a classificação, tendo sido treinada com os dados dos Censos 2011 e os resultados da primeira etapa do Autocomplete dos Censos 2021; de referir que, para além das profissões e atividades, foram incluídas variáveis auxiliares de codificação como as tarefas executadas e os principais bens produzidos. O processo foi totalmente desenvolvido em Python (Tensorflow, Keras) e a arquitetura escolhida foi a Long Short-Term Memory (LSTM), uma versão melhorada da Recurrent Neural Network (RNN). O conhecimento e experiência do processo de codificação automática, atualmente em produção no INE, bem como as várias experiências que os técnicos foram desenvolvendo nas suas áreas de atuação, permitiram alcançar com sucesso os objetivos pretendidos.

Resultados do processo de codificação dos Censos 2021



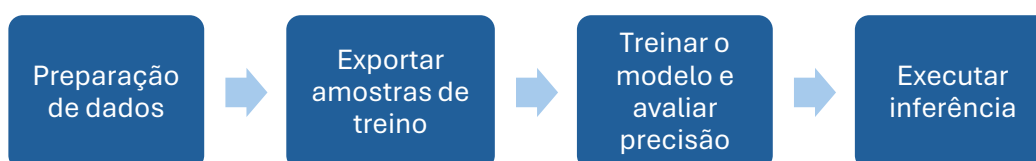
Para disponibilizar à sociedade uma ferramenta que permite a utilização do processo de codificação através do *Autocomplete*, desenvolvido para os Censos 2021, o INE criou três API que podem ser acedidas [aqui](#).

[INE, NEWS 59 – “A experiência de classificação nos Censos 2021 utilizando algoritmos de Deep-Learning”](#)

➤ Inteligência Artificial ao Serviço da Geoinformação Estatística

O Instituto Nacional de Estatística (INE) tem vindo a reforçar a sua capacidade de integração entre dados estatísticos e geoespaciais, promovendo o desenvolvimento da sua Infraestrutura de Informação Geográfica (IIG) através da utilização de tecnologias de Inteligência Artificial (IA). No âmbito do projeto 2023-PT-GEOS, cofinanciado pela União Europeia, o INE desenvolveu modelos de Aprendizagem Profunda (*Deep Learning* - DL) destinados a melhorar a localização dos edifícios, mapeando as áreas de implantação (ou "footprints") de edifícios individuais a partir de imagens de alta resolução. O desenvolvimento destas técnicas irá facilitar a avaliação da exatidão da informação sobre novos edifícios recebida de fontes de dados externas.

A metodologia empregue incluiu a seleção de áreas geográficas diversas, com distintos graus de urbanização, representativas do território de Portugal Continental. A recolha de amostras foi realizada de forma a garantir a representatividade da heterogeneidade territorial e a robustez do modelo a treinar. As amostras foram delineadas em software de Sistemas de Informação Geográfica (SIG) e incluíram um total superior a 120 000 áreas de implantação de edifícios. Estas amostras foram exportadas para serem utilizadas no treino de dois modelos de DL: o Mask R-CNN (*Mask Region-Based Convolutional Neural Network*) e o SAM (*Segment Anything Model*).



Após o treino dos modelos com os algoritmos mencionados, foi realizada uma avaliação da precisão usando métricas padrão, como a Precisão (*Precision*) e a Interseção sobre a União (IoU - *Intersect Over Union*). A análise de erro foi conduzida em 10 Regiões de Interesse (ROI), cada uma abrangendo uma área de 10 quilómetros quadrados, correspondendo a 8 % da área total dos dados de treino.



Figure 1. Áreas de implantação de edifícios em Santa Maria da Feira (APU), SAM



Figure 2. Áreas de implantação de edifícios em Santa Maria da Feira (APU), Mask R-CNN



Figure 2. Áreas de implantação de edifícios em Póvoa de Varzim (AMU), Mask R-CNN



Figure 3. Áreas de implantação de edifícios em Póvoa de Varzim (AMU), SAM



Figure 4. Áreas de implantação de edifícios em Torres Novas (APR), Mask R-CNN



Figure 5. Áreas de implantação de edifícios em Torres Novas (APR), SAM

Os resultados experimentais de acordo com a métrica da Precisão (proporção de edifícios modelados que são subsequentemente identificados com exatidão) revelaram uma precisão entre 81% e 94% no Mask R-CNN e entre 79% e 93% no caso do SAM. Por outro lado, a métrica IoU, que mede a qualidade de uma previsão pela área de interseção ou união entre a área de implantação do edifício prevista pelo modelo e a área de implantação de referência, apresentou valores entre 62 % e 75 % para o Mask R-CNN e entre 60 % e 83 % para o SAM.

Em conclusão, este projeto demonstrou a aplicabilidade e eficácia da metodologia proposta na monitorização do edificado, permitindo obter dados mais atuais e fiáveis. Esta abordagem atua em complementaridade com a informação administrativa do SIOU e contribui para o reforço do controlo e melhoria da qualidade dos dados. O trabalho futuro neste domínio consistirá em expandir a utilização deste modelo para apoiar o processo de produção estatística, nomeadamente para melhorar, conceptual

e topologicamente, a representação e a localização do edificado da IIG, que servirá de suporte essencial ao Censos de 2031 e à geointegração de dados administrativos. O uso de Inteligência Artificial e Observação da Terra reforça a autonomia e eficiência na atualização da Base Geográfica de Edifícios.

Campoverde, C., Koeva, M., Persello, C., Maslov, K., Jiao, W., & Petrova-Antonova, D. (2024). *Automatic building roof plane extraction in urban environments for 3D city modelling using remote sensing data*. Remote Sensing, 16(8), 1386. <https://doi.org/10.3390/rs16081386>

Ji, S., Wei, S., & Lu, M. (2019). *Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set*. IEEE Transactions on Geoscience and Remote Sensing, 57(1), 574–586. <https://doi.org/10.1109/TGRS.2018.2858817>

Li, J., Huang, X., Tu, L., Zhang, T., & Wang, L. (2022). *A review of building detection from very high-resolution optical remote sensing images*. GIScience & Remote Sensing, 59(8), 1199–1225. <https://doi.org/10.1080/15481603.2022.2101727>

Oscó, L. P., Wu, Q., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Junior, J. M. (2023). *The segment anything model (SAM) for remote sensing applications: From zero to one shot*. International Journal of Applied Earth Observation and Geoinformation, 124, 103540. <https://doi.org/10.1016/j.jag.2023.103540>

➤ **One-stop-shop for Artificial Intelligence and Machine Learning for Official Statistics Project**

One-stop-shop for Artificial Intelligence and Machine Learning for Official Statistics Project (AIML4OS) é um projeto ESSnet que envolve 16 países, incluindo Portugal. Teve início em 1 de abril de 2024 e terá uma duração de 4 anos. O seu objetivo principal é criar uma plataforma única para a Inteligência Artificial (IA) e o Machine Learning (ML) aplicados às estatísticas oficiais.

Esta plataforma irá oferecer um ponto de entrada único para o Sistema Estatístico Europeu (ESS), permitindo o acesso a um conjunto de ferramentas e capacidades integradas para a implementação de soluções baseadas em IA e ML. Através desta plataforma, serão disponibilizadas orientações e assistência adaptadas à produção estatística para facilitar a utilização de IA/ML em quadros metodológicos e de execução apropriados, promovendo a aplicação segura e eficiente dessas tecnologias.

O projeto será implementado por meio de 13 pacotes de trabalho (WPs), dos quais 6 são dedicados ao apoio à implementação e 7 são focados em casos de uso práticos. O INE Portugal está envolvido em 3 dos WPs transversais (WP2, WP4 e WP6), e em quase todos os casos de uso; WP7 Earth Observation, WP8 Editing Focus, WP9 Imputation Focus, WP11 Applying ML for estimating firm-level supply chain networks, WP12 Large Language Models e WP13 Synthetic Data.

Estes casos de uso permitirão que as soluções de IA/ML sejam aplicadas a situações reais no contexto das estatísticas oficiais e os utilizadores podem explorar atualizações detalhadas, aceder a recursos e manter-se informados sobre os eventos futuros relacionados com o projeto no site: <https://cros.ec.europa.eu/dashboard/aiml4os>

➤ **A informação Empresarial Simplificada**

A Informação Empresarial Simplificada (IES) foi criada em 2006 no âmbito de um programa governamental de simplificação e modernização da Administração Pública, denominado Programa SIMPLEX.

Esta medida, que resultou de um esforço conjunto de todas as entidades da administração pública envolvidas (o Ministério das Finanças, o Ministério da Justiça, o Instituto Nacional de Estatística e o Banco de Portugal), agrega o cumprimento de quatro obrigações legais, anteriormente dispersas, pelas empresas num único ato. Com a Informação Empresarial Simplificada, as obrigações das empresas perante estas entidades da Administração Pública, nomeadamente o depósito das contas anuais, são integralmente cumpridas por via eletrónica e realizadas num único momento.

Esta medida teve um impacto significativo tanto nas empresas como nas diferentes entidades da Administração Pública envolvidas.

No caso do INE, a integração na IES no sistema de produção estatística foi profundamente transformadora, em particular no domínio das estatísticas económicas.

As principais vantagens da Informação Empresarial Simplificada incluem: uma cobertura completa do universo empresarial – 450.000 em comparação com as 50.000 empresas inquiridas anteriormente; a redução do tempo de disponibilização da informação de 12 para 6,5 meses; a informação ser recebida automaticamente por via eletrónica; um aumento significativo no detalhe da informação; e a eliminação do Inquérito Harmonizado às Empresas, um dos inquéritos mais dispendiosos realizados pelo Instituto Nacional de Estatística.

Realça-se também o papel fundamental que o acesso aos dados da totalidade das empresas teve no processo de dimensionamento amostral de um conjunto significativo de inquéritos, nomeadamente os inquéritos setoriais (Indústria, Construção, Comércio a Retalho, Serviços, etc.). Estes inquéritos beneficiam dos dados da Informação Empresarial Simplificada, tanto no dimensionamento da amostra (amostras mais eficientes e de menor dimensão) como na produção estatística, impactando assim a qualidade e garantindo a coerência com as estatísticas estruturais das empresas.

Os dados da Informação Empresarial Simplificada, integrados ao nível da empresa com outros dados administrativos e de inquéritos, aumentam as possibilidades de produção estatística e investigação estatística (sem impor encargos de declaração adicionais às empresas). Os exemplos de microdados integrados com a Informação Empresarial Simplificada incluem dados administrativos, como os dados do e-Fatura e o IVA da Autoridade Tributária, a Declaração Mensal de Remunerações da Segurança Social, e dados de inquéritos como o Inquérito Comunitário à Inovação, dados do Comércio Internacional de Bens e do Inquérito à Utilização de Tecnologias da Informação e Comunicação nas empresas.

Apesar das inúmeras vantagens, identificam-se algumas limitações, como sejam o adiamento unilateral, por parte do Ministério das Finanças, do prazo de submissão dos dados, em algumas edições; a implementação das mudanças necessárias no Sistema de Informação ser um processo moroso, dificultado por obstáculos burocráticos como a aprovação de decretos-lei e portarias e o desenvolvimento tecnológico.

Desde 2019, dando continuidade ao processo de simplificação iniciado em 2006 que levou à criação da IES, está a ser planeada uma nova simplificação. O objetivo é pré-preencher as Demonstrações Financeiras da IES com dados extraídos do Ficheiro Normalizado de Auditoria para Fins Fiscais (SAF-T) relativos à contabilidade, permitindo eliminar quadros e campos da atual IES, nos casos em que a informação possa ser obtida diretamente da contabilidade das empresas através do SAF-T (*Standard Audit File for Tax Purposes*). Este ficheiro eletrónico normalizado, geralmente em formato XML, é utilizado internacionalmente para a transmissão de dados fiscais e contabilísticos das empresas às autoridades tributárias, tendo sido Portugal o primeiro país a adotar este sistema, em 2008.

Apesar de esta medida não ter sido ainda implementada por ter sido considerada demasiado exigente, espera-se poder dar continuidade à evolução do sistema a breve trecho.

[Ana Chumbau; Sofia Rodrigues, 2018, “Simplified Business Information: the way forward”](#)

➤ e-Fatura

O sistema e-Fatura é uma plataforma da Autoridade Tributária e Aduaneira (AT), criada em janeiro de 2014, que permite a comunicação e o registo eletrónico de faturas, sendo um dos seus principais objetivos combater a fraude e evasão fiscal.

Desde 2013, a lei exige a comunicação eletrónica das faturas à Autoridade Tributária e abrange todas as pessoas singulares ou coletivas com sede, estabelecimento estável ou domicílio fiscal em Portugal. Os dados gerados pelo e-Fatura cobrem todas as transações entre empresas (B2B) e entre empresas e consumidores (B2C).

Desde maio de 2020, o INE tem recebido dados mensais — cerca de 100 milhões de registos por mês — sobre transações comerciais e despesas pessoais declaradas em faturas eletrónicas, de acordo com um protocolo estabelecido entre a Autoridade Tributária e o INE.

O tratamento dos dados do e-Fatura desempenhou um papel essencial como prova do sucesso da estratégia do INE em matéria de dados administrativos.

No que diz respeito a este tipo de dados provenientes de fontes externas, o INE está a seguir os mesmos procedimentos de validação já implementados para os inquéritos. O Departamento de Recolha e Gestão de Dados e o Departamento de Metodologia e Sistemas de Informação trabalham em estreita colaboração na identificação, análise e tratamento de fontes de dados externas, com particular ênfase nas fontes de dados administrativos. O objetivo é facilitar a sua disponibilidade e integração na produção estatística, criando bases de dados consolidadas, melhoradas e enriquecidas que sirvam as necessidades dos vários utilizadores internos e externos. Os procedimentos criados, adaptados dos já existentes para os inquéritos - por exemplo, produção de notas metodológicas, fichas de caracterização, e documentos com o estabelecimento de níveis de serviço para utilizadores internos e externos foram inicialmente definidos para o e-Fatura e estão agora a ser aplicados a todas as outras fontes administrativas.

O INE documentou os fluxos de processamentos dos dados do e-Fatura, tendo sido incorporadas significativas melhorias quer no enriquecimento da base de dados com novas variáveis provenientes de fontes do INE, quer na identificação e tratamento de *outliers* e na consistência dos dados.

Desde 2024, para uma divisão específica da CAE (Divisão 68 – Atividades imobiliárias), o INE substituiu a recolha por inquérito mensal pela utilização exclusiva da informação do e-Fatura. Esta decisão resultou na retirada das empresas desse setor da amostra do inquérito mensal.

Os microdados do e-Fatura estão disponíveis para a comunidade de investigação desde dezembro de 2023, juntamente com um relatório de caracterização, que identifica a estrutura dos dados, as definições das variáveis disponibilizadas e os procedimentos mais relevantes aplicados. O feedback regular aos fornecedores de dados está em fase de desenvolvimento.

A característica única do e-Fatura de ter informação tanto do vendedor como do comprador traz muitas oportunidades e desafios para explorar e integrar com outros dados.

Na perspetiva do comprador, esta fonte de dados permite estudar e produzir novas estatísticas sobre o comportamento do consumidor. Na perspetiva do vendedor, um objetivo central é substituir a comunicação mensal de volume de negócios em todos os inquéritos dos Indicadores de Curto Prazo às empresas. Adicionalmente, os dados foram melhorados e enriquecidos com dados estruturais do INE (como código de atividade e região) e podem ser integrados com outros microdados a nível empresarial.

Apesar das limitações identificadas – relacionadas com a necessidade de limpeza e padronização dos dados brutos e com a garantia da atempada transmissão da informação por parte das empresas e da Autoridade Tributária –, o potencial transformador do e-Fatura é inegável.

O INE está fortemente empenhado em ultrapassar estes desafios, nomeadamente através do desenvolvimento de metodologias mais robustas para melhoria da qualidade, e da coordenação interinstitucional. O objetivo principal continua a ser o alargamento (gradual) da substituição da recolha em inquéritos dos Indicadores de Curto Prazo, utilizando exclusivamente os microdados transacionais

do e-Fatura, concretizando assim a estratégia de simplificação da carga estatística e de otimização da produção estatística.

[2025 Portugal \(2025 eInvoicing Country Sheet\) - eInvoicing Country Fact sheets -](#)

2024; Conferência da Qualidade; Portugal, A., et al; [Administrative Data Quality Challenges through the Lens of Invoice](#)

2021; Moreira, A., et al, ["The use of electronic invoice data in COVID time"](#) – a case study of e-Invoice data usage in Portugal

➤ **Contas Satélite em Portugal**

Nos últimos 20 anos, seguindo a sua estratégia de inovação e de informação mais detalhada, o INE tem vindo a desenvolver e consolidar um vasto conjunto de contas satélite. Este percurso foi iniciado com a publicação da conta satélite da saúde (2004) e do turismo (2005). As contas satélite complementam o sistema de contas nacionais, oferecendo uma visão mais detalhada e permitindo uma análise mais aprofundada de setores, considerados relevantes, da economia e da sociedade.

