



**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

# Amostragem no âmbito da investigação científica - Porquê e para quê?

A investigação científica é um processo sistemático de recolha de dados que permite resolver problemas ligados ao conhecimento dos fenómenos do mundo real em que vivemos, mediante a obtenção de respostas para questões de investigação específicas, permitindo, ainda, confirmar hipóteses sobre as relações presumidas entre fenómenos naturais. Deste modo, a investigação permite criar novos conhecimentos, contribuindo para a expansão do conhecimento científico, sendo importante consciencializar o leitor para a importância da Ciência Estatística no decurso desse processo. Na verdade, a Estatística tem aplicações nas mais diversas áreas científicas (Engenharia, Economia, Psicologia, Sociologia, Medicina, entre outras) e muitos profissionais a consideram como uma ferramenta básica no seu trabalho diário. Dada a vasta quantidade de dados com que somos confrontados diariamente, até mesmo para o cidadão comum é fundamental compreender a informação estatística divulgada pelos meios de comunicação social. É curioso referir que Wells, sublinhou, já em 1903, a importância da literacia Estatística para as gerações futuras, defendendo a ideia de que “No futuro, o pensamento estatístico será tão necessário para a cidadania eficiente como saber ler e escrever”.

Em sua essência, a Estatística é uma ciência que apresenta processos próprios para recolher, apresentar e interpretar adequadamente conjuntos de dados, sejam eles numéricos ou não, sendo uma mais-valia a nível da tomada de decisões e da investigação científica. Os dados podem ser obtidos por meio de uma população ou de uma amostra (parte da população selecionada para ser observada/medida/ entrevistada).

Antes de prosseguir com a leitura deste texto, é importante que o leitor compreenda a distinção entre Estatística Descritiva e Estatística Inferencial. A Estatística Descritiva visa sintetizar uma série de valores da mesma natureza, organizando-os e descrevendo-os através de tabelas, gráficos e do cálculo de medidas de tendência central (média, moda, mediana), medidas de dispersão (e.g., mínimo, máximo, desvio padrão, variância, coeficiente de variação), percentis, quartis, decis e medidas de achatamento e curtose. A Estatística Inferencial tem como objetivo a extrapolação dos resultados (obtidos com a estatística descritiva) para a população, isto é extrapolar resultados da amostra para a população (fazer inferência). Neste âmbito, abrange a estimação de parâmetros (pontual e por intervalos) e os testes de hipóteses.

Num recenseamento (Censo) a população é estudada de forma exaustiva, observando-se todos os elementos que constituem a população. Em contrapartida, uma sondagem é um estudo estatístico realizado a partir de uma amostra. Os problemas de um processo de amostragem ou de sondagem compreendem a escolha de uma amostra (plano de sondagem); e a estimação de um ou vários valores característicos da população (e.g., nível médio de satisfação dos clientes de uma empresa, taxa de desemprego da população residente

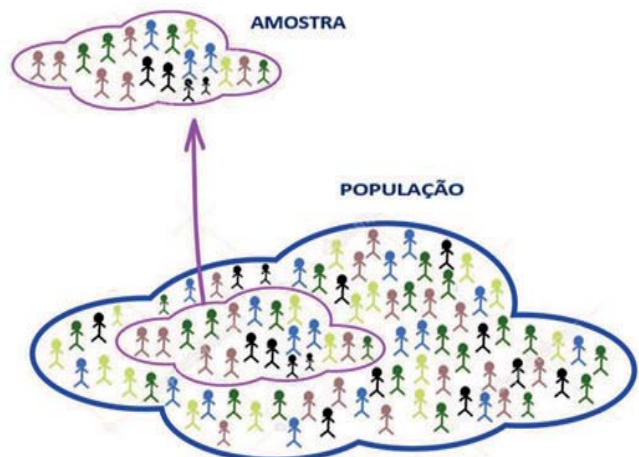
num país), através da construção de estimadores, com base nas observações obtidas.

A opção pela seleção de uma amostra, em vez da realização de um censo, é efetuada, em geral quando a população é infinita, ou considerada como tal, não podendo assim ser analisada na íntegra; a população é grande e, por conseguinte, a realização de um censo é muito cara e requer muito tempo e muito esforço; as unidades que constituem a amostra para inspeção, são destruídas, como acontece frequentemente no contexto do controlo de qualidade; ou alguns elementos da população são inacessíveis, por diversas causas. Em geral, no caso da realização de uma sondagem, há uma menor ocorrência de erros humanos, resultantes de observações incorretas ou incompletas, de deficiências no trabalho dos entrevistadores ou a nível do tratamento de dados, devido ao menor volume de dados. O papel das sondagens é relevante na produção de informação estatística, algumas vezes no decurso de uma investigação científica, na sociedade atual, sendo mesmo considerado como um recurso estratégico essencial para a vida das organizações e para a sociedade em geral, surgindo na comunicação social em diversos contextos (e.g., inquéritos de opinião/sondagens à opinião pública, escolhas políticas, medidas de audiência de canais de televisão, previsões eleitorais, estudos de mercado, utilizados com o intuito de descobrir as preferências das pessoas em relação a certos produtos/serviços).

A amostragem é um conjunto de procedimentos através dos quais se seleciona uma amostra de uma população. Os métodos de amostragem integram-se, essencialmente, em duas categorias: i) amostragem aleatória ou probabilística (todos os elementos da população têm uma probabilidade calculável e superior a zero de integrar a amostra); e ii) amostragem não aleatória ou empírica (não probabilística). Segundo essa definição, a amostragem probabilística implica um sorteio com regras bem determinadas, sendo utilizado um determinado mecanismo probabilístico a nível da seleção dos elementos da população a serem incluídos na amostra, e cuja realização só será possível se a população for finita e totalmente acessível.

O recurso à amostragem probabilística tem a vantagem de permitir controlar e minimizar o erro amostral, por conseguinte deve ser utilizada sempre que possível, especialmente no que se reporta à produção de estatísticas oficiais. No entanto, há situações em que é necessário o recurso a métodos não probabilísticos, nomeadamente no caso de indisponibilidade de uma base de sondagem (lista exaustiva e sem duplicações dos elementos da população alvo, ou seja, é o conjunto de elementos que estão diretamente listados como unidades na base ou que podem ser identificados a partir desta) ou de não existirem meios financeiros, ou outros, suficientes para implementar uma amostragem probabilística de qualidade.

Os problemas de amostragem podem ser mais ou menos complexos, dependendo das populações e das variáveis consideradas pertinentes no contexto do estudo. Na indústria, as amostras



são frequentemente selecionadas com vista ao controlo da qualidade dos produtos e materiais, pelo que em geral os problemas de amostragem são mais simples de resolver. Por outro lado, em pesquisas sociais, económicas ou de opinião, a complexidade dos problemas de amostragem é maior, pelo que deve ser dada uma especial atenção à caracterização da população e à escolha do método de seleção da amostra, a fim de evitar