



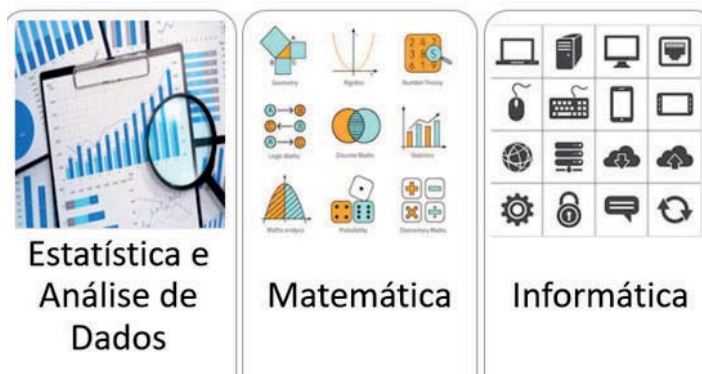
Por: Áurea Sousa
Professora Auxiliar do Departamento de Matemática e Estatística
da Faculdade de Ciências e Tecnologia da Universidade dos Açores
aurea.st.sousa@uac.pt

Os métodos de Estatística e de Análise de Dados, a Matemática e a Informática na atualidade

Na era do Big Data, em que são armazenados grandes volumes de dados (estruturados e não-estruturados), muitas vezes de forma automática através da internet, as competências desenvolvidas pelos indivíduos ao longo da vida são essenciais para uma boa inserção no mercado de trabalho, conferindo-lhes uma atuação mais eficiente e eficaz a nível profissional, razão pela qual estas competências são cada vez mais valorizadas. Muitas dessas competências podem ser desenvolvidas em ambiente escolar e não só (e.g., interação com colegas, familiares ou amigos; internet; livros). A Estatística, a Matemática, a Informática, o domínio de línguas, a habilidade de comunicação e a criatividade estão entre as competências para a vida. Os investigadores e até mesmo as instituições públicas e privadas têm vindo a recorrer, cada vez mais, a modelos matemáticos e estatísticos, por forma a nortear a tomada de decisões, com base na extração de conhecimento útil a partir de dados, sendo relevante o desenvolvimento de competências no âmbito da extração, tratamento e análise de dados. Consequentemente, o termo Ciência de Dados (Data Science) é cada vez mais utilizado na atualidade, visando a extração de conhecimento a partir dos dados para apoio à tomada de decisões. No entanto, mesmo na posse de informações desta índole, a interpretação dos resultados obtidos e a tomada de decisões devem ser levadas a cabo à luz de todo o contexto vigente (por exemplo, no contexto empresarial, pode ser preponderante o conhecimento relativo ao mercado e à legislação, entre outros factores internos ou externos).

Na sociedade do conhecimento e da informação, as bases de dados de grandes dimensões são cada vez mais frequentes, pelo que o seu armazenamento e a sua análise, através de métodos estatísticos e de análise de dados, é essencial. A Estatística pode ser utilizada em todas as áreas do conhecimento (estudos de mercado, estudos de satisfação, hábitos de consumo, pesquisa médica, controlo de qualidade, experimentação agrícola, Economia e Gestão, futebol, previsão de resultados eleitorais, entre outras), onde há recolha e/ou registo de dados, contendo alguma incerteza e variabilidade, contribuindo para o desenvolvimento do conhecimento nessas áreas, no contexto da investigação científica e técnica.

Na atualidade, a carência a nível da existência de estatísticos profissionais ainda é notória, dada a grande pertinência a nível da potencial cooperação entre os indivíduos com estas competências e os profissionais das mais diversas áreas do conhecimento, e a crescente procura dessas soft skills no contexto do mercado de trabalho, cada vez mais competitivo. A título de exemplo, é de referir que a Estatística pode contribuir para os avanços científicos, inclusivamente a nível das Ciências da Saúde, ajudar a nível da solução económica de diversos problemas administrativos, a nível do planeamento de medidas de intervenção no âmbito das políticas públicas, contribuindo também para a melhoria da qualidade de vida dos cidadãos e para uma melhor gestão dos recursos disponíveis. Neste contexto, é de salientar a forte contribuição da Ciência Estatística a nível da tomada de decisões informadas e da argumentação acerca da neces-



sidade e plausibilidade da adoção de determinadas políticas sociais, envolvendo inclusivamente a afetação de recursos financeiros.

Algumas noções matemáticas (e.g., análise combinatória, cálculo diferencial e integral, cálculo matricial, linearização de funções não lineares) foram e são essenciais para o desenvolvimento da Estatística Matemática. Porém, os desenvolvimentos informáticos também têm sido determinantes para a evolução desta ciência, daí a ligação entre a Matemática, a Estatística e a Informática, sendo de salientar a necessidade inerente à contínua criação de algoritmos que permitam a transformação de dados em conhecimento. Assim, embora a Estatística Multivariada tenha surgido por volta de 1901, foram os sucessivos avanços computacionais que permitiram e impulsionaram o seu desenvolvimento e a aplicação prática destes e de outros métodos de Estatística e de Análise de Dados, utilizando software apropriado. Sabia, por exemplo, que aos algoritmos de Análise de Clusters e de Análise em Componentes Principais são técnicas matemáticas, com grande fundamentação na Álgebra e na Geometria? A Álgebra Linear é também encontrada, por exemplo, na Análise de Regressão Linear Múltipla e na Análise de Variância Multivariada (MANOVA), e estes são apenas alguns exemplos. Por outro lado, as competências em Informática também contribuem para uma mais eficiente análise de dados e, na realidade, o software disponível encontra-se em contínuo desenvolvimento, existindo ainda muito a fazer no âmbito da Estatística Computacional (por exemplo, a nível da validação de classes e de partições, no âmbito da Análise Classificatória, muitos dos índices existentes ainda não estão disponíveis na generalidade dos packages (profissionais) de Estatística e Análise de dados e necessitam de ser programados).