Portugal tem-se afirmado internacionalmente pelo dinamismo e abordagem inovadora no desenvolvimento e compilação de contas satélite, fundamentais para melhorar o conhecimento, a monitorização e a formulação de políticas públicas nas respetivas áreas. A inovação portuguesa caracteriza-se não só pelo pioneirismo nalguns domínios, pela sofisticação metodológica, pela abrangência de domínios cobertos, mas também pela capacidade de articular parcerias institucionais e de responder às necessidades concretas dos utilizadores.

Entre as contas satélite mais inovadoras, destacam-se:

A **Conta Satélite da Economia Social** é reconhecida como referência internacional, especialmente na Europa, por oferecer uma visão integrada do setor da economia social. A principal inovação reside na capacidade de detalhar diferentes famílias de entidades e os seus estatutos jurídicos, com especial atenção às Instituições Particulares de Solidariedade Social (IPSS). Foi desenvolvida em colaboração com a Cooperativa António Sérgio para a Economia Social (CASES), o que permitiu alinhar o desenho da conta com as necessidades dos atores do setor e reforçar a sua utilidade para o planeamento e avaliação de políticas públicas.

A **Conta Satélite do Mar** foi pioneira a nível mundial, tendo sido desenvolvida em parceria com a Direção-Geral de Política do Mar. Foi concebida como uma ferramenta de monitorização da Estratégia Nacional para o Mar. A sua relevância mantém-se, sendo frequentemente referenciada pela OCDE como exemplo de boas práticas. A colaboração institucional foi essencial para garantir que os indicadores produzidos respondem às exigências de planeamento estratégico e reporte internacional do setor.

A **Conta Satélite do Desporto** destaca-se pela segmentação inovadora da informação por tipo de entidade, permitindo isolar, por exemplo, as sociedades desportivas e evidenciar a heterogeneidade do setor. O projeto foi desenvolvido em articulação com o Instituto Português do Desporto e Juventude, assegurando uma abordagem ajustada às especificidades do setor e às necessidades dos decisores políticos e operadores desportivos.

A **Conta Satélite do Turismo** introduziu a aplicação de matrizes input-output para estimar o impacto monetário indireto da atividade turística na economia. Esta metodologia permite uma análise mais abrangente dos efeitos do turismo, por captar as interações com outros setores económicos. A inovação aqui reside na capacidade de quantificar efeitos multiplicadores, oferecendo aos utilizadores — entidades públicas ou operadores privados — uma ferramenta robusta para avaliação de políticas.

Este modelo de trabalho, baseado em colaboração interinstitucional e em respostas adaptadas às necessidades dos utilizadores, tem sido determinante para o sucesso das contas satélite em Portugal.

➤ **Utilização de dados administrativos: o exemplo do Índice de Preços da Habitação (IPHab)**

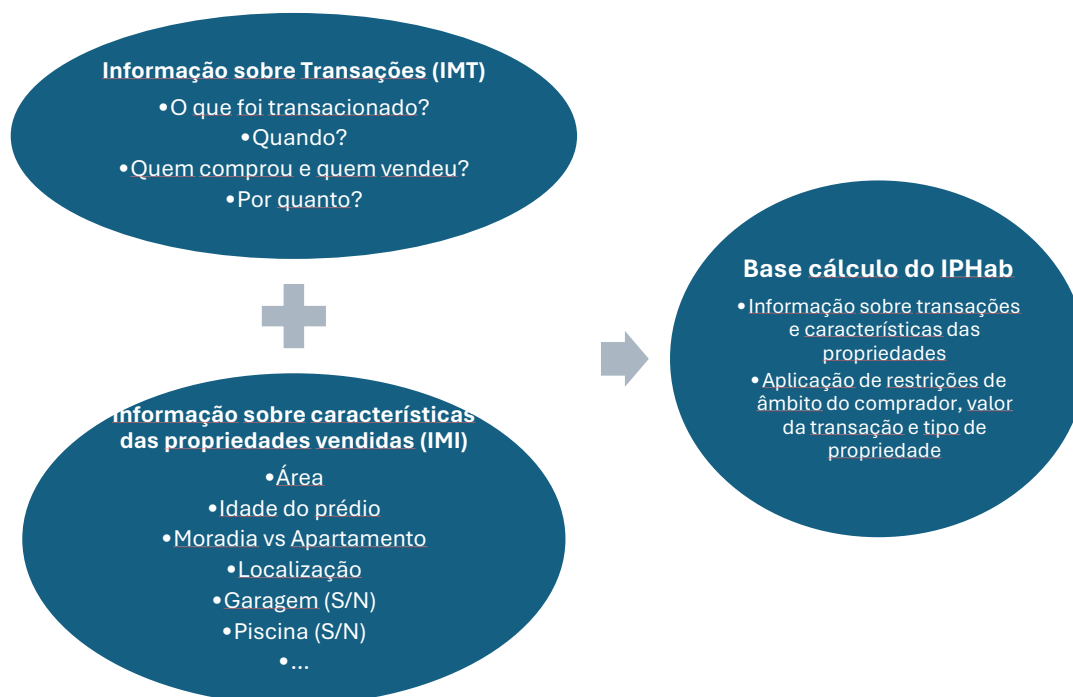
O mercado imobiliário em geral, e da habitação em particular, não tiveram a devida cobertura de informação estatística, nomeadamente quanto ao comportamento dos preços, apesar da relevância que apresentam na atividade e no património do país¹, bem como no orçamento de qualquer família. A última crise financeira e do *subprime*, colocou o tema dos preços do mercado imobiliário, prioritariamente da habitação, na agenda dos principais agentes económicos, sejam governos, bancos centrais, empresas ou Famílias. Definir a política de habitação de um país, avaliar potenciais desequilíbrios macroeconómicos, atuar na política monetária, identificar oportunidades de investimento, ou o momento certo para comprar habitação, dependem de informação fiável, atempada e regular sobre o comportamento deste mercado.

A expressão “não existem duas casas iguais!” sintetiza a complexidade deste “produto”, composto por múltiplas características, algumas subjetivas e dependentes do gosto de cada proprietário. Esta situação torna mais complexa a tarefa de avaliar o preço de cada habitação que é vendida em cada momento, na medida em que o mesmo encontra-se intimamente relacionado com fatores como a área, a qualidade da construção, o estado de conservação, o tipo de construção, elementos complementares de conforto, ou a localização.

Assim, medir corretamente a evolução dos preços da habitação ao longo do tempo exige a construção de índices de preços que considerem os efeitos das diferenças de qualidade das habitações transacionadas, de modo que a variação de preços apurada reflita as reais alterações de preços e não a alteração das características das habitações transacionadas entre os momentos no tempo que estamos a comparar.

Em 2014, o INE iniciou a divulgação do Índice de Preços da Habitação (IPHab), que mede a evolução dos preços dos alojamentos familiares adquiridos pelas Famílias em Portugal. O IPHab é compilado a partir da informação do Imposto Municipal Sobre as Transmissões Onerosas de Imóveis (IMT) e do Imposto Municipal sobre Imóveis (IMI), cedida pela AT. O IMT possibilita identificar nomeadamente a habitação transacionada, o momento da transação, os intervenientes e o valor da transação. Por outro lado, a informação do IMI, permite conhecer um conjunto alargado de características da propriedade, como a localização, área, idade do edifício, tipo de construção, existência de garagem, piscina ou elevador, por exemplo.

¹ O stock líquido de edifícios e alojamentos representa cerca de 65% do total dos Ativos não financeiros produzidos, tendo a habitação um peso de 48%.



A associação entre os dois fluxos de informação, permite construir uma base de dados com informações sobre a transação, mas sobretudo relativa às características das habitações transacionadas em cada momento. A metodologia subjacente ao cálculo do IPHab segue a regulamentação europeia em vigor², baseia-se na estimação de uma relação funcional entre o logaritmo dos preços de transação dos alojamentos e as suas características, enquadrando-se no âmbito dos modelos probabilísticos de “preços hedónicos”. Esta estimação é realizada trimestralmente com dados de dois trimestres adjacentes, para o conjunto das transações, sendo possível, por esta via, controlar as diferenças qualitativas das habitações transacionadas e estimar uma taxa de variação de preços ajustada de efeitos de alteração de qualidade.

A produção e divulgação do IPHab com estas características técnicas, com recurso à integração de duas grandes bases de dados externas e não totalmente padronizadas, e assente numa metodologia de cálculo inovadora, representou um enorme desafio, tendo estabelecido um marco fundamental no que concerne à importância e institucionalização da informação administrativa enquanto input essencial na produção de estatísticas oficiais, sendo igualmente um estudo de caso no contexto europeu, em termos de desenvolvimento dos indicadores de preços do imobiliário.

Este mesmo conjunto de informação e metodologia de cálculo foi utilizada para criar, em 2017, o Índice de Preços das Propriedades Comerciais (IPPCom), o qual tem como propósito medir a evolução dos preços das propriedades comerciais transacionadas no território nacional.

Instituto Nacional de Estatística (2022). *Índice de Preços da Habitação*. Documento Metodológico. INE. <https://smi.ine.pt/DocumentacaoMetodologica/Detalhes/1699>

Eurostat; International Labour Organization; International Monetary Fund; OECD; UNECE; World Bank. (2013). *Handbook on Residential Property Prices Indices (RPPIs)*. Available on <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-12-022>

➤ Censos com dados Administrativos

Os recenseamentos da população e da habitação representam um dos pilares do Sistema Estatístico de qualquer país no que se refere à contagem e caracterização da população e do respetivo parque habitacional, aos níveis nacional, regional e local, produzindo indicadores essenciais para a definição de políticas públicas e para o processo de tomada de decisão.

² [Regulamento \(UE\) 2016/792](#) e [Regulamento de execução \(UE\) 2023/1470](#)

O INE tem em curso um programa de transformação dos Censos para um modelo mais eficiente com recurso a informação administrativa. Pela sua dimensão, este é porventura o mais ambicioso projeto do INE no que se refere à integração de dados provenientes de fontes administrativas, correspondendo a mais uma etapa do percurso trilhado nos últimos anos. Este projeto insere-se no quadro de desenvolvimento da Infraestrutura Nacional de Dados que dá corpo à estratégia do INE de integração e criação de valor para a sociedade a partir de diferentes fontes de dados.

Nos últimos anos, o INE tem desenvolvido um trabalho de investigação aprofundado sobre o potencial da informação administrativa para substituir o modelo censitário tradicional (baseado na recolha através de um inquérito exaustivo à população), alinhando-se com uma tendência internacional crescente, já adotada por um número considerável de países que privilegiam as fontes administrativas como base para a produção de estatísticas censitárias.

A Base de População Residente (BPR) constitui o elemento central do projeto Censos com Dados Administrativos para a área da população: é constituída pela população residente em Portugal e cobre um conjunto de características geográficas, demográficas e socioeconómicas, resultantes da integração de informação administrativa proveniente de diversas fontes da Administração Pública.

Em termos metodológicos, a construção da BPR tem início após um processo exaustivo de avaliação da qualidade dos dados e de integração de identificadores numéricos em cada fonte administrativa³, seguindo-se posteriormente a incorporação da restante informação. Para cada ano, por referência à data de 31 de dezembro, é gerada uma tabela de indivíduos candidatos a residentes em Portugal. Após a constituição deste universo, procede-se à aplicação de um conjunto de regras metodológicas designadas por Índices de residência. Estas regras consistem na verificação da presença do indivíduo em diversas bases de dados administrativas, como, por exemplo: exercício de atividade profissional, frequência do sistema de ensino, declaração de rendimentos fiscais, inscrição em centros de emprego ou outras formas de participação social e económica. Este procedimento é utilizado de forma transversal pelos países que utilizam fontes administrativas para aferir a população residente, na medida em que é essencial assegurar a residência efetiva em território nacional, garantindo, assim, uma aproximação ao conceito de população residente utilizado nas operações censitárias tradicionais.

Um dos princípios fundamentais que norteia este projeto é a confidencialidade da informação. O INE assegura que todos os dados administrativos são objeto de rigorosos mecanismos de segurança na sua transmissão, armazenamento e tratamento, em conformidade com o princípio do segredo estatístico, consagrado na Lei do Sistema Estatístico Nacional, e com as disposições do Regime Geral de Proteção de Dados (RGPD). Assim, garante-se que a informação disponibilizada pelos organismos públicos é utilizada exclusivamente para fins estatísticos, não sendo possível a identificação de um indivíduo em particular ou a utilização por terceiros.

Os trabalhos desenvolvidos para a construção da BPR ao longo de diferentes anos demonstram que a informação administrativa tem um elevado potencial para suportar a produção de estatísticas de natureza censitária. No entanto, em algumas áreas, essa informação ainda não está disponível ou apresenta limitações em termos de cobertura e detalhe, o que evidencia a necessidade de continuar a aprofundar e complementar as fontes existentes.

Neste sentido, a realização dos Censos 2021 revelaram-se uma oportunidade única para aferir os resultados do trabalho desenvolvido no âmbito do projeto BPR, permitindo avaliar a sua robustez enquanto alternativa ao modelo censitário tradicional. A comparação entre os Censos 2021 e a Base de População Residente confirmou a consistência dos resultados obtidos pela BPR, ao mesmo tempo que permitiu identificar oportunidades de melhoria no seu processo de produção.

Segundo a estimativa da BPR, em 2020 residiam em Portugal 10 434 817 indivíduos, valor que difere apenas 0,6% dos resultados ajustados dos Censos 2021⁴, evidenciando uma convergência entre os dois

³ A não existência de um identificador numérico único em Portugal e transversal às diferentes fontes administrativas coloca desafios acrescidos nesta área

⁴ Os resultados dos Censos 2021 com data de referência a 19 de abril de 2021 foram ajustados, através de recuo demográfico, a 31 de dezembro de 2020

modelos e reforçando a fiabilidade da informação administrativa como base para a produção estatística, não obstante os desvios serem mais significativos à medida que o detalhe geográfico se torna mais fino.

De referir ainda, que no âmbito da comparação Censos 2021 e BPR foi possível identificar 9 522 517 indivíduos comuns às duas bases, correspondendo a 92,1% da população recenseada nos Censos 2021 e a 91,7% da população da BPR. Os resultados relativos às taxas de igualdade para as variáveis de caracterização dos indivíduos, cuja ligação entre BPR e Censos foi possível, mostram valores particularmente elevados, acima de 97% para as variáveis demográficas: Sexo, Idade, Estado civil, Naturalidade e Nacionalidade.

As variáveis socioeconómicas — como a Condição perante o trabalho, Situação na profissão, Ramo de atividade, Profissão e Nível de escolaridade completo — apresentam taxas de correspondência mais baixas, entre 60% e 90%, o que pode ser explicado por diferenças conceptuais, nomeadamente nos momentos de referência utilizados, nos critérios de classificação ou no grau de atualização da informação, entre outros fatores que influenciam a comparabilidade dos dados. De referir também que o nível de preenchimento destas variáveis na BPR é ainda parcial, por não existir informação administrativa passível de classificar a generalidade da população residente em Portugal nestas dimensões.

A informação administrativa atualmente integrada na BPR evidencia-se como altamente relevante para a caracterização da população residente em Portugal. No entanto, subsistem áreas críticas que continuam a ser objeto de desenvolvimento, com vista a assegurar as condições necessárias para uma transição sólida e sustentada para um modelo censitário baseado em fontes administrativas.

A concretização deste projeto depende, de forma decisiva, da colaboração contínua e empenhada das entidades da Administração Pública que disponibilizam dados administrativos, sendo de destacar o contributo das que já participam ativamente neste processo, colaboração essa que se revela imprescindível para a modernização do Sistema Estatístico Nacional e para o reforço da confiança nas estatísticas oficiais.

O INE mantém-se fortemente empenhado em aprofundar este caminho, procurando superar as condicionantes identificadas e avançando na integração de informação relativa a indivíduos, agregados familiares e alojamentos.

[INE, 2019, Statslab – “Censos com dados administrativos”](#)

[Joana Araújo, Cecília Cardoso, Cláudia Pina, Paula Paulino, 2024 “ Census data and Administrative Data in Portugal: Results and challenges”](#)

➤ **Integração de dados administrativos em Inquéritos às Famílias**

A recolha de informação através de inquéritos às famílias enfrenta atualmente diversos desafios, incluindo taxas elevadas de não resposta, elevado recurso a respostas por proxy -cerca de 40% no Inquérito às Condições de Vida e Rendimento (ICOR) 2024 -, imprecisões nas respostas e relutância em fornecer informações sobre rendimento, riqueza e despesas pessoais. Para mitigar estes desafios, o INE tem implementado uma estratégia inovadora de integração⁵ de dados administrativos nos inquéritos, representando uma mudança significativa face à prática anterior, em que a informação era recolhida por entrevista e os dados administrativos eram utilizados apenas durante a fase da calibração estatística.

