

Estatística: alguns alertas muito atuais



Por: Osvaldo Silva
Professor Auxiliar do Departamento
de Matemática e Estatística
da Faculdade de Ciências e Tecnologia
da Universidade dos Açores
osvaldo.dl.silva@uac.pt

A Estatística é uma ciência capaz de nos fazer entender os factos com base nas informações obtidas a partir dos dados, as estatísticas. Depois de produzidas, divulgadas e disseminadas, as estatísticas podem provocar diferentes interpretações naqueles que as recebem, acomodando, assim, informações diversas e distintas. Só com uma correta apreensão e com a assimilação dos significados que lhes estão na origem é que os respetivos recetores das mesmas poderão fazer o seu adequado uso. Se não houver esse cuidado, acabam por serem mal utilizadas, atribuindo-lhes mais ou menos importância do que a que realmente têm. As estatísticas não podem ser vistas como medidas abstratas, mas sim como observações deduzidas da realidade na forma de medidas, com recurso à modelização. Por exemplo dois homens são convidados a comer dois frangos assados. O primeiro comeu os dois frangos, enquanto o segundo não comeu nenhum frango. O indicador da média dirá que cada homem comeu um frango. Para muitos isso pode causar uma certa descrença na Estatística e nas estatísticas obtidas para retratarem a realidade.

Vivemos numa época com fome de dados e de tempo, onde a sua recolha assume valores gigantescos e impensáveis desde o número de pesquisas no Google, a reprodução de vídeos no Youtube ou a publicação de fotografias nas redes sociais. Com tantos dados, crescem as expectativas do que fazer com eles. As expectativas são muito elevadas e muitos alegam que, com recurso ao big data (Mega dados ou grandes dados), conseguimos resolver todos os problemas que nos afligem. Os big data são analisados por algoritmos desenvolvidos pelo Homem e não podemos nem devemos menosprezar o facto de que errar é humano. Podemos ter um algoritmo com muita boa performance, mas se os dados estiverem incorretos o resultado obtido de nada nos serve. Nesse caso, não é o algoritmo que é bom ou mau, mas antes a forma como o usamos. Os algoritmos nunca poderão descrever de uma forma totalmente objetiva a realidade, por mais fiáveis que possam ser os dados e por mais avançada que possa ser a inteligência artificial. Não nos podemos esquecer de que são os programadores de informática que decidem, através da programação, o que é certo e o que é errado, sendo essas decisões

morais tomadas por essas pessoas. Neste contexto, devemos interrogar-nos sobre os objetivos a serem alcançados por esses algoritmos, por exemplo: Andam à procura da descoberta da verdade ou à procura do lucro? Pretende-se segurança ou eficiência? Todos esses dilemas morais devem ser devidamente pensados e refletidos, para que não hajam muito dissabores num futuro próximo.

Os dados que estão constantemente a ser recolhidos podem parecer registos passivos da realidade, mas já estão a condicionar essa realidade e a alterar cada vez mais o nosso mundo. Esses algoritmos criados que deveriam retratar a realidade estão, muitas vezes, a substituí-la. Por exemplo, na atribuição de créditos bancários, com a criação de perfis de crédito, pessoas com determinadas características podem ter mais dificuldades na obtenção de empréstimo do que outras, pelo que essas pessoas estão a ser logo à partida discriminadas, potenciando assim ainda maiores desigualdades sociais e económicas. Ao nível da segurança, com base nos algoritmos criados para identificar potenciais criminosos, por exemplo nos Estados Unidos da América, verifica-se uma conexão entre os jovens negros pobres e a criminalidade. Assim, a ação policial incide a sua atuação principalmente em locais com indivíduos que se inserem nessa descrição. A criação de perfis étnicos, propiciou que muitos inocentes fossem detidos, engrossando assim as estatísticas e promovendo uma forte associação entre a cor da pele e a criminalidade.

Devemos todos envidar esforços para tentar reconhecer quando a Estatística está a ser usada de forma abusiva e avaliar, de um modo o mais imparcial possível, os nossos julgamentos. A forma de colocar questões é de crucial importância num inquérito por amostragem. De modo intencional ou não, por vezes a formulação empurra os inquiridos numa determinada direção. Um bom questionário coloca questões de forma neutra, devendo haver muito cuidado na forma como as questões são colocadas. Numa determinada investigação empírica, que se serve de uma amostra, o mais importante é que a mesma seja representativa da população em estudo. O uso de um processo de amostragem aleatório tem o objetivo de se garantir uma amostra representativa da população. O problema é quando muitas das pessoas selecionadas não querem participar no estudo. Quando muitas pessoas dizem que não, essa amostra aleatória deixa de ter qualquer valor. Assim, questões mal colocadas, exclusões e não-respostas são algumas das razões pelas quais as sondagens podem reproduzir a realidade com menos precisão do que a que aparentam. Por essa razão, qualquer sondagem tem sempre uma margem de incerteza. Os resultados têm de ser encarados como uma reprodução imprecisa da realidade, é como se olhássemos através de uma janela de vidro uma roseira no jardim, sem nun-



ca termos a possibilidade de tocar nem cheirar uma rosa dessa roseira.