Qualquer projeto de investigação científica deve ter como ponto de partida uma boa revisão bibliográfica sobre o estado da arte relativamente à temática em estudo. A clarificação dos objetivos da pesquisa e a elaboração das hipóteses de investigação devem ser efetuadas de uma forma refletida, tendo em atenção todas as informações disponíveis e o contexto da investigação. O foco nas questões de investigação permitirá direcionar mais tarde a

escolha dos métodos estatísticos a serem utilizados, aquando da análise dos dados, a qual deverá dar resposta a cada uma dessas questões.

Na grande maioria dos casos, é impraticável o estudo de toda a população, por diversas razões (e.g., financeiras, tempo, logística), pelo que se procede frequentemente à recolha de uma amostra (parte da população), a qual deverá ser representativa da população. Quando isso não acontece, a amostra diz-se enviesada e a sua utilização é desaconselhável, por poder levar (e geralmente leva) a interpretações erróneas dos resultados obtidos. Para que a Inferência Estatística seja válida, é essencial a utilização de métodos adequados para a seleção da amostra, a identificação da população alvo, evitando-se a falta de aleatoriedade e o erro a nível do cálculo da dimensão da amostra; sendo também essencial a aplicação de métodos apropriados aquando da análise dos dados recolhidos, respeitando os pressupostos para a aplicação dos mesmos, pelo que a literacia estatística (capacidade de entender e compreender os resultados estatísticos) é indispensável também nesse âmbito (isto é, a utilização de software não dispensa a intervenção humana). Assim, o processo de investigação empírica requer um planeamento no que se refere à forma de seleção da amostra, sendo importante antever a utilização de determinados métodos de análise de dados, já antes do início da parte empírica da investigação.

A forma como um questionário é estruturado e organizado é um fator que condiciona, em muito, a análise posterior que pode ser feita a partir dos dados recolhidos, desde a informação que será recolhida até à possibilidade de aplicação de determinados métodos estatísticos. Contudo, muitas vezes há solicitações relativas ao tratamento e análise de dados previamente recolhidos, sem a intervenção de um estatístico na fase inicial da investigação, pelo que é importante alertar e consciencializar a sociedade em geral para o facto de que, no decurso de um estudo empírico, o papel do estatístico deve começar muito antes da recolha e da análise dos dados. Na verdade, a sua contribuição, logo desde a fase de planeamento do estudo, é geralmente uma mais valia, rumo à excelência da investigação científica, conferindo uma maior credibilidade ao processo investigativo.

A relevância da Estatística em cursos das mais diversas áreas científicas e até mesmo a nível do Ensino Básico e Secundário tem vindo a ser cada vez mais reconhecida, sendo de salientar a importância do desenvolvimento da literacia estatística, do raciocínio estatístico e do pensamento estatístico para o pleno exercício da cidadania. Contudo, alguns estudantes recém-chegados ao Ensino Superior manifestam inicialmente alguma ansiedade face às unidades curriculares obrigatórias da área de Estatística, muito provavelmente, devido a vivências experimentadas no decurso da aprendizagem da Matemática no Ensino Básico e/ou Secundário. Os estudantes com mais dificuldades na área de Estatística tendem, geralmente, a associá-la, muitas vezes inconscientemente, a números, a fórmulas matemáticas e a cálculos, por vezes fastidiosos, já que a nível do ensino Básico e Secundário esta temática é lecionada no âmbito da Matemática. Na realidade, há que tirar partido das potencialidades tecnológicas, pelo que a utilização de calculadoras científicas e de computadores deve ser incentivada, libertando a mente dos estudantes para que estes possam colocar também o foco na interpretação dos resultados obtidos (por exemplo, embora seja importante saber como calcular o desvio padrão, não é menos importante saber interpretar os valores obtidos para este coeficiente, utilizando o computador ou um software apropriado), aproximando os estudantes daquilo que é feito habitualmente no meio profissional. Nesse sentido, e de forma a promover o desenvolvimento de competências estatísticas, os jovens devem ser incentivados a terem um papel ativo a nível do processo de ensino e aprendizagem, fomentando-se a sua participação e o seu envolvimento a nível da construção de questionários, da recolha de dados, da formulação de questões e de hipóteses de investigação, da interpretação e da discussão dos resultados obtidos, com vista à sua motivação e ao seu sucesso a nível do desenvolvimento da literacia estatística, no âmbito de uma aprendizagem contextualizada. Por outras palavras, é necessário envolvê-los no processo de ensino e aprendizagem da Estatística, com vista ao desenvolvimento do seu potencial a nível de habilidades e competências nesta área, incluindo o desenvolvimento do pensamento e do raciocínio estatístico, fomentando-se sempre o sentido crítico (Pensar pela sua cabeça!). Este é um compromisso que deverá ser assumido por todos, para que a médio/longo prazo a sociedade seja constituída por cidadãos mais aptos a avaliar criticamente as informações (e as contrainformações) veiculadas diariamente, pelos meios de comunicação social, por instituições públicas e privadas e por inúmeros artigos respeitantes a estudos empíricos das mais diversas áreas.

Os diversos níveis de ensino, em Portugal, ainda apelam excessivamente à memorização, mas há que repensar se não deverá ser dada mais atenção ao desenvolvimento do raciocínio, das competências a nível da procura de informações e da criatividade? E não se esqueça de que somente com o contributo de todos é que é possível combater a literacia estatística! E você já fez, está a fazer ou pensa fazer a sua parte?