A utilização de informação administrativa sobre rendimentos e despesas baseia-se num processo de record linkage entre os indivíduos da amostra de cada operação estatística e a Base da População Residente (BPR). A integração é realizada a partir de variáveis que identificam alojamentos e respetivos residentes, permitindo posteriormente associar os microdados provenientes de diversas entidades, nomeadamente a Autoridade Tributária (AT), a Segurança Social e o Instituto dos Registos e Notariado (IRN). De referir que a partir de 2024, no formulário eletrónico foi incluída uma questão (não

⁵ No contexto deste documento, entende-se por integração a substituição de informação obtida por recolha direta por informação igual disponível em dados administrativos

obrigatória) sobre o Número de Identificação Fiscal (NIF) de todos os indivíduos do agregado. Após registo e validação do NIF, este é automaticamente encriptado, garantindo que o INE não conhece o NIF original. Esta solução confere maiores garantias de segurança e privacidade ao respondente, facilitando a resposta a esta questão, fundamental para o cruzamento com os dados administrativos fiscais. Nos casos em que a ligação direta é possível, prescinde-se dos métodos de record linkage.

Desde 2021, tem sido possível emparelhar cerca de 89% e 95% das amostras à BPR, sendo que a parte não ligada se deve sobretudo a indivíduos que passam a integrar o agregado familiar em momento posterior à disponibilização da BPR.

O processo de integração de dados administrativos consiste, de forma genérica, em associar a cada indivíduo da amostra informações sobre rendimento, despesa ou outros atributos obtidos administrativamente, sendo realizado um trabalho prévio de harmonização das definições utilizadas no regulamento de cada operação estatística com as correspondentes definições nos dados administrativos. Este processo de harmonização assume particular importância no caso dos rendimentos, devido à existência de rendimentos isentos de imposto que não estão sujeitos a declaração no Modelo 3 do IRS nem na Declaração Mensal de Remunerações da AT (DMR_AT), mas que integram o conceito estatístico de rendimento disponível das famílias.

Desde 2023 que os rendimentos e contribuições sociais obrigatórias provenientes do Anexo A do Modelo 3 do IRS e da respetiva Nota de Liquidação são integrados nas componentes do rendimento das operações estatísticas do ICOR.

A Figura 1 ilustra o trabalho de comparação das distribuições de rendimento bruto do trabalho por conta de outrem (TCO) na amostra emparelhada tendo em conta as duas fontes, uma descrição mais pormenorizada sobre a integração de dados administrativos no ICOR pode ser encontrada aqui (artigo Q2024).

Figura 1 – Rendimentos TCO: Informação administrativa vs. informação recolhida por inquérito

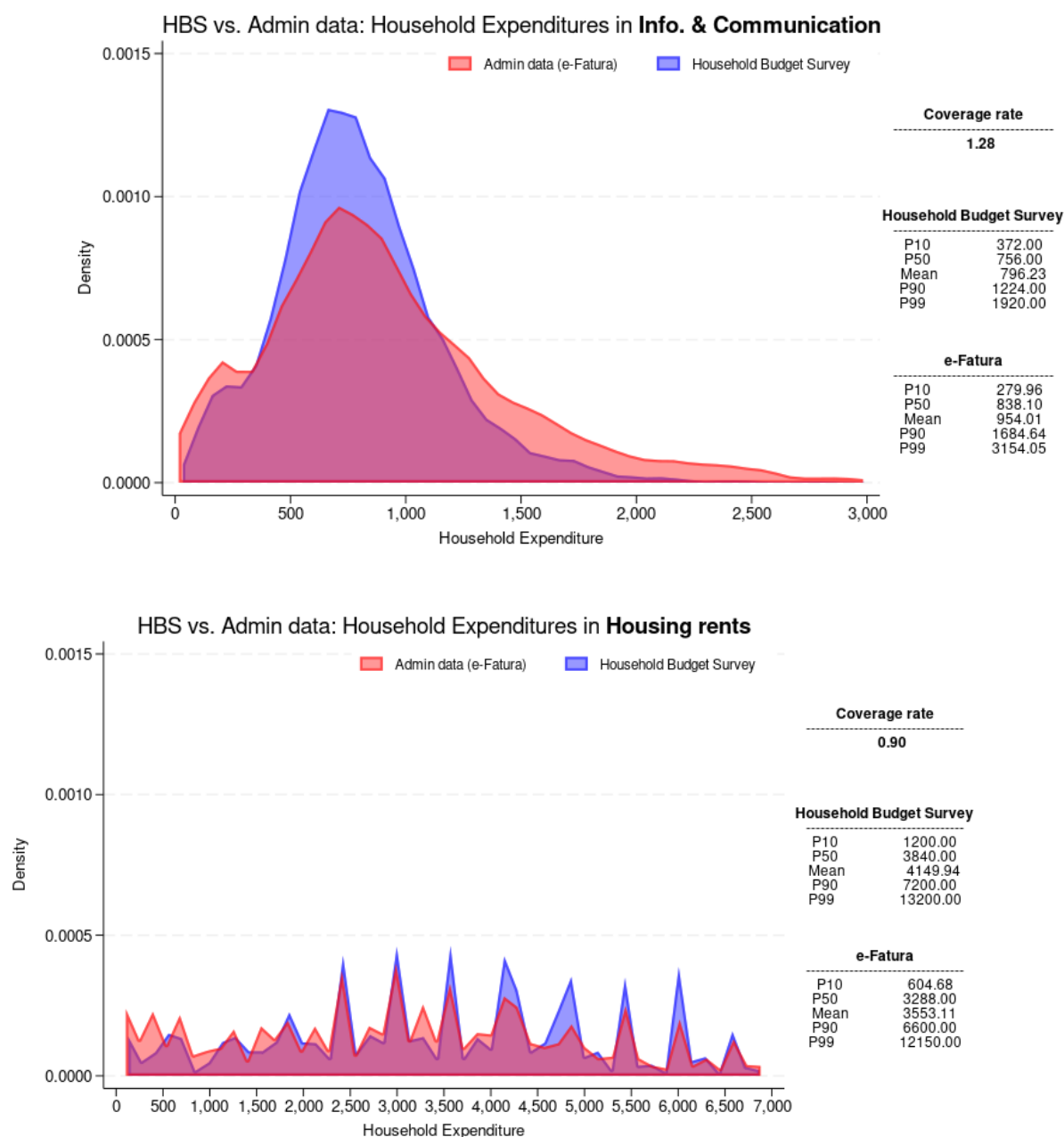


Fonte: (artigo Q2024)

Na edição do ICOR de 2025 (referente aos rendimentos de 2024), prevê-se a integração dos rendimentos dos trabalhadores independentes (Modelo 3 do IRS, Anexo B), bem como de algumas prestações sociais pagas pela Segurança Social, nomeadamente as pensões sociais de velhice. Na próxima edição do Inquérito às Despesas das Famílias (IDF), o INE prevê alargar a integração a determinadas categorias de despesa registadas no sistema e-Fatura que sejam regulares e mediadas por contrato em que o emitente esteja obrigado a incluir o NIF do consumidor, tais como água, eletricidade e telecomunicações. Prevê-se igualmente a integração das despesas com rendas de habitação permanente, com base nos Recibos Eletrónicos de Renda/Modelo 44. Para este efeito, o INE tem desenvolvido estudos exploratórios a partir

da amostra emparelhada do IDF 2022/2023 com os dados do e-Fatura e da informação da AT sobre rendas, com o objetivo de comparar as despesas recolhidas por inquérito com as correspondentes informações administrativas, conforme ilustrado na Figura 2. No estudo (artigo UNECE) são apresentados resultados adicionais para outras categorias de despesa, bem como uma explicação pormenorizada deste exercício.

Figura 2 - Despesas recolhidas pelo IDF 2023 vs. despesas obtidas através dos dados administrativos



Fonte: (artigo UNECE)

Por último, destacam-se os estudos de viabilidade em curso, com vista à eventual integração de informação mensal sobre remunerações líquidas proveniente das DMR-AT no Inquérito ao Emprego (IE), a partir do emparelhamento das amostras do IE com dados administrativos relativos ao período de 2019 a 2022. Os resultados deste exercício podem ser consultados nos estudos (artigo seminar ENBES) e (artigo workshop do LFS). Importa ainda referir que não é apenas a informação proveniente da AT ou da Segurança Social que tem sido integrada nos inquéritos. Na fase de validação da informação recolhida por entrevista, o INE tem igualmente recorrido à informação do IRN, nomeadamente da Base de Dados de Identificação Civil, para construir as relações de parentesco entre os indivíduos da amostra

emparelhada, bem como para obter outros atributos relevantes, como a atividade económica do empregador (segundo a Classificação Portuguesa das Atividades Económicas – CAE).

Em síntese, a aposta do INE na utilização de dados administrativos, não apenas para fins censitários, mas também como integração nos inquéritos às famílias, constitui uma inovação estruturante na recolha e tratamento da informação estatística. Esta estratégia permite reduzir custos, diminuir a carga sobre os inquiridos, aumentar a precisão e cobertura da informação, melhorando a qualidade dos dados e alinhando-se com as recomendações internacionais em matéria de modernização e inovação estatística (European Commission, 2022; Eurostat, 2013).

No futuro, o INE pretende consolidar e aprofundar esta abordagem, reforçando a integração dos dados administrativos relativos ao rendimento e à despesa, mas também à riqueza das famílias, bem como expandir a sua aplicação a novas dimensões, tais como prestações sociais, habilitações e situação face ao emprego, entre outras.

European Commission (2022); Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Better assessing the distributional impact of Member States' policies, 494 final. [European Commissionhttps://ec.europa.eu/social/BlobServlet?docId=26123&langId=en](https://ec.europa.eu/social/BlobServlet?docId=26123&langId=en)

Eurostat (2013); "The use of registers in the context of EU-SILC: challenges and opportunities"; Edited by Markus Jäntti, Veli-Matti Törmälehto and Eric Marlier; <https://ec.europa.eu/eurostat/documents/3888793/5856365/KS-TC-13-004-EN.PDF>

INE 2023; P. Cruz, D. Leite, A. Portugal, F. Ribeiro, S. Rodrigues, D. Santos; "Measuring Monthly Net Incomes from linked Labour Force Survey Data with Administrative Records", Statistics Portugal, European Network for Better Establishment Statistics (ENBES); <https://drive.google.com/file/d/1w6drNqJFSsUDAO3duoWwwNobJSIc5RTJ/view>

INE 2024; E. Góis, C. Farinha Rodrigues, D. Gomes, D. Leite, M.M. Pinho); "Integration of income administrative data into the Portuguese household income distribution: a first national experience using employees' income tax data"; European Conference on Quality in Official Statistics (Q2024); <https://airdrive.eventsair.com/eventsairwesteuprod/production-leading-public/ddb4edb2dc144553a755435a1aabb9e4>

➤ Carta de Equipamentos e Serviços de Interesse Geral (CE-SIG)

Em 2025, o INE disponibilizou a aplicação interativa Carta de Equipamentos e Serviços de Interesse Geral (CE-SIG). Esta iniciativa é um dos produtos estatísticos desenvolvidos no âmbito do projeto Coesão Territorial e Serviços Sociais de Interesse Geral (POAT 2020) que visa apoiar a monitorização do ciclo de Programação da Política de Coesão 2021-2027.

Enquadrada no StatsLab – Estatísticas em desenvolvimento, esta aplicação disponibiliza ferramentas de pesquisa e análise espacial numa plataforma WebSIG, que permite a consulta de equipamentos e serviços de interesse geral, de natureza institucional pública ou privada, com resultados localizados em base pontual.

A aplicação CE-SIG permite também observar o nível de provisão de serviços de interesse geral às populações e acompanhar os padrões territoriais de desenvolvimento. Disponibiliza ainda informação de caracterização dos equipamentos e dos serviços que prestam, dos territórios nos quais se encontram localizados, e informação sobre geografias de acessibilidade e procura dos equipamentos/serviços. Adicionalmente, o utilizador tem acesso a geografias de acessibilidade e de procura dos equipamentos/serviços, nomeadamente áreas de serviço (com base em distâncias físicas e distâncias tempo a pé e em automóvel ligeiro) e dois tipos de áreas de influência (potencial e efetiva) e respetivas métricas associadas.

Os dados de base relativos aos equipamentos e serviços no âmbito do CE-SIG provêm de diversas fontes, e são sujeitos a validação, tratamento e integração com as restantes bases existentes no INE, aproveitando as potencialidades da informação associada à Infraestrutura Nacional de Dados.

Nesta primeira divulgação foram disponibilizados equipamentos de quatro domínios setoriais: Cultura, abrangendo os museus para o ano de 2023; Educação, cobrindo os estabelecimentos de ensino superior no ano letivo 2023/24; Proteção civil, que cobre os corpos de bombeiros do ano 2023; e Saúde, que abarca os estabelecimentos hospitalares em 2022.



Fonte: INE, Carta de Equipamentos e Serviços de Interesse Geral (CE-SIG)

A aplicação disponibiliza, para cada equipamento georreferenciado, elementos para a sua caracterização, como serviços oferecidos, recursos humanos e materiais disponíveis. Adicionalmente, apresentam-se indicadores territoriais e de contexto para os diferentes níveis geográficos, como população residente, a densidade populacional, a superfície, ou o número de agregados domésticos e alojamentos familiares. O mapeamento desta informação permite a visualização das áreas geográficas servidas pelos equipamentos georreferenciados, proporcionando uma nova perspetiva do território.

Esta divulgação corresponde à primeira apresentação de um projeto que deverá ter atualizações futuras, procedendo-se à atualização anual da informação disponibilizada e ao alargamento da oferta de informação a novos domínios setoriais e de equipamentos, bem como a outros modos de transporte.

Paralelamente, e tirando partido da informação produzida na CE-SIG, são disponibilizados indicadores de acessibilidade a equipamentos e serviços de interesse geral no âmbito do projeto IAssLocal - Indicadores de Assimetria ao nível Local e Inter-regional. O IAssLocal, desenvolvido também no contexto do projeto Coesão Territorial e Serviços Sociais de Interesse Geral (POAT 2020), tem como objetivo disponibilizar novos indicadores para a caracterização da diversidade socioeconómica dos territórios, aproveitando igualmente as potencialidades da informação associada à IND – Infraestrutura Nacional de Dados do INE. Estes indicadores apresentam informação a nível municipal e regional, abrangendo áreas como o mercado de arrendamento, a aquisição de habitação, as taxas de esforço associadas aos empréstimos bancários para aquisição de imóveis, o rendimento familiar, os rendimentos dos trabalhadores, a demografia e agora também a acessibilidade a equipamentos.

[INE, 2025, Destaque “Carta de Equipamentos e Serviços de Interesse Geral \(CE-SIG\) – Aplicação Interativa](#)

4. Conclusões

O percurso de inovação no Instituto Nacional de Estatística (INE), conforme detalhado nos exemplos apresentados, comprova que a inovação é uma necessidade estratégica e uma transformação holística que transcende a mera atualização tecnológica. O cerne desta estratégia é a Infraestrutura Nacional de Dados (IND), que institucionalizou a abordagem por processos e a integração de dados, garantindo estatísticas com maior relevância, oportunidade e granularidade.

Tendo em conta os exemplos apresentados, podem distinguir-se quatro principais pilares nesta estratégia de modernização, nomeadamente:

- i) a reengenharia organizacional e de processos, em que se destacam o modelo centralizado de gestão de dados e o Sistema Integrado de Gestão, que reforçam a qualidade, segurança e capacidade de resposta do INE perante crises inesperadas (ex: COVID-IREE);
- ii) a cultura de colaboração, organizacional e externa, através de iniciativas como a ligação com a academia ou o projeto meia-hoRa, que fomentam uma cultura interna de partilha e *upskilling*, e apoiam o desenvolvimento da IND;
- iii) a adoção de novas tecnologias e métodos, integrando técnicas avançadas como o *webscraping* para o IPC, o *Deep Learning* para a classificação automática nos Censos 2021, e a IA no domínio da geoinformação, que torna o INE mais eficiente e sustentável na produção estatística;
- iv) o aproveitamento cada vez mais intensivo de fontes externas, como a IES e o e-Fatura, integrando-as com fontes de recolha direta. Destacam-se a aposta de desenvolvimento dos Censos administrativos e de novos produtos e serviços. Esta abordagem não só permite reduzir significativamente a carga estatística sobre os respondentes como também o desenvolvimento de novos indicadores (Índice de Preços da Habitação); a melhoria da qualidade e cobertura dos inquéritos às famílias (ICOR e IDF) e novos produtos e serviços (CE-SIG).