Numa época em que as fake news inundam os órgãos de comunicação social (dos jornais online às redes sociais), os cidadãos que têm a habilidade do ‘pensamento estatístico’ – de fazer a leitura e análise, além de entenderem o significado de informações estatísticas diversas, por exemplo, gráficos, medidas, tabelas, análises, conseguem com mais facilidade verificar se houve manipulação dos dados. Os erros associados a uma qualquer investigação podem ser de diferentes ordens e ocorrer em diferentes momentos: na conceção da investigação, na tradução dos conceitos associados às perguntas pertencentes ao questionário; no desenho e na seleção da amostra; na identificação e na localização dos participantes; e em tantas outras situações. Os erros podem ser considerados como involuntários ou mesmo acontecer de forma propositada. No caso de existirem erros, inevitáveis, apesar do controlo efetuado, há a necessidade de dar-lhes um imediato tratamento, com base numa limpeza dos dados, numa análise exploratória e na validação dos dados.

A Estatística é uma ciência poderosa, mas infelizmente às vezes é utilizada por pessoas, sem escrúpulos, que adulteram os dados e forjam estatísticas com a aparência de verdade para atingirem os seus nebulosos objetivos, como sejam desde vantagens económicas, políticas ou de índole meramente pessoal. Só com uma sociedade em que a literacia Estatística faça parte integral da educação e formação dos indivíduos, e em que estes desenvolvam adequadamente as capacidades de pensar e de racionar de forma crítica, é que poderemos combater esta praga das fake news. Devemos estar todos em alerta, porque a massificação dessas notícias falsas tem objetivos bem definidos, o de dividir e confundir a opinião pública, para que uns poucos consigam reinar a seu belo prazer, sem dar cavaco a ninguém. Já temos muitos exemplos na atualidade do seu modus operandis. A receita é simples, basta acederem à tecnologia e misturar alguns dados com aparência de que retratam a realidade que querem transmitir para difundirem a mensagem pretendida. Num tempo em que

o acesso à informação é instantâneo, a mensagem é difundida rapidamente e o cidadão comum na ânsia de estar informado é, em grande parte, aprisionado mentalmente por essas notícias. Tenha cuidado para não cair nessas “ratoeiras”!

A Estatística exige, para além de conhecimentos técnicos associadas ao processo investigativo, experiência sobre a temática em estudo e a capacidade de diálogo e de comunicação com outros interlocutores, com as mais variadas formações e visões sobre o que está em análise, e ainda uma boa dose de bom senso aquando da interpretação e discussão dos resultados obtidos. Temos de ser bons observadores daquilo que se pretende analisar, de modo a não deturpar a realidade para não se chegar à mesma conclusão da célebre frase dita por Obi-Wan Kenobi para Luke Skywalker no filme Guerra nas Estrelas: Episódio VI – O Retorno de Jedi. “Luke, você descobrirá que muitas das verdadeiras que nos prendemos, dependem do nosso ponto de vista.”

A transparência deve estar sempre associada a todo o processo de investigação. Devemos averiguar quem apresenta essa informação e se tem algum interesse associado ao resultado apresentado. Devemos também procurar informação adicional noutras fontes. Devemos perceber quais os conceitos teóricos que estão a ser avaliados e se estes se encontram devidamente mensurados. É importante também perceber como foram recolhidos os dados e se as análises utilizadas são as mais adequadas. É ainda relevante verificar se há erros grosseiros na apresentação dos resultados (por exemplo, se os gráficos têm algum eixo que distorce os resultados ou se existem valores aberrantes (outliers), caso em que o uso da média dará uma informação pouco credível).

A melhor atitude consiste em ler com atenção, conferir a pertinência do que está sendo dito e averiguar a veracidade das notícias, recorrendo, sempre que possível, a outras fontes. O problema é que isto dá trabalho! Agora está avisado! Não coma gato por lebre! Não reclame depois quando se deixar levar, inadvertidamente, por um resultado em que quatro e quatro somaram qualquer coisa diferente de oito.