Em suma, a inovação no INE é um compromisso contínuo com a eficiência, a credibilidade e a relevância da estatística oficial.

Este sucesso é inseparável da colaboração institucional, nomeadamente com as entidades da Administração Pública que fornecem os dados essenciais para a concretização da IND e a modernização do Sistema Estatístico Nacional.



Not Just Another Black Box: AI in the Age of Trusted Statistics

Sónia Quaresma, sonia.quaresma@ine.pt

Instituto Nacional de Estatística, I.P.

What Has Generative AI Brought to the Table for Official Statistics?

The emergence of Generative AI (GenAI), and particularly Large Language Models (LLMs), has introduced a wave of innovation across sectors, and official statistics is no exception. These technologies offer new capabilities that can transform how statistical agencies process data, build infrastructure, and interact with users. However, with their strengths come significant limitations and challenges that must be carefully weighed, especially in a field where accuracy, transparency, and accountability are foundational.

Some of the most relevant powerful capabilities that could support innovation and efficiency in various stages of statistical production include:

- **Extracting Structure from Unstructured Data** GenAI can analyse free-text inputs (like survey comments or administrative records) and extract relevant structured variables, enabling statistical use of previously untapped data sources.
- **Processing Multiple Data Modalities** Modern GenAI tools can interpret not just text, but also audio, images, and video, potentially expanding the scope of data collection and integration (e.g., image classification, speech-to-text processing in interviews, video metadata analysis).
- **Support for Causality and Pattern Detection** While not a substitute for formal causal modelling, LLMs can surface relationships or hypotheses from large datasets that may guide further statistical analysis or exploratory work.
- **Semantic Interoperability** GenAI helps harmonise data from different sources by aligning schema definitions, resolving naming conflicts, and disambiguating contextual meanings, key for cross-country or multi-source statistical projects.
- **Contextual Aggregation and Disambiguation** By understanding the broader linguistic context, LLMs can aggregate values intelligently or clarify ambiguous terms (e.g., differentiating “income” as personal or household, depending on use case).
- **Coding Made Easier** LLMs can generate statistical code (e.g., in R, Python, SQL) based on natural language prompts, translate code between languages, or suggest efficient workflows, accelerating development and prototyping.
- **Synthetic Data Generation** GenAI can produce realistic synthetic datasets useful for testing methods, system development, or training purposes, while helping protect the privacy of real individuals.

Despite its potential, GenAI also brings a set of significant risks that are particularly sensitive in the field of official statistics, where trust, transparency, and accuracy are paramount:

- **Uncontrolled Uncertainty and Hallucinations** LLMs may produce plausible sounding but entirely incorrect or invented content, known as hallucinations, posing a risk of misinformation if outputs are not validated.

- **Bias** Inherited from their training data, LLMs may produce biased or unfair outputs, which can conflict with the principles of impartiality and fairness that underpin official statistics.
- **Lack of Transparency** LLMs is often considered “black boxes” making it difficult to understand or trace the rationale behind specific outputs.
- **Numerical Inaccuracy** GenAI models are poor at performing calculations or interpreting numeric data accurately, making them unsuitable for tasks requiring exact values or arithmetic precision.
- **Limited Reasoning Capabilities** While LLMs can reproduce patterns seen in training data, they often fail at logical reasoning, deduction, and generalising to unseen situations, limiting their utility in analytical tasks requiring inference or abstraction.
- **Suggestive but Not Self-Solving** LLMs can suggest code structures or outline problem-solving steps but cannot execute code or evaluate its correctness. They are assistive, not autonomous decision-makers.
- **High Computational Cost** Running GenAI models, especially at national scale, requires significant infrastructure, energy, and maintenance, raising questions of cost-efficiency and sustainability.
- **Lack of Theoretical Guarantees** Unlike traditional statistical models, GenAI tools typically lack the formal mathematical foundations needed for reproducibility, error bounds, or inference guarantees.
- **Scalability Challenges with Big Data** GenAI models are not inherently optimised for large-scale data processing. Integrating them into big data pipelines can be costly and technically demanding.
- **Vulnerability to Prompt Hacking** LLMs can be manipulated through carefully crafted inputs to produce misleading or inappropriate outputs, raising concerns about reliability, especially in open-ended tools.
- **Privacy and Confidentiality Risks** If not properly controlled, LLMs may unintentionally reproduce sensitive information seen during training, which could violate confidentiality policies fundamental to official statistics.

Should We Still Make the Case for Using LLMs?

Given the many limitations outlined, from hallucinations and reasoning gaps to confidentiality risks, it's natural to question whether LLMs have a place in the rigorous, standards-driven world of official statistics. However, the answer is not to reject these tools, but to use them wisely and within clearly defined boundaries.

LLMs should not replace traditional statistical processes or analytical models. Instead, they should be seen as complementary tools that, when applied thoughtfully, can support specific tasks within the statistical workflow, particularly those involving language, classification, or user interaction. Their greatest value lies in assisting experts, enhancing accessibility, and helping users make better sense of the growing volume and complexity of information available.

Aligning Use with Statistical Principles

To ensure their use is both beneficial and responsible, the application of LLMs in official statistics must be carefully aligned with the core principles that guide statistical production:

Accuracy and Reliability: Outputs must be verified, contextualised, and never used as the sole source of truth for statistical indicators.

Relevance and Timeliness: LLMs can help speed up interaction and document processing but must rely on curated, up-to-date sources to avoid spreading outdated information.

Transparency and Impartiality: Clear documentation of the models, training data, and limitations is essential to maintain trust and ensure fair use.

Explainability: Mechanisms should be implemented to trace how conclusions were reached, particularly when used to assist decision-making or public communication.

Accessibility: Well-designed LLM interfaces can make statistical information more approachable, especially for non-expert users.

Confidentiality: All implementations must rigorously safeguard private data, preventing the leakage of sensitive information from training or interactions.

INE recognises the need to evolve alongside new data ecosystems, technologies, and user expectations. Innovation is not optional, it is essential. That said, not all innovation is equally appropriate. The key is responsible innovation: introducing new tools like LLMs where they can genuinely add value, without compromising the core values of official statistics.

From Potential to Practice: Prototyping with LLMs at INE

The potential of Large Language Models (LLMs) in official statistics is not just theoretical, it is being actively explored through practical experimentation. At INE, several promising use cases have emerged, including:

Summarising historical documents: LLMs can assist in synthesising large volumes of previously published material, such as statistical reports, metadata, and methodological notes, making it easier to extract relevant insights and provide contextual overviews.

Helping users navigate published content: With the increasing amount of information available on INE's portal, LLMs could help users find relevant tables, publications, or concepts by interpreting natural language queries and mapping them to structured resources.

Supporting metadata management: LLMs may support harmonisation of terminology, the completion of metadata fields, or the alignment of definitions across sources, improving coherence and reducing manual work.

These ideas were explored in practice during a collaborative hackathon hosted by INE in June 2025, as part of Work Package 12 of the AIML4OS project (Artificial Intelligence and Machine Learning for Official Statistics). The event brought together participants from multiple national statistical institutes - Portugal, France, the Netherlands, Sweden, Ireland, and Norway - working in a shared environment enabled by SSPCloud.

Three functional prototypes were developed:

- A tool that uses LLMs to detect and extract thematic content from statistical websites;
- A dissemination summary generator capable of producing concise content in both local languages and English;
- A tool to extract quantitative data from PDF documents, helping to structure information that is often locked in unstructured formats.

These prototypes reflect INE's approach to innovation: using AI not as a replacement for expert processes, but as a support tool for improving usability, efficiency, and accessibility, always within a framework of transparency, quality, and ethical standards.

Bringing It All Together: A Prototype for Smarter Dissemination

Among the several promising prototypes developed during the AIML4OS hackathon, one stood out for its relevance and adaptability to statistical production workflows: the Dissemination Summary Prototype, led by the Portuguese team.

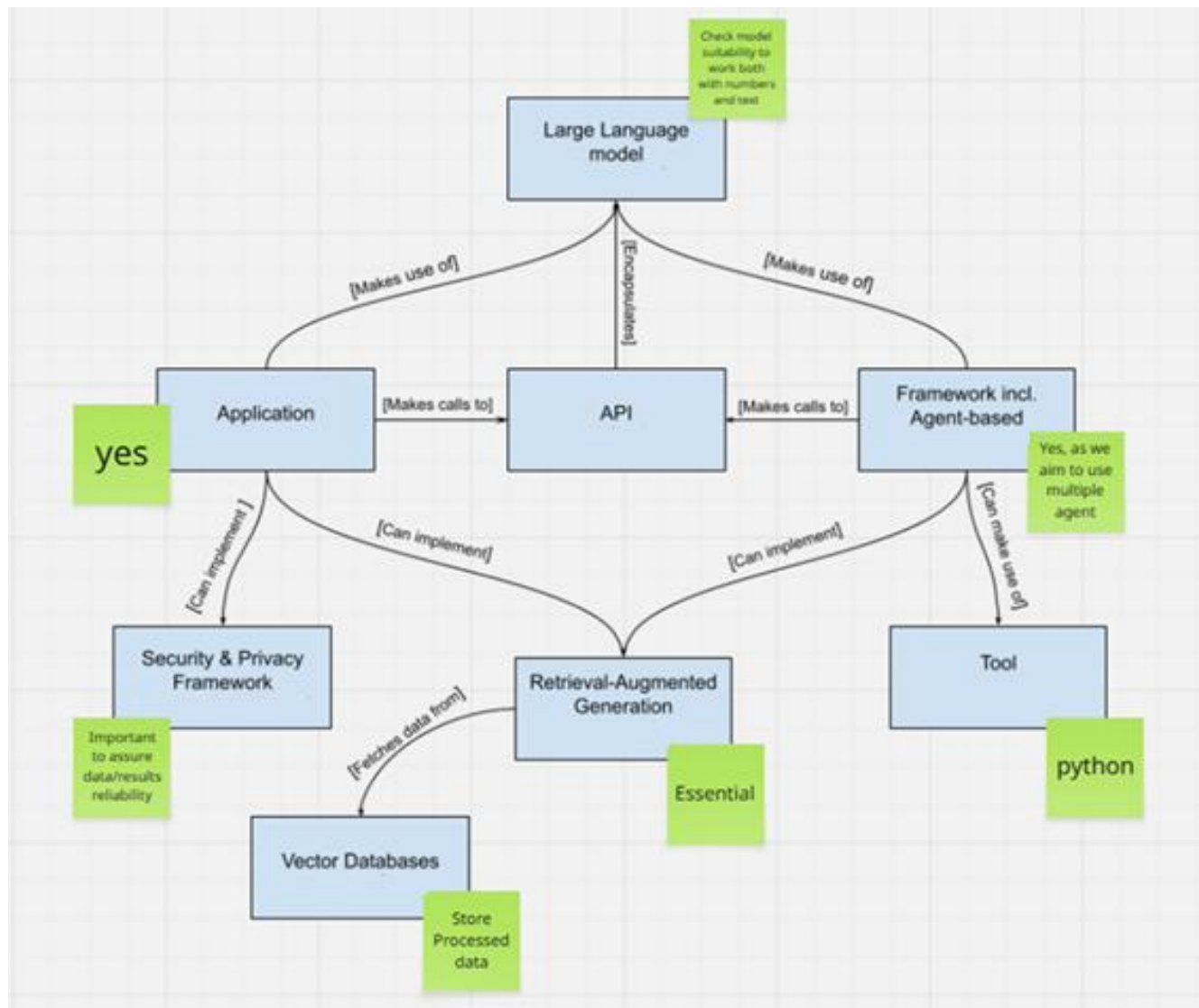


Figure 1 - Example of Architecture integrating LLMs in Statistical Production

At the heart of this prototype is a common yet time-consuming task faced by many statistical offices, summarising analytical reports into clear, concise overviews for publication. These summaries, sometimes of historical data, are typically required in both the local language and English. The challenge is not just linguistic; it's about ensuring consistency, clarity, and contextual fidelity across domains.

The Portuguese team set out to explore whether Large Language Models (LLMs) could streamline this process without compromising on quality or control. The result was a simple, effective AI-powered workflow designed with reusability and institutional constraints in mind.

The system begins by ingesting a report—usually a PDF authored by a subject-matter expert—and extracting its content for semantic processing. A Retrieval-Augmented Generation (RAG) pipeline is then deployed, enabling the LLM to work with the actual content of the report rather than relying on pretraining alone. From this, the model generates a short summary in the desired languages, along with a set of tags that describe the document's thematic areas.

This kind of automation could offer real value in statistical dissemination: accelerating editorial workflows, improving consistency across bilingual outputs, and ensuring thematic metadata is applied

more systematically. The prototype was evaluated on dimensions such as reusability, feasibility, and compatibility with on-premise infrastructure, and scored high on all fronts.

Under the hood, the system uses LangChain, a specific Python framework to build applications using LLMs, to orchestrate the process. It may optionally integrate with ChromaDB, a database that may store the keywords and terms found in the documents, as vectors supporting the search of high-dimensional embeddings (e.g. from text), enabling fast and relevant semantic search. For language generation, several serving options are supported depending on the deployment context. These include a local instance of Ollama running Llama 3.3, a Mistral model hosted within the SSPCloud infrastructure, or paid access to OpenAI models. This flexible architecture allows the team to explore and balance performance, cost, and infrastructure constraints across different environments.

More than just a tech demo, this prototype offers a glimpse into how AI can be embedded thoughtfully into statistical processes, augmenting expert work, not replacing it. It reflects a pragmatic approach to innovation at INE: solving concrete problems, testing real tools, and doing so in a way that's scalable across national contexts.

As we continue to explore AI's place in official statistics, prototypes like this remind us that progress doesn't always require revolution. Sometimes, it's about making the everyday easier, smarter, and more consistent.



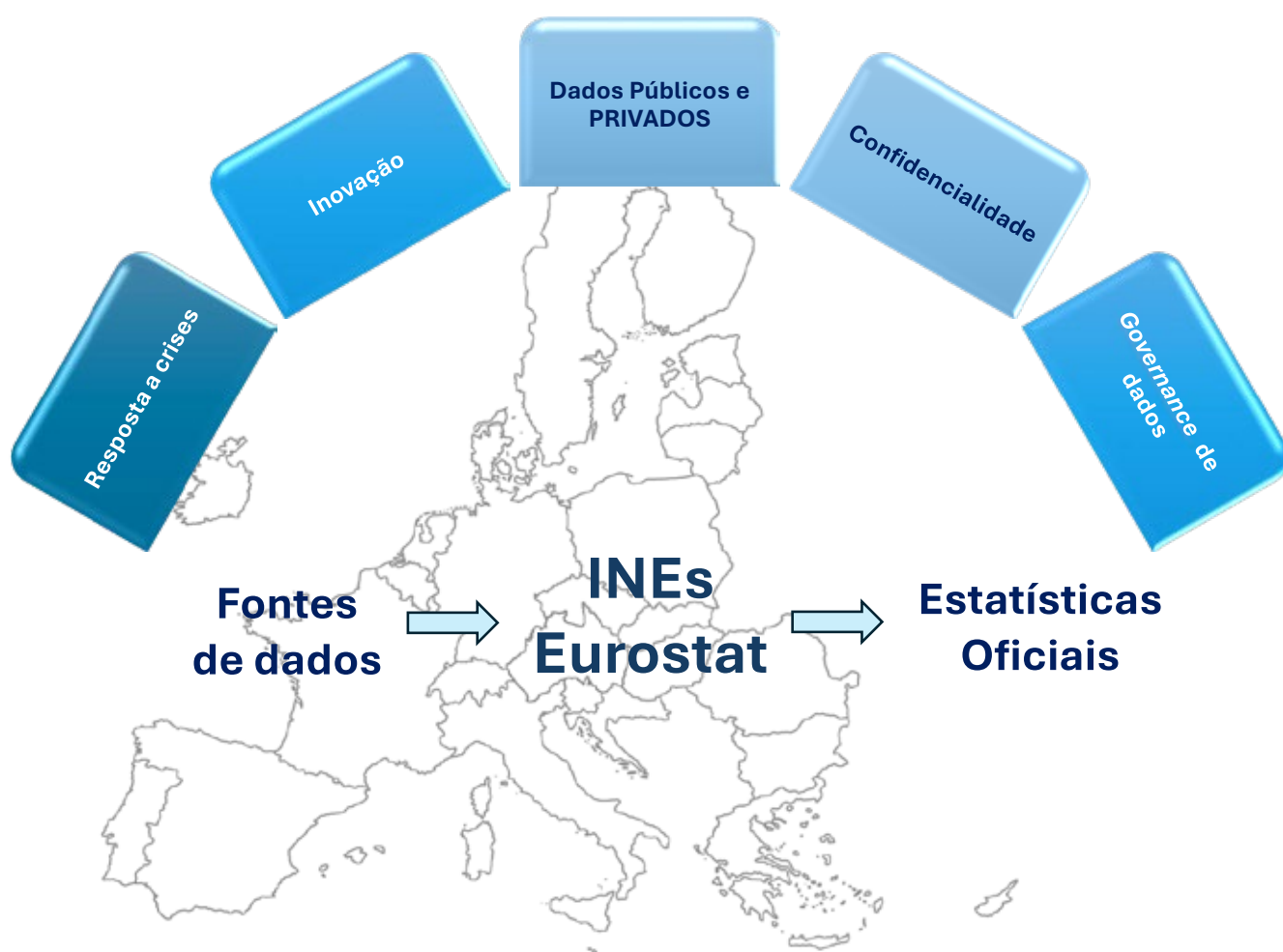
Revisão do Regulamento das Estatísticas Europeias

Desafios e oportunidades para os sistemas estatísticos

Instituto Nacional de Estatística, I.P.

O **Regulamento (CE) n.º 223/2009** do Parlamento Europeu e do Conselho, sobre as **Estatísticas Europeias**, foi alterado, pela segunda vez, pelo **Regulamento (UE) 2024/3018** do Parlamento Europeu e do Conselho, de 27 de novembro de 2024.

Esta recente revisão surge num contexto de transformação digital e de aumento da procura por estatísticas mais ágeis, pertinentes e baseadas em fontes diversificadas — um desafio e uma oportunidade clara para os sistemas estatísticos nacionais e europeu e para os profissionais da estatística.



Nova era para as Estatísticas

O que muda

- Recolha estatística em contexto de crise

O Eurostat poderá desencadear **ações estatísticas urgentes** para responder a situações de crise (como pandemias ou desastres naturais), com participação voluntária dos Estados-Membros e possibilidade de apoio financeiro. Esta capacidade de resposta rápida será essencial para garantir **estatísticas oportunas em contextos excecionais**.

- Reforço do acesso a dados administrativos

Os Institutos Nacionais de Estatística (INE) veem o seu direito reforçado de **acesso gratuito e facilitado a dados de entidades públicas**, para efeitos estatísticos. Este acesso poderá melhorar a **eficiência na produção estatística** e reduzir a carga sobre os respondentes.

- Abertura aos dados privados

O novo regulamento cria bases legais para o **acesso a dados detidos por entidades privadas**, com garantias de proporcionalidade, minimização de encargos e salvaguardas legais. Abre-se assim a porta à integração de fontes de dados alternativas, com potencial para enriquecer indicadores existentes e criar novos.

- Introdução das “estatísticas em desenvolvimento”

É reconhecida uma nova categoria de produtos: **estatísticas em desenvolvimento a nível europeu**. Estas podem utilizar métodos inovadores ou novas fontes de dados, e serão identificadas como tal, permitindo a sua **experimentação e validação científica**, antes de integrarem o corpo das estatísticas oficiais.

- Papel reforçado dos INE na governação de dados

Os INE poderão participar ativamente na **definição de normas nacionais de qualidade, interoperabilidade e partilha de dados**, assumindo um papel relevante na **infraestrutura de dados do país**.

Implicações para a comunidade estatística

Estas alterações representam uma **evolução significativa do Sistema Estatístico Europeu (SEE)** e têm impacto nos **Sistemas Estatísticos Nacionais**. A comunidade profissional é chamada a:

- Adaptar práticas e métodos à integração de novas fontes de dados (privadas, digitais, administrativas).
- Reforçar o compromisso com a **qualidade, transparência e confidencialidade**.
- Participar ativamente no desenvolvimento de **novos indicadores e metodologias**.
- Colaborar com entidades públicas e privadas para melhorar o ecossistema de dados nacional.

Confidencialidade garantida

O novo regulamento reforça a proteção dos dados individuais, assegurando que o acesso alargado a novas fontes de dados **não compromete a confidencialidade**, que permanece um pilar inegociável do trabalho estatístico.

Conclusão

A recente revisão do Regulamento (CE) n.º 223/2009 assinala uma **nova fase para as estatísticas oficiais europeias**, em que inovação, integração de dados e agilidade ganham destaque — sem abdicar do rigor científico nem da ética profissional.

Pode aceder à versão consolidada do Regulamento (CE) n.º 223/2009 relativo às estatísticas europeias na seguinte ligação:

<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A02009R0223-20241226>



Manual do Processo de Produção Estatística do Instituto Nacional de Estatística

Instituto Nacional de Estatística, I.P.

1. Introdução

O Manual do Processo de Produção Estatística (MPPE) do Instituto Nacional de Estatística (INE) é um instrumento estruturante do processo de produção das estatísticas oficiais. A sua organização assenta no modelo internacional Generic Statistical Business Process Model (GSBPM), desenvolvido pela United Nations Economic Commission for Europe (UNECE).

Sublinha-se a importância estratégica do MPPE para a garantia da qualidade do processo da produção estatística, tanto no contexto do Sistema Estatístico Nacional (SEN) como do Sistema Estatístico Europeu (SEE).

O INE enquanto autoridade estatística nacional, assume o compromisso com a produção de estatísticas oficiais de qualidade sendo o *Manual do Processo de Produção Estatística (MPPE) do INE*, na sua versão 3.1 (2025), o instrumento de referência que descreve em detalhe o processo de produção estatística desde a conceção e operacionalização à avaliação de uma operação estatística, sendo ajustável a qualquer tipo de operação estatística:

- Recenseamentos ou Inquéritos por amostragem;
- Estatísticas derivadas (obtidas por agregação e/ou integração de variáveis das estatísticas primárias e de outras fontes de informação);
- Baseadas em fontes externas, nomeadamente as administrativas.

2. O MPPE e o GSBPM

O MPPE identifica e documenta de forma sistematizada as fases, os subprocessos e as principais tarefas que caracterizam o processo de produção estatística, para qualquer operação estatística, independentemente da metodologia utilizada ou da sua complexidade, adotando como estrutura de base o modelo de referência internacional Generic Statistical Business Process Model (GSBPM v5.1) da United Nations Economic Commission for Europe (UNECE).

O alinhamento com o GSBPM é total, pelo que o MPPE está organizado em oito fases principais (*Fig. 1*). As primeiras sete sequenciais e iterativas, que refletem o ciclo de vida completo de uma operação estatística:

1. Identificação de necessidades (Specify Needs)
2. Especificações (Design)
3. Desenvolvimento (Build)
4. Recolha (Collect)
5. Processamento (Process)
6. Análise (Analyse)
7. Disseminação (Disseminate)
8. Avaliação (Evaluate)

Cada fase é composta por um conjunto de subprocessos uniformizados (que totalizam 44), que especificam de forma detalhada as atividades essenciais em cada fase do ciclo de vida de uma operação estatística.

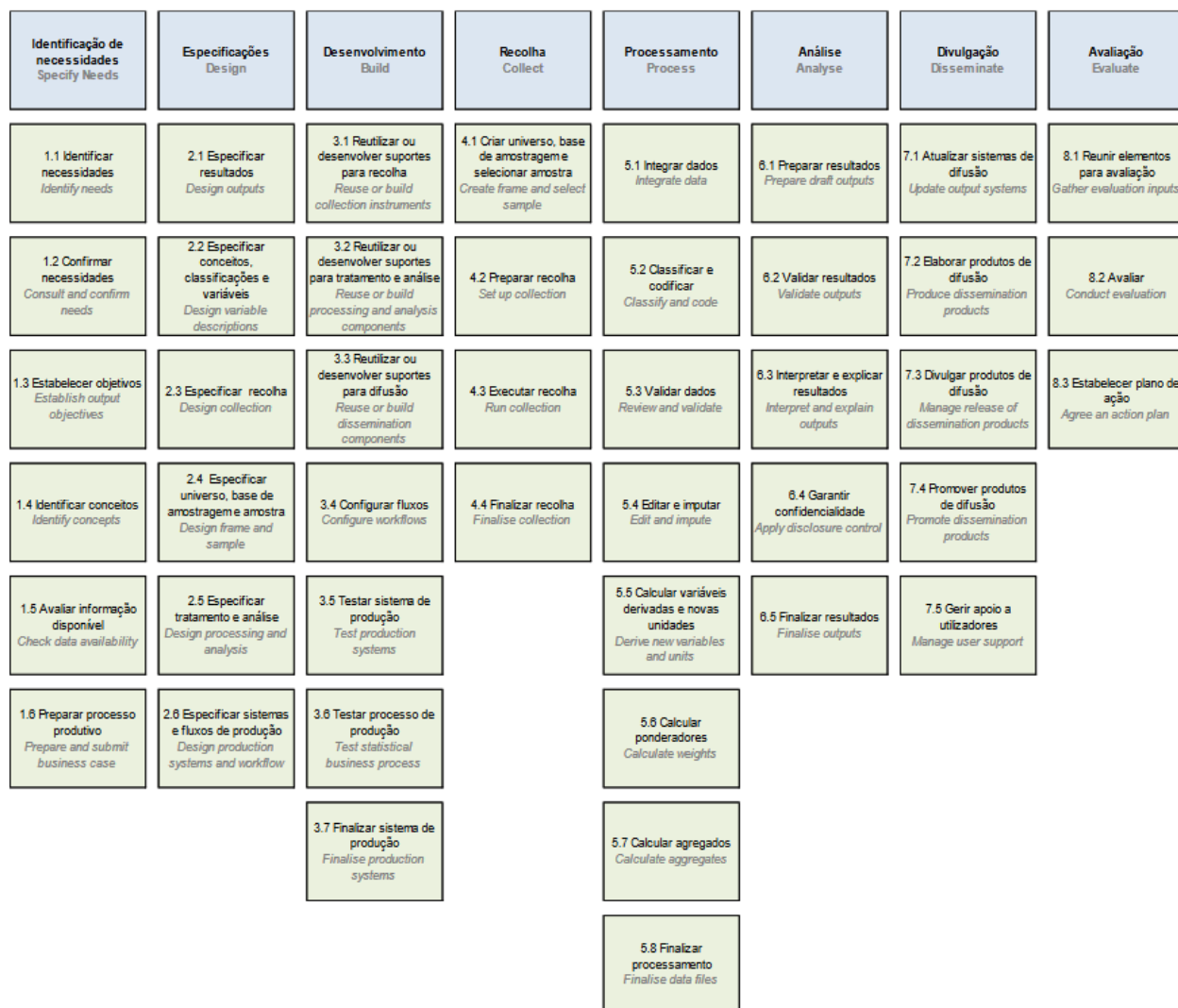
O MPPE do INE adiciona ao GSBPM, um terceiro nível de detalhe compostos por tarefas concretas que detalham a execução de cada fase/subprocesso. Estas tarefas encontram-se mapeadas numa matriz que também identifica as unidades orgânicas participantes, isto é, responsáveis e intervenientes, e os outputs documentais associados, facilitando a gestão das responsabilidades e promovendo a eficiência na execução de tarefas entre as unidades orgânicas.

Inclui também exemplos práticos sobre a sua aplicabilidade em diversos tipos de operações estatísticas, de modo a ilustrar de uma forma mais direta e clara, a sua relevância e utilização.

O MPPE incorpora ainda os princípios da gestão da qualidade, em alinhamento com a ISO 9001, para a Gestão da Qualidade, adotando o ciclo PDCA (Plan-Do-Check-Act) como abordagem base para a melhoria contínua do processo estatístico e assegura a consistência entre planeamento, execução e avaliação de uma operação estatística.

Neste sentido, o MPPE é um pilar do Sistema de Gestão Integrado (SGI) do INE (que integra a Gestão da Qualidade e a Segurança de Informação), sendo o processo “Produção Estatística” deste sistema inteiramente suportado pelo MPPE que assegura conceptualmente a conformidade, a rastreabilidade e a melhoria dos processos.

Fig. 1 - Fases e Subprocessos do Processo de Produção Estatística, INE, MPPE versão 3.1 em 2025



O MPPE, como documento de referência, constitui um garante da coerência metodológica e da comparabilidade internacional do processo de produção estatística, encontrando-se alinhado com os princípios do Código de Conduta para as Estatísticas Europeias (European Statistics Code of Practice) e com o disposto no Regulamento (CE) n.º 223/2009 do Parlamento Europeu e do Conselho, na sua redação atual, relativo às estatísticas europeias - o qual estabelece o quadro jurídico para o desenvolvimento, produção e disseminação de estatísticas europeias, assegurando a sua qualidade, independência e fiabilidade.

O Código de Conduta para as Estatísticas Europeias, adotado pelo Comité do Sistema Estatístico Europeu, estrutura-se em 16 princípios que abrangem o quadro institucional, os processos e os produtos estatísticos. O MPPE constitui um instrumento que dá suporte e operacionaliza vários desses princípios, como por exemplo:

- **Princípio 4 – Compromisso com a qualidade:** o MPPE promove a identificação sistemática de pontos fortes e fracos e a melhoria contínua do processo de produção estatística; a documentação das fases do processo de produção estatística garante a reprodutibilidade e controlo de qualidade

dos procedimentos adotados; A estruturação do processo estatístico e a sua formalização no MPPE contribuem para a clareza das operações e a comunicação eficaz com utilizadores especializados.

- **Princípio 7 – Metodologia sólida:** ao basear-se no GSBPM o MPPE assegura que o processo estatístico segue padrões harmonizados e reconhecidos.
- **Princípio 10 – Eficácia na utilização dos recursos:** a documentação exaustiva das fases do processo de produção estatística garante a reprodutibilidade e controlo de qualidade dos procedimentos adotados.

3. Formação e desenvolvimento de competências

O MPPE é também um instrumento importante na formação de novos técnicos, sendo utilizado, na capacitação inicial através do Curso de formação específico para a carreira de técnico superior especialista em estatística do INE; e para estruturar o conhecimento institucional e garantir a aplicação homogénea de boas práticas.

4. Conclusão

O MPPE é um documento estruturante no INE, que assegura e promove:

- a existência de uma linguagem comum para stakeholders, assim como para os intervenientes no processo estatístico, facilitando a comunicação, a compreensão mútua e a articulação eficaz entre as diferentes partes envolvidas;
- o alinhamento com os padrões do Sistema Estatístico Europeu;
- a sistematização e documentação dos procedimentos do INE;
- a otimização da eficiência, da rastreabilidade e da qualidade do processo estatístico;
- a promoção de sinergias criadas no decurso do processo produtivo, ao nível das várias tarefas e das respetivas responsabilidades.

O seu papel estruturante no SGI assegura que a missão do INE – produzir estatísticas de qualidade, independentes e relevantes para a sociedade – se concretiza com rigor, transparência e valor científico. O seu alinhamento com o Código de Conduta para as Estatísticas Europeias reforça a confiança dos utilizadores de dados oficiais, servindo de base sólida para a retenção do conhecimento. Reveste-se também de adaptabilidade e a relevância contínua, acompanhando a evolução do GSBPM e ainda as alterações naturais do processo produtivo.

5. Trabalhos futuros

Em junho de 2025, foi publicada a nova versão do modelo Generic Statistical Business Process Model (GSBPM v5.2), que introduz atualizações significativas para refletir a evolução dos processos estatísticos, nomeadamente:

- A crescente utilização de novas fontes de dados;
- A digitalização dos processos estatísticos;
- A interligação mais profunda entre os modelos GSBPM, GAMS (Generic Activity Model for Statistical Organisation) e GSIM (Generic Statistical Information Model).

Entre as principais alterações introduzidas nesta versão, destacam-se:

- A introdução do conceito de "processos integrados de dados", refletindo a necessidade de tratamento coordenado e eficiente de múltiplas fontes;
- O reforço da ligação ao ciclo de vida da informação, promovendo a coerência entre os processos estatísticos e a gestão da informação;
- A revisão terminológica e redefinição de vários subprocessos, ajustando-os à realidade da ciência de dados e à digitalização dos fluxos de trabalho;
- A integração explícita de abordagens baseadas em inteligência artificial, machine learning, sensores e web scraping, acompanhando os avanços tecnológicos e metodológicos.

Estas inovações darão origem a uma revisão do MPPE em vigor à data no INE.



ALEA – Um marco na promoção da literacia estatística: uma visão em 3 partes

O ALEA e a produção de Estatísticas Oficiais

Pedro Campos, *pedro.campos@ine.pt*

Instituto Nacional de Estatística, I.P.

A missão do Instituto Nacional de Estatística (INE) não se esgota na produção de informação estatística oficial. Inclui também a sua divulgação de forma clara, acessível e útil para a sociedade. Promover a compreensão dos números que descrevem o país é, de facto, um passo essencial para aproximar os cidadãos da realidade estatística e, assim, reforçar a literacia estatística.

Foi precisamente neste espírito que nasceu o ALEA (Ação Local de Estatística Aplicada). A génese deste projeto remonta a uma visita de dois professores da Escola Secundária Tomaz Pelayo (ESTP) ao INE, no final da década de 1990. O encontro gerou uma ideia simples, mas poderosa: criar um espaço que unisse a experiência pedagógica da escola e o conhecimento técnico do INE, colocando a Estatística ao serviço da aprendizagem e da cidadania.

Dessa colaboração pioneira, que logo a seguir envolveu o Ministério da Educação, atualmente representado através da Direção-Geral dos Estabelecimentos Escolares (DGEstE), surgiu uma plataforma que, ao longo dos anos, se consolidou como referência em Portugal e a nível internacional. O ALEA não é apenas um repositório de conteúdos e recursos educativos, mas também um espaço de partilha entre professores, alunos, investigadores e profissionais de estatística. Com jogos, exemplos interativos, glossários e materiais pedagógicos, tornou-se num verdadeiro laboratório de exploração estatística, onde o rigor técnico do INE encontra a criatividade do ensino.

Ao apoiar e dinamizar o ALEA, o INE cumpre uma dimensão essencial da sua missão: a **promoção da literacia estatística**. Mais do que produzir dados, trata-se de garantir que esses dados são compreendidos, questionados e utilizados de forma informada. O ALEA é, assim, a expressão prática de como a Estatística Oficial se aproxima das escolas, das famílias e da sociedade civil, promovendo cidadãos mais críticos e conscientes no uso da informação.

O ALEA, desde o seu início, tem Consultoria Científica da Prof. Eugénia Graça Martins.

O ALEA na Educação para a Matemática

Emília Oliveira, *ecmo.estp@gmail.com*

Escola Secundária de Tomaz Pelayo

A Escola Secundária Tomaz Pelayo (ESTP) é coautora do ALEA. Desde a sua génese, a ESTP colabora na criação e atualização de conteúdos. O ALEA foi dos primeiros sítios portugueses a disponibilizar conteúdos e dados de Estatística, de acordo com o currículo nacional, e este propósito mantém-se. Com a revisão curricular das Aprendizagens Essenciais (AE) de Matemática para o Ensino Básico e para o

Ensino Secundário, atualmente em vigor, vários conteúdos foram já adaptados/atualizados e novos recursos foram criados. Por exemplo, várias ActivAlea's - fichas com atividades, exercícios, comentários e propostas destinadas a sistematizar a formação básica de estatística, foram renovadas (ficha n.º 31 "Organização de dados quantitativos utilizando Tabelas dinâmicas no Excel", ficha n.º 21 "Presidentes da República Portuguesa", ...).

Entre os recursos criados mais recentemente, destacam-se as "TarefALEA's" de Estatística e de Probabilidade. Este novo espaço do ALEA disponibiliza algumas sugestões de tarefas de Estatística, com as quais se pretende responder a algumas questões estatísticas, utilizando dados oficiais produzidos pelo INE, Eurostat, Banco de Portugal, entre outras entidades, e sistematizados pela PORDATA. Todas as tarefas disponibilizadas têm uma estrutura comum. Inicialmente, são sugeridos/apresentados os passos essenciais para a organização e o tratamento dos dados, entretanto recolhidos, seguindo-se um conjunto de atividades e questões adaptadas ao nível de ensino ou ao ano de escolaridade dos alunos.

Com os dados da temática que se pretende abordar com os alunos, trabalham-se conceitos (unidade de observação, variável estatística, categorias da variável, tipo de dados, ...), trabalha-se a organização e representação de dados, trabalha-se o cálculo das medidas descritivas e também a interpretação dessas medidas na caracterização da distribuição dos dados. As tarefas de Probabilidade seguem um princípio semelhante: procurar responder a questões ou desafios, em que se pretende desenvolver, além do conceito clássico - ou de Laplace - de Probabilidade, que só se aplica em situações "raras" de equilíbrio dos acontecimentos elementares, a noção intuitiva de "Probabilidade de um acontecimento", tendo em conta a regularidade estatística observada numa série de repetições do fenómeno aleatório, que conduza à realização desse acontecimento. Uma das ações que continua a contar com a participação de alunos do ensino básico e secundário de várias escolas do país são os Desafios do ALEA - problemas sobre assuntos do quotidiano, baseados em notícias publicadas em órgãos de comunicação social. Tem sido um dos recursos que consubstancia um desígnio do ALEA: a promoção da literacia estatística em Portugal.

O ALEA e a Educação em Portugal

João Gonçalves, joao.goncalves@dgeste.mec.pt;
Franck de Almeida, franck.almeida@dgeste.mec.pt

Direção-Geral dos Estabelecimentos Escolares

Em 2024, celebraram-se os 25 anos do ALEA, projeto que marcou profundamente a educação estatística em Portugal. Esta efeméride constitui também uma oportunidade para destacar a colaboração contínua entre o ALEA — inicialmente através da Direção Regional de Educação do Norte (DREN) — e, atualmente, com a Direção-Geral dos Estabelecimentos Escolares (DGEstE).

Desde o seu início, o ALEA afirmou-se como uma plataforma inovadora, promovendo a literacia estatística junto de alunos e professores, através de recursos acessíveis, atividades interativas e desafios pedagógicos. A DREN foi uma das primeiras entidades a reconhecer o potencial transformador deste projeto, apoiando a sua implementação nas escolas da região Norte e incentivando a sua integração nas práticas letivas.

Com a reorganização dos serviços do Ministério da Educação e a transição para a DGEstE, esta parceria não só se manteve como se fortaleceu. A DGEstE, em articulação com as suas Direções de Serviços Regionais, tem continuado a colaborar com o ALEA na dinamização de concursos escolares, na

formação contínua de docentes e na produção de materiais didáticos que aproximam a estatística do cotidiano dos alunos.

Ao longo destes 25 anos, milhares de estudantes participaram em atividades promovidas pelo ALEA, desenvolvendo competências essenciais para a cidadania informada: interpretar gráficos, analisar dados, questionar conclusões. Estas competências são hoje mais relevantes do que nunca, num mundo onde a informação circula em grande volume e velocidade e onde a capacidade de pensar criticamente é um pilar da democracia.

A DGEstE orgulha-se de ter contribuído para este percurso de excelência. Através do apoio do Ministério da Educação, Ciência e Inovação, da mobilização das escolas e da valorização do trabalho dos professores, tem sido possível consolidar uma cultura de literacia estatística que ultrapassa os limites da sala de aula.

Celebrar os 25 anos do ALEA é, por isso, celebrar também o compromisso da escola pública com uma educação de qualidade, inovadora e centrada no desenvolvimento integral dos alunos. É reconhecer o valor da colaboração entre instituições, da continuidade das boas práticas e da aposta na formação de cidadãos mais preparados para os desafios do presente e do futuro.

Que esta parceria continue a inspirar novas gerações e a reforçar o papel da estatística como ferramenta de compreensão do mundo.



Capacitação em Ciência de Dados para o Sistema Estatístico Nacional e Administração Pública

Pedro Campos, *pedro.campos@ine.pt*
Jorge Magalhães, *jorge.magalhaes@ine.pt*

Instituto Nacional de Estatística, I.P.

Introdução

Num contexto em que os dados assumem um papel central na formulação de políticas públicas, na gestão da administração e na tomada de decisão informada, o Instituto Nacional de Estatística (INE) tem vindo a investir na formação para a capacitação em Ciência de Dados dirigida ao Sistema Estatístico Nacional (SEN) e à Administração Pública.

Esta iniciativa contempla um conjunto de cursos com apoio do Plano de Recuperação e Resiliência (PRR) e tem como objetivo dotar técnicos e dirigentes de competências práticas e conceptuais que lhes permitam integrar novas fontes de dados, aplicar metodologias inovadoras de análise e explorar ferramentas avançadas de ciência de dados. São abordados temas que vão desde os fundamentos de programação em R e Python, à visualização e comunicação de resultados em Sistemas de Informação Geográfica, ou em *dashboards*, passando por técnicas de *machine learning*, inteligência artificial e análise de redes.

Mais do que um programa de formação, trata-se de uma estratégia estruturante que reforça a modernização e inovação do SEN, promovendo a qualidade, a eficiência e a relevância das Estatísticas Oficiais. Ao mesmo tempo, esta aposta na capacitação em ciência dos dados contribui para aproximar a administração de práticas baseadas em evidência, potenciando a utilização de dados no desenho e monitorização de políticas públicas.

Com esta aposta, que envolve de forma intensa o departamento de Recursos Humanos, o INE reafirma o seu compromisso em estar na linha da frente da transformação digital e científica da produção estatística, apoiando uma Administração mais atual e informada, ágil e preparada com os recursos necessários para os desafios da sociedade contemporânea.

Capacitação em Ciência de Dados: Áreas e Cursos de Formação

O plano de formação em Ciência de Dados promovido pelo INE para o Sistema Estatístico Nacional e a Administração Pública está organizado em sete áreas estratégicas, que refletem as diferentes dimensões da produção, gestão e utilização de dados. A cada área estão associados cursos específicos, concebidos para responder às necessidades de iniciação, aprofundamento ou especialização dos participantes.

A – Ciência de Dados para a Gestão

Formações que introduzem o papel da ciência de dados nas organizações, abordando dimensões estratégicas, competências técnicas, recursos tecnológicos e exemplos de aplicabilidade.

B – Matemática e Estatística

Cursos de introdução ou aprofundamento em estatística descritiva, probabilidade, amostragem, testes de hipóteses, modelos estatísticos e *machine learning*.

C – Bases de Dados

Formações dedicadas a bases de dados estruturadas e não estruturadas, modelização, análise e integração de dados, com foco em SQL e DDL.

D – Ferramentas de Análise e Manipulação de Dados

Capacitação prática em linguagens e softwares amplamente usados em estatística oficial e ciência de dados.

E – Recolha, Tratamento e Integração de Dados

Formações sobre técnicas de preparação, modelização, correção e imputação de dados, assegurando qualidade e integridade da informação (*cursos ainda em desenvolvimento para esta área*).

F – Comunicação e Visualização

Cursos que exploram metodologias de visualização gráfica e tabular, *dashboards*, infografias e *reporting*, para uma comunicação eficaz de resultados.

G – Segurança da Informação e Proteção de Dados

Formações introdutórias às problemáticas, políticas e boas práticas de segurança da informação e proteção de dados pessoais.

Ações de formação

A oferta formativa cobre de forma integrada desde os aspetos fundamentais da Estatística até às competências técnicas avançadas. Na área da Ciência de Dados para a gestão, destaca-se o *Workshop em Ciência de Dados para o Top Management*, dirigido a dirigentes e decisores. Já a área de Matemática e Estatística apresenta uma ampla gama de cursos, incluindo análise multivariada de dados, estatística descritiva e multivariada, desenho amostral, econometria e modelos de previsão, bem como tópicos mais avançados como *AI for Business* e *Network Analytics and Deep Learning*. Na área de Bases de Dados, os cursos vão do nível iniciante ao avançado em bases relacionais e SQL, assegurando uma base sólida para trabalhar com dados estruturados.

Complementando estas dimensões, a área de Ferramentas de análise e manipulação de dados aprofunda competências práticas em linguagens e softwares fundamentais: R, Python (em três níveis de progressão), STATA e Git. A componente de Comunicação e Visualização foca-se em ferramentas de *business intelligence* (Business Objects R4 e Power BI), mapas temáticos e análise espacial, assegurando que os resultados são apresentados de forma clara e eficaz. Finalmente, a área de Segurança da Informação e Proteção de Dados promove a consciencialização e boas práticas em cibersegurança, gestão da informação e privacidade, reforçando a confiança e o cumprimento das obrigações legais. No seu conjunto, esta oferta traduz-se num percurso de capacitação abrangente e alinhado com as necessidades atuais das Estatísticas Oficiais e da Administração Pública.

A - Ciência de Dados para a gestão			D - Ferramentas de análise e manipulação de dados		
	A1	Workshop em Ciência de Dados para o <i>Top Management</i> (curso de meio dia)		D1	R para Ciência de Dados I
B - Matemática e Estatística				D2	R para Ciência de Dados II
	B1	Análise de Dados com SPSS e R		D3	Python para estatísticas oficiais - Basic Data Science
	B3	AI for Business		D4	Python para estatísticas oficiais - Intermediate Data Science
	B4	Network Analytics and Deep Learning		D5	Python para estatísticas oficiais - Advanced Data Science
	B9	Métodos analíticos e modelos de Previsão (PRR)		D11	Programação em STATA (PRR)
	B10	Estatística Multivariada (PRR)		D10	Introdução ao Git
	B11	Introdução à Econometria (PRR)	F - Comunicação e visualização		
	B12	Econometria Aplicada (PRR)		F1	Business Intelligence - Business Objects R4
	B13	Desenho Amostral		F2	Business Intelligence - Power BI
	B8	Curso de Estatística Descritiva		F4	Mapas Temáticos e Visualização Espacial
C - Bases de Dados				F5	Análise Espacial
	C1	Base de Dados Beginner	G - Segurança da Informação e proteção de dados		
	C2	SQL Beginner		G1	Segurança da Informação: Boas práticas
	C3	Base de Dados Expert		G2	Gestão de Segurança da Informação
	C2	SQL - Expert		G3	Segurança da Informação e a Privacidade

A realização nos dois primeiros anos: 2023 e 2024

Em 2023, realizaram-se 19 ações de formação, das quais 17 internas e 2 externas, envolvendo 226 formandos (140 internos e 86 externos), num total de 4.593 horas de formação. Participaram 173 indivíduos, dos quais 60% homens e 40% mulheres. Além do INE participaram entidades da Administração Pública, nomeadamente órgãos delegados do INE no âmbito do Sistema Estatístico Nacional.

Em 2024, realizaram-se 25 ações de formação, das quais 24 internas e 1 externa (ministrada pela Universidade Católica Portuguesa), envolvendo 315 formandos (278 internos e 37 externos), num total de 6.467 horas de formação. Participaram 212 indivíduos, dos quais 38% homens e 62% mulheres.

Estes resultados confirmam a consolidação da estratégia de capacitação em ciência de dados do INE, não só reforçando competências internas, mas também promovendo o envolvimento de parceiros externos com o objetivo de conduzir uma estratégia integrada que modernize e inove o SEN, elevando a qualidade, a eficiência e a relevância das Estatísticas Oficiais em Portugal.



• Artigos em Revistas

- Brilhante, M.F., Gomes, M.I., Mendonça, S., Pestana, D. and Santos, R. (2025) - Meta-Analysis of Genuine and Fake p-Values. *J. Statistical Theory and Practice* **19**, 29.
<https://doi.org/10.1007/s42519-025-00445-3>
- Brites, Nuno M.; Braumann, Carlos A. (2025) - Moments and probability density of threshold crossing times for populations in random environments under sustainable harvesting policies. *Computational Statistics* **40** (6), 3191–3203.
<https://doi.org/10.1007/s00180-022-01237-0>
- Miranda, C., Souto de Miranda M. & Gomes, M.I. (2025) - A new proposal for robust estimation of the extremal index. *J. Statist. Comput. and Simul.* **95**:5, 949-966.
<https://doi.org/10.1080/00949655.2023.2300700>

• Capítulos Livros

- Título: A control chart for zero-inflated semi-continuous data.
Autores: Figueiredo, F., Figueiredo, A. e Gomes, M.I.
Ano: 2025
Dimotikalis, Y. and Skiadas, C.H. (Eds.) - *Data Analysis and Related Applications 5: Models, Methods and Techniques, Volume 13*, Chapter 5, pp.65-77, ISTE, Wiley.
<https://doi.org/10.1002/9781394401604.ch5>
- Título: Further Results on Location Invariant Estimation of the Weibull Tail Coefficient.
Autores: Gomes, M.I., Caeiro, F. e Henriques-Rodrigues, L.
Ano: 2025
Dimotikalis, Y. and Skiadas, C.H. (Eds.), *Data Analysis and Related Applications 5: Models, Methods and Techniques, Volume 13*, Chapter 6, pp.79-94, ISTE, Wiley.
<https://doi.org/10.1002/9781394401604.ch6>
- Título: Delta method estimation for SDE mixed models: Sensitivity to random parameter distribution.
Autores: Jacinto, Gonçalo; Filipe, Patrícia A.; Braumann, Carlos A.
Ano: 2025
Advances in Mathematical Modeling in Science, Engineering and Social Sciences (Editors: Bezzeghoud, M., Carapau, F., Minhós, F., Vaidya, A.), Springer. Aceite para publicação.
- Título: Optimal harvesting in randomly varying environments: Sensitivity of profit and effort to population and economic parameters.
Autores: Brites, Nuno M.; Braumann, Carlos A.
Ano: 2025
Advances in Mathematical Modeling in Science, Engineering and Social Sciences (Editors: Bezzeghoud, M., Carapau, F., Minhós, F., Vaidya, A.), Springer. Aceite para publicação.
- Título: The Role of Resampling Methods and Generalised Means in Statistical Extreme Value Theory (EVT).
Autora: Gomes, M.I.
Ano: 2025
Brazil-Portugal Mathematics, (Editors: M. Abreu, H. Bursztyn & J. Picado), Coimbra Mathematical Texts 4, ISBN 978-3-031-98454-9, Springer Nature, Switzerland, pp. 135-157.
https://doi.org/10.1007/978-3-031-98455-6_6

• Livros

Título: *Python e Exploração de Dados*

Autor: Jorge Caiado

Ano: 2024

Editora: Edições Sílabo. Lisboa

ISBN: 978-989-561-379-3.

Título: *Inteligência Artificial: Guia para Iniciantes*

Autor: Jorge Caiado

Ano: 2025

Editora: Edições Sílabo. Lisboa

ISBN: 978-989-561-414-1.

Título: *XXVII Congresso SPE25 / Sociedade Portuguesa de Estatística – Programa e Resumos*

Editores: Carlos Sousa, Clara Cordeiro, Conceição Ribeiro, M. Helena Gonçalves, Nelson Antunes e
Sílvia Pedro Rebouças

Ano: 2025

Editora: Edições SPE – Sociedade Portuguesa de Estatística

ISBN: 978-972-8890-52-0.

Depósito Legal: 555257/25

Título: *O guia do viajante para uma aprendizagem de máquina responsável*
Versão para o programa R.

Autores: Przemyslaw Biecek, Anna Kozak e Aleksander Zawada

Ano: 2025

Editora: Edições SPE – Sociedade Portuguesa de Estatística

ISBN: 978-972-8890-51-3.

Depósito Legal: 554460/25

Livro, na versão inglesa, disponível em: <https://betaandbit.github.io/RML>



• Prémios Estatístico Júnior



Estes são os resultados dos **Prémios Estatístico Júnior 2025**.

Na categoria de Ensino Secundário, os trabalhos distinguidos foram:

1. *EstatíNews*, desenvolvido pelos alunos António Gonçalves, Leonor Aguiar e Mariana Abilheira, sob orientação do professor Eduardo Cunha, do Agrupamento de Escolas de Barcelos;
2. *DataNews*, desenvolvido pelos alunos Eva Eira, Filipa Gomes e Valeria Giraldes, sob orientação do professor Eduardo Cunha, do Agrupamento de Escolas de Barcelos;
3. *MathsJournal*, desenvolvido pelos alunos José Alves, Miguel Pereira e Miguel Lopes, também sob orientação do professor Eduardo Cunha, do Agrupamento de Escolas de Barcelos.

A Direção da SPE, congratulou-se com a qualidade, criatividade e rigor estatístico demonstrado nos trabalhos apresentados, reconhecendo o seu contributo relevante para a promoção da literacia estatística e para o estímulo da utilização crítica de dados reais no contexto educativo.

Este ano, o número de candidaturas duplicou em relação à edição anterior, sinal do crescente interesse por esta iniciativa junto das escolas e dos professores.



Sobre a teoria de modelos espaciotemporais para séries temporais de contagens

Ana Martins, a.r.martins@ua.pt

*Departamento de Eletrónica, Telecomunicações e Informática,
Instituto de Engenharia Eletrónica e Informática de Aveiro,
Universidade de Aveiro*

Os modelos estatísticos para séries temporais de contagens, isto é, valores inteiros não negativos, têm particular relevância quando as séries temporais observadas têm contagens baixas, o que ocorre em diversas áreas, desde a saúde às finanças. Dado que os modelos contínuos assentam no pressuposto da distribuição Normal e, sendo esta uma distribuição simétrica não é capaz de acomodar o comportamento das contagens baixas, que têm uma distribuição assimétrica.

A investigação em modelos de séries temporais para contagens tomou impulso no início da década de 80 com a introdução do operador de filtragem binomial (em inglês, binomial thinning operator) pelos autores Steutel & Van Harn (1979). Este operador tornou-se bastante popular e afirmou-se como uma alternativa à multiplicação nos modelos autorregressivos (AR), garantindo assim que, juntamente com um processo de inovação discreto, o processo seria de valor inteiro – daí a sua designação de modelos inteiros autorregressivos (INAR). Assim, iniciou-se o desenvolvimento de uma classe muito rica, pois estes modelos, ao contrário dos ARMA contínuos, não têm uma especificação única. Esta classe são os modelos inteiros autorregressivos e de médias móveis – INARMA.

No contexto univariado há já variadas contribuições, mas o desenvolvimento de modelos multivariados tem sido mais parco, fruto dos desafios de desenvolvimento teórico, mas também de questões práticas de implementação de estratégias de estimação eficientes. Assim, este trabalho contribuiu para o desenvolvimento de modelos multivariados, em específico, modelos que também contemplam a dimensão espacial. Esta nova classe denomina-se por modelos inteiros espaciotemporais ARMA (STINARMA) e é inspirada nos modelos contínuos STARMA introduzidos por Pfeifer & Deutsch (1980). Nestes modelos, a informação espaciotemporal é introduzida no modelo através de uma matriz de vizinhança que é incorporada no operador de filtragem binomial.

Os modelos STINARMA (Martins et. al, 2023) foram estudados detalhadamente nas suas componentes puramente de médias móveis (STINMA) e puramente autorregressiva (STINAR). Posteriormente, resultados para a classe completa combinando componentes AR e MA, foram obtidos. Para cada um destes modelos os momentos até à segunda ordem, incluindo as funções de autocovariância e autocorrelação espaciotemporais, bem como propriedades probabilísticas foram estudadas. A estimação de parâmetros foi também abordada neste trabalho, nomeadamente através do Método dos Momentos, dos Mínimos Quadrados Condicionais e da Máxima Verosimilhança Condicional. O desempenho dos métodos de estimação em amostras finitas foi avaliado através de estudos de simulação. Finalmente, estes modelos foram utilizados para estudar a associação entre contagens diárias de admissões

hospitalares respiratórias e fatores ambientais no distrito de Aveiro, recorrendo a modelos STINAR não estacionários com parâmetros variáveis no tempo.

Referências

- Martins A., Scotto, M. G., Weiß, C. H & Gouveia S. (2023) Space-time integer-valued ARMA modelling for time series of counts, *Electronic Journal of Statistics*, 17(2), 3472-3511.
- Pfeifer P. E., Deutsch S. J. (1980). A Three-Stage Iterative Procedure for Space-Time Modeling, *Technometrics*, 22 (1), 35-47.
- Steutel, F. W., Van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability, *The Annals of Probability*, 7 (5), 893-899.

Breve CV

Ana Martins, galardoada com o Prémio SPE 2024, é Mestre em Matemática e Aplicações e Doutora em Matemática Aplicada pela Universidade de Aveiro. Contudo, o seu percurso académico iniciou-se nas ciências da vida e da saúde – Licenciatura em Bioquímica e Mestre em Saúde Pública, pela Universidade do Porto. Atualmente desempenha funções de investigadora no Instituto de Engenharia Eletrónica e Informática de Aveiro. O prémio SPE 2024 foi atribuído no âmbito do seu trabalho de doutoramento (SFRH/BD/143973/2019) desenvolvido sob a orientação de Sónia Gouveia (Universidade de Aveiro), Manuel G. Scotto (Instituto Superior Técnico) e Christian Weiss (Helmut Schmidt University).

.





Neural Bayes estimation and selection for complex bivariate extremal dependence models

Lídia André, lidiamandre@gmail.com

School of Mathematical Sciences, Lancaster University

Simular um conjunto de dados a partir de um modelo é geralmente fácil e computacionalmente eficiente, mesmo quando a função de verosimilhança não pode ser formulada ou quando o seu ajuste é computacionalmente caro. Este é o caso de vários modelos que estão disponíveis na literatura de extremos multivariados, em particular modelos de caudas mais flexíveis. Exemplos destes modelos incluem os que permitem alternar entre duas classes importantes de dependência: asymptotic dependence e asymptotic independence (e.g., Wadsworth et al., 2017, Huser and Wadsworth, 2019 e Engelke et al., 2019); ou modelos capazes de ajustar juntamente a região central e caudal de dados multivariados sem necessidade de definir um vetor de thresholds (e.g., André et al., 2024). Nestes modelos, a função de verosimilhança é conhecida; contudo, computacionalmente custosa visto que é composta por funções que são obtidas através de integração numérica e funções inversas.

Para ajudar na implementação deste tipo de modelos complexos, investigamos abordagens que não dependem da função de verosimilhança, mas sim que utilizam redes neuronais para realizar a inferência dos parâmetros dos modelos. Para isso, exploramos as propriedades dos estimadores neuronais Bayesianos (NBEs do inglês neural Bayes estimators) propostos por Sainsbury-Dale et al. (2024) para a inferência paramétrica dos vários modelos acima mencionados.

Devido à ausência do ajuste da função de verosimilhança no processo de inferência, os critérios de informação clássicos, como o AIC e o BIC, não podem ser utilizados na seleção do modelo mais apropriado para ajustar os dados em questão. Deste modo, propomos também usar redes neuronais como classificadores neuronais Bayesianos (NBCs do inglês neural Bayes classifiers) para a seleção de modelos.

Disponibilizamos, assim, um conjunto de ferramentas para a seleção e inferência de modelos, que não depende da função de verosimilhança e é amortizada (i.e., permite a sua utilização repetida sem custo computacional adicional). Este conjunto de ferramentas permite então que o modelo mais adequado seja selecionado usando o NBC, e subsequentemente os seus parâmetros são estimados usando o NBE.

Através de estudos de simulação extensivos, mostramos que os NBEs têm um desempenho global bastante bom, embora exibam, nalguns casos, mais bias do que os estimadores de máxima verosimilhança (MLEs do inglês maximum likelihood estimators); no entanto, os NBEs também apresentam, por vezes, menor variabilidade. Além disso, ao estimar quantidades de cauda de interesse, este bias não se mostrou problemático, uma vez que a estrutura geral da dependência extremal parece ser bem capturada pelos NBEs treinados. Quando comparados, o NBC e o BIC mostraram um desempenho semelhante, tanto na escolha do modelo mais adequado entre dois como quatro modelos.

Aplicamos os NBCs e NBEs no estudo do comportamento extremal das alterações das flutuações do campo geomagnético horizontal em três localizações distintas, e mostramos que a nossa metodologia é capaz de obter estimações rápidas e fiáveis.

Este trabalho está disponível em André et al. (2025) e encontra-se em fase de aceitação pela revista *Extremes*.

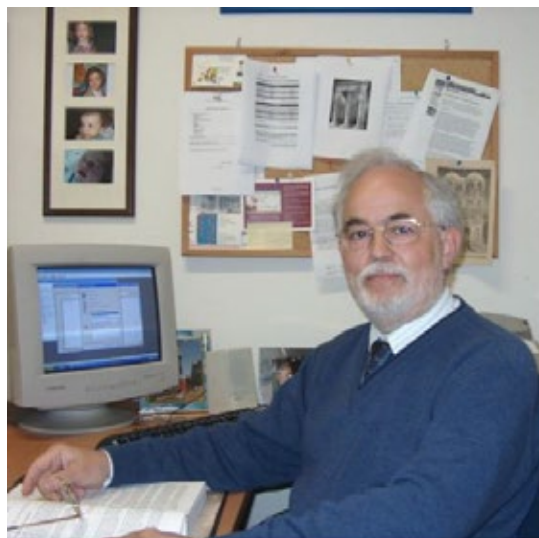
Referências

- André, L., Wadsworth, J., and O'Hagan, A. (2024). Joint modelling of the body and tail of bivariate data. *Computational Statistics and Data Analysis*, 189:107841.
- André, L. M., Wadsworth, J. L., and Huser, R. (2025). Neural Bayes estimation and selection for complex bivariate extremal dependence models. *arXiv preprint arXiv:2503.23156*.
- Engelke, S., Opitz, T., and Wadsworth, J. (2019). Extremal dependence of random scale constructions. *Extremes*, 22:623–666.
- Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444.
- Sainsbury-Dale, M., Zammit-Mangion, A., and Huser, R. (2024). Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician*, 78(1):1–14.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B*, 79:149–175.

Breve CV

Lídia André é Licenciada em Matemática Aplicada à Economia e Gestão pelo ISEG, Mestre em Estatística e Investigação Operacional pela Faculdade de Ciências da Universidade de Lisboa, e Doutora em Estatística pela Universidade de Lancaster, Reino Unido. Entre o mestrado e o doutoramento foi investigadora bolsista pelo IPMA e a Universidade NOVA de Lisboa. Após o doutoramento e até setembro de 2025, foi investigadora pós-doutoral na Universidade de Namur, Bélgica.





O Prémio Carreira SPE foi instituído pela Direção da Sociedade Portuguesa de Estatística em 2013. Na altura foi uma surpresa, para além da Direção apenas os organizadores do Congresso da SPE, em Aveiro, estavam a par do segredo e foram encarregados de recorrer aos malabarismos que fossem necessários para garantir que os premiados iriam ao Congresso. O primeiro contemplado foi, naturalmente, o Professor Bento Murteira, uma personalidade ímpar, que partilhou generosamente o seu conhecimento enciclopédico da Estatística com todos nós, cuidando sempre de inculcar a visão de que a Estatística não é apenas ciência e técnica, tem também vertentes filosóficas que apoiam a transformação de informação em conhecimento. Esta honra conferida pela SPE foi realçada no ano passado, comemorando a sua obra como Estatístico Centenário, como mereceria ser cognominado.

Podem obter-se informações mais detalhadas e o regulamento, na página da SPE, nomeadamente em [Prémio Carreira SPE](#) e em [Premio_Carreira-v2.pdf](#).

O Prémio Carreira SPE em 2025, por deliberação unânime do Júri (presidido pelo Presidente da SPE e tendo como vogais os anteriores premiados) foi atribuído aos sócios Fernando Rosado e Maria Eugénia da Graça Martins.

A Professora Maria Eugénia da Graça Martins, licenciada em Matemática Aplicada (FCUL, 1970) e com um Doctorat de 3eme Cycle de Mathématiques Appliquées (Grenoble, 1979), obteve o grau de Doutor em Probabilidade e Estatística da Universidade de Lisboa (1983), com uma dissertação sobre leis estáveis para somas, orientada por Dinis Pestana.

Posteriormente trabalhou também em estabilidade para extremos, publicando nomeadamente Graça Martins, M. E. and Pestana, D. D. (1985). "The extremal limit theorem extensions", in W. Grossman, J. Mogyoródi, I. Vincze and W. Wertz (eds), *Probability Theory and Mathematical Statistics with Applications*, D. Reidel, Dordrecht, 143-153.

Graça Martins, M. E. and Pestana, D. D. (1987). Nonstable Limit Laws in Extreme Value Theory, in M. L. Puri, J. P. Vilaplana . and W. Wertz (eds.) *New Perspectives in Theoretical and Applied Statistics*, Wiley, New York, 449-457.

Cedo se tornou a figura de proa do ensino da cadeira inicial de Probabilidade e Estatística de diversos cursos da FCUL, e a interlocutora natural de todas as instituições envolvidas no ensino Básico e Secundário de Probabilidade e Estatística. Um excelente Manual está disponível – [IPE](#)

Na Sociedade Portuguesa de Estatística, podemos considerá-la a embaixadora junto daquelas instituições, e a consistente animadora, com Manuela Neves e Luísa Loura, do Prémio Estatístico Júnior, e do forte investimento da SPE no despertar de vocações jovens, nomeadamente no apoio a [ALEA](#) –

Ação Local de Estatística Aplicada – um projeto conjunto do Agrupamento de Escolas Tomaz Pelayo, de Santo Tirso e do Instituto Nacional de Estatística e o Explorística, projeto criado pela Sociedade Portuguesa de Estatística com o apoio da Ciência Viva.

Como reconhecimento do seu mérito já em 1995 recebeu da Sociedade Portuguesa de Estatística o Prémio António Raúl Simões Neto, por ser a continuadora daquele ilustre pedagogo, pelo trabalho distinto que tem realizado na área de Ensino da Estatística, como investigadora, como docente de cadeiras de iniciação, e como formadora de professores; e o Departamento de Estatística e Investigação Operacional da FCUL, promoveu uma homenagem em 2010, realizando um colóquio intitulado O Ensino da Estatística do Básico ao Secundário, tendo como oradores Luísa Loura, Luís Gouveia, João Pedro da Ponte, Ana Vieira Lopes, Paula Teixeira, Emília Oliveira, Pedro Campos, João Branco e Dinis Pestana.

O Professor Fernando Rosado licenciou-se em Matemática Aplicada (FCUL, 1971), obteve o grau de Doutor em Probabilidade e Estatística (Universidade de Lisboa, 1984), orientado pelo Professor J. Tiago de Oliveira, e obteve o título de Professor Agregado em 1991.

Trabalhou consistentemente sobre outliers, publicando grande número de trabalhos e orientando dissertações de Mestrado e de Doutoramento nessa área. No XIV Congresso da Sociedade Portuguesa de Estatística foi regente do curso *Outliers em Dados Estatísticos*, disponível em [ODE](#).

No ensino pós-licenciatura, especializou-se também na Análise de Dados Multivariados.

Foi Presidente da Sociedade Portuguesa de Estatística de 2000 a 2005, contribuindo significativamente para o seu reconhecimento internacional, nomeadamente junto do International Statistical Institute, de que foi um dos primeiros portugueses a ser membro eleito. Em 2005 editou o *Memorial da Sociedade Portuguesa de Estatística*, e assumiu a partir de então a edição do Boletim da SPE, que transformou num instrumento de valorização do trabalho dos outros, registando a evolução da Estatística em Portugal. A persistência com que tem inspirado os números temáticos do Boletim da SPE, publicado regularmente na Primavera e no Outono de cada ano, mostram grande acuidade na perceção das novas tendências de investigação em Portugal.

O Presidente da SPE, logo após a deliberação do Júri, publicou na página da SPE uma nota detalhada sobre as carreiras dos premiados (<https://spe2025.mozello.site.com/premio-carreira/>).

Certamente todos concordarão que a Eugénia Graça Martins tem tido um protagonismo reconhecido no que se refere ao ensino da Estatística, nomeadamente pré-universitária, e que o Fernando Rosado tem tido uma ação espetacular no que se refere à visibilidade e coesão da comunidade estatística portuguesa, sendo assim duas personalidades que se distinguiram na projeção da Sociedade Portuguesa de Estatística não só em Portugal como junto de toda a comunidade lusófona.

Por isso a SPE honra-se ao atribuir o Prémio Carreira SPE 2025 ao Fernando Rosado e à Maria Eugénia da Graça Martins.

Dinis Pestana

Manuela Neves



• Um “Conflito de Interesses” *versus* um “testemunho de premiado”

Numa agenda, muito apertada – que reflete a grande vitalidade da comunidade estatística portuguesa – a sessão da entrega dos Prémios Carreira SPE 2025, como noticiado, decorreu durante o Congresso SPE25 e foi organizada pelos Professores Dinis Pestana e Manuela Neves.

Além dos organizadores, também tomaram a palavra, os Professores Luís Meira Machado, Presidente da SPE, Alice Martins, Carvalho e Silva, Emília Oliveira, Susana Rosado; e os Premiados Fernando Rosado e Maria Eugénia Graças Martins.

Como reconhecimento, a efetuar após um benefício recebido, a gratidão reclama do Editor que no presente item do Boletim SPE, se inclua esta mensagem que apenas deseja ser um “testemunho de premiado”. De facto, trata-se de um conflito de interesses, entre Editor e Premiado e que deve ser declarado; mas que foi decidido pela inclusão – em humilde partilha com toda a comunidade.

Dada a apertada agenda, com base nas palavras abaixo, escritas na primeira pessoa, tentei expressar-me, muito grato, na cerimónia de entrega dos Prémios Carreira SPE 2025. Aqui ficam “dadas à estampa”.

Muito obrigado a todos por esta vivência.

E agora, é o momento que eu devo preencher com uma palavra que sintetize o meu estado de alma. Vou usar duas: Humildade e Gratidão. Humildade para, desde logo, confundir a soberba e Gratidão pelo valor e qualidade do benefício recebido. O reviver de uma longa carreira científica e profissional associada à densidade dos pensamentos, emoções e memórias – avós, pais, filhas, netos – que neste momento me preenchem é incompatível com o pouco tempo que vos desejo tomar.

O “primeiro momento” iluminou-me, com uma “dedicatória aos meus pais” – alentejanos humildes que, com muito esforço e a partir de um Alentejo Profundo, conseguiram construir as bases para gerar três ramos familiares “mais à frente”.

Tempus fugit!

Avancemos.

De facto, sei muito bem o valor do tempo em congressos e do cumprimento de uma agenda repleta. É uma alegria que assim seja; para felicidade de todos aqueles que, como eu, vivem apaixonadamente estes momentos fundamentais na partilha científica nacional. Até hoje, não falhei nenhum dos 27 congressos SPE; tal como alguns dos presentes nesta sala. Parabéns ao Presidente da Comissão Científica do XXVII Congresso, Prof Luís Meira Machado, Presidente da SPE e à Comissão Organizadora Local, na pessoa das Profs Clara Cordeiro e Conceição Ribeiro. Sei avaliar, como excelente, o valor do vosso trabalho. Neste ano temos o prazer de viver uma reunião científica que muito honra a tradição da SPE.

Ora bem! E quando menos se espera, surge um Prémio Carreira que, deste o primeiro momento, considero como um novo desafio na minha vida. Para a família, os colegas e amigos, o meu reconhecimento por este benefício, exige, na realidade, uma palavra que é, gratidão a todos.

É enorme a Gratidão que sinto por este belo momento da minha vida que muito me honra.

Desejo também dizer: Parabéns Eugénia, a quem mais uma vez felicito pelo Prémio Carreira SPE que recebeste. Por muitas razões ele é muito bem merecido; entre outras, pela excelência da tua competência e do teu protagonismo e liderança nacional e internacional em tudo o que refere Ensino e Aprendizagem da Estatística no currículo, nos planos de estudos e aprendizagens essenciais no ensino da Estatística e não só; também na pedagogia e métodos. É de enorme valor a tua contribuição para uma melhor Literacia Estatística. Por tudo isto muito obrigado Professora Eugénia. Continua! Continua como tens feito; pois a tua sabedoria e o teu legado devem ser transmitidos.

Agora é devida uma palavra muito especial para o Prof Dinis Pestana pela generosidade da apresentação – grande amigo e companheiro de tantas lides académicas. A diversidade e a excelência, na sua vida científica, iniciada nas literacias da Faculdade de Letras e que para sorte dos Matemáticos, Probabilistas e Estatísticos se converteu numa paixão com dedicação plena que lhe garantiu ser a referência que todos reconhecemos – desde o Minho ao Algarve e Ilhas – no todo nacional e internacional. E tudo isto também eu tive o enorme prazer de acompanhar de perto.

Muito Obrigado! Dinis e, se me permites, nestas palavras e com o mesmo carinho incluo também a Ivette e, obviamente, na minha mente surge uma vasta lista de nomes de alunos, colegas e amigos que, nestes momentos, intensamente mais sentimos.

Muito Obrigado! também para a Prof. Manuela por tão generosas palavras, também de grande amiga.

E por último, o mais difícil e como dizia no início, este momento também envolve uma filha e, embora já tenhamos ouvido, de outrem, a invocação da casualidade na geração filial; a Doutora Susana, não é por acaso que ela é minha filha. É-o, porque muito querida; como todas!

E, na minha gratidão, agora ainda mais emocionada, permitam-me, também, que partilhe este Prémio com a Lúcia – aquela que há quase seis décadas é companheira-esposa e mãe de um triunvirato de mulheres, também já mães, que muito nos orgulham.

Muito obrigado, Susana, pelo teu “desempenho”.

A Prof. Susana que, com muita generosidade e muito Amor, brilhantemente me expôs perante todos vós pertence a uma geração de Estatísticos com raízes na FCUL e nos seus Professores – num Departamento que é uma Raiz na Estatística portuguesa; plantada por muitos daqueles “estrangeiros, estrangeirados e os outros” que os Profs Dinis Pestana e Rui Santos glosaram num documento, sobre as “Raízes e a História da Estatística em Portugal”. Este excelente opúsculo está disponível também na página da SPE; com ampla divulgação no mundo da científico e já com honras de Academia das Ciências.

A Prof Susana Rosado na Faculdade de Arquitetura, tenta transmitir aos futuros arquitetos que para além da sua ciência fundamental também devem incluir um pouco de coisas com números, como bom instrumento de trabalho. As notícias dizem-me que tem tido o merecido sucesso. E como ela, tantos outros iniciados no pioneirismo da FCUL, na promoção da Estatística em Portugal e que, também a partir das mais diversas universidades e empresas, por todo o mundo, vemos trilhar os mais diversos caminhos da Ciência e da Vida.

Este momento, tão belo na minha vida; uma vez mais, traz até mim, a continuidade e o valor de uma paixão de sempre, um lema de “passagem de testemunho” que cada vez mais, convictamente me apaixona – Transmitir duas das coisas mais duradouras que os seres humanos podem deixar em herança: Raízes e Asas.

E, por falar em Raízes e para terminar seja-me tolerado um olhar para muito longe e focar os meus *Outliers* no meu Alentejo Profundo de onde migrei há 60 anos.

Então, da referida autobiografia *Septuagesimus Annum*, aqui vai o Mote de “as minhas décimas” para:

... terras do Grande Lago...

A tranquilidade, em que me revejo,
do viver simples, no Alentejo;
no meu ser, com humildade,
da Força dos Menores fez verdade.
Eis boa razão para não esquecer
O belo lugar que me viu nascer.

Muito obrigado!
Bem hajam!

Fernando Rosado



Edições SPE - Mini Cursos

Título: O guia do viajante para uma aprendizagem de máquina responsável:
Versão para o programa R

Autores: P. Biecek, A. Kozak e A. Zawada

Ano: 2025.

Título: *Análise Estatística de Dados Financeiros*

Autores: C. Amado, C. Nunes, A. Sardinha

Ano: 2019.

Título: *Uma introdução à Meta-Análise*

Autora: Maria de Fátima Brilhante

Ano: 2017.

Título: *Estatística Bayesiana*

Computacional – uma introdução

Autores: M. Antónia Amaral Turkman e

Carlos Daniel Paulino

Ano: 2015.

Título: *Análise de Valores Extremos: Uma Introdução*

Autoras: M. Ivette Gomes, M. Isabel Fraga Alves e Cláudia Neves

Ano: 2013.

Título: *Modelos com Equações Estruturais*

Autora: Maria de Fátima Salgueiro

Ano: 2012.

Título: *Análise de Dados Longitudinais*

Autoras: Maria Salomé Cabral e

Maria Helena Gonçalves

Ano: 2011

Título: *Uma Introdução à Estimação Não-Paramétrica da Densidade*

Autor: Carlos Tenreiro

Ano: 2010

Título: *Análise de Sobrevida*

Autoras: Cristina Rocha e

Ana Luísa Papoila

Ano: 2009

Título: *Análise de Dados Espaciais*

Autoras: M. Lucília de Carvalho e

Isabel C. Natário

Ano: 2008

Título: *Introdução aos Métodos Estatísticos Robustos*

Autores: Ana M. Pires e João A. Branco

Ano: 2007

Título: *Outliers em Dados Estatísticos*

Autor: Fernando Rosado

Ano: 2006

Título: *Introdução às Equações Diferenciais Estocásticas e Aplicações*

Autor: Carlos Braumann

Ano: 2005

Título: *Uma Introdução à Análise de Clusters*

Autor: João A. Branco

Ano: 2004

Título: *Séries Temporais – Modelações lineares e não lineares*

Autoras: Esmeralda Gonçalves e Nazaré Mendes Lopes

Ano: 2003 (2ª Edição em 2008)

Título: *Modelos Heterocedásticos. Aplicações com o software Eviews*

Autor: Daniel Muller

Ano: 2002

Título: *Inferência sobre Localização e Escala*

Autores: Fátima Brilhante, Dinis Pestana, José Rocha e

Sílvio Velosa

Ano: 2001

Título: *Controlo Estatístico de Qualidade*

Autoras: Maria Ivette Gomes; Fernanda Figueiredo e Maria Isabel Barão

Ano: 1999 (2.ª edição, revista e aumentada, 2010)

Título: *Tópicos de Sondagens*

Autor: Paulo Gomes

Ano: 1998



Retrospectiva do Boletim SPE - Tema Central

Disponíveis em <https://www.spestatistica.pt/publicacoes/categoria/boletim-da-spe>

Primavera de 2025 – Teoria das Filas de Espera

Outono de 2024 – Replicabilidade e Controlo de Confidencialidade

Primavera de 2024 – *Rising Stars*

Outono de 2023 – Educação (e) Estatística

Primavera de 2023 – PORTSEA – um mar de Extremos em Portugal

Outono de 2022 – Prémios na Sociedade Portuguesa de Estatística

Primavera de 2022 – Liderança Estatística

Outono de 2021 – *Machine Learning* e Inteligência Artificial

Primavera de 2021 – Especial Covid: a Estatística ao serviço da sociedade

Outono de 2020 – 40 anos SPE: De onde viemos? Onde estamos? Para onde vamos?

Primavera de 2020 – INE–85 anos de estatísticas a servir o país

Outono de 2019 – Estatística nas Ciências da Saúde

Primavera de 2019 – Séries Temporais de Valor Inteiro

Outono de 2018 – Equações diferenciais estocásticas e algumas aplicações

Primavera de 2018 – Estatística Multivariada – perspetiva no século XXI

Outono de 2017 – O Tema Central da Estatística - um novo olhar

Primavera de 2017 – Incerteza em Engenharia

Outono de 2016 – O Tema Central da Estatística

Primavera de 2016 – Séries Temporais e suas aplicações

Outono de 2015 – Estatística em Genética

Primavera de 2015 – Estatística no Desporto

Outono de 2014 – Estatística no Ensino Básico e Secundário

Primavera de 2014 – (Um) Ano Internacional da Estatística

Outono de 2013 – A “Escola Bayesiana” em Portugal

Primavera de 2013 – Estatística não - paramétrica

Outono de 2012 – Métodos Estatísticos em Medicina

Primavera de 2012 – Estatística no Ensino Superior Politécnico

Outono de 2011 – Análise de Sobrevivência

Primavera de 2011 – Sondagens e Censos

Outono de 2010 – Estatística Espacial

Primavera de 2010 – Data Mining - Prospeção (Estatística) de Dados?

Outono de 2009 – Modelos Económétricos

Primavera de 2009 – Investigação (em) Estatística

Outono de 2008 – Processos Estocásticos

Primavera de 2008 – ALEA - Um sítio do nosso mundo

Outono de 2007 – Bioestatística

Primavera de 2007 - A “Escola de Extremos” em Portugal

Outono de 2006 – Ensino e Aprendizagem da Estatística



O MUNDO DA ESTATÍSTICA

ORGANIZAÇÃO PARTICIPANTE



Federation of European National
Statistical Societies

Índice

Editorial	2
Mensagem do Presidente	4
Notícias	5
<i>Enigmística</i>	16
INE – 90 anos de rigor e inovação ao serviço da Sociedade	
Preâmbulo	
<i>Instituto Nacional de Estatística, I.P.</i>	17
A Inovação no INE – 20 exemplos dos últimos 20 anos	
<i>Instituto Nacional de Estatística, I.P.</i>	19
Not Just Another Black Box: AI in the Age of Trusted Statistics	
<i>Sónia Quaresma</i>	48
Revisão do Regulamento das Estatísticas Europeias	
Desafios e oportunidades para os sistemas estatísticos	
<i>Instituto Nacional de Estatística, I.P.</i>	53
Manual do Processo de Produção Estatística do Instituto Nacional de Estatística	
<i>Instituto Nacional de Estatística, I.P.</i>	56
ALEA – Um marco na promoção da literacia estatística: uma visão em 3 partes	
O ALEA e a produção de Estatísticas Oficiais	
<i>Pedro Campos</i>	61
O ALEA na Educação para a Matemática	
<i>Emília Oliveira</i>	61
O ALEA e a Educação em Portugal	
<i>João Gonçalves e Franck de Almeida</i>	62
Capacitação em Ciência de Dados para o Sistema Estatístico Nacional e Administração Pública	
<i>Pedro Campos e Jorge Magalhães</i>	64
<i>Ciência Estatística</i>	67
Artigos, Capítulos de Livros e Livros	67
<i>Prémios Estatístico Júnior 2025</i>	69
<i>Prémio SPE 2024</i>	
<i>Ana Martins</i>	70
<i>Prémio SPE 2025</i>	
<i>Lídia André</i>	72
<i>Prémios Carreira SPE – Fernando Rosado e Maria Eugénia Graça Martins</i>	
Prémios Carreira SPE 2025	
<i>Dinis Pestana e Manuela Neves</i>	74
Um “Conflito de Interesses” versus um “testemunho de premiado”	
<i>Fernando Rosado</i>	76
<i>Edições SPE - Minicursos</i>	78
<i>Boletim através do Tema Central</i>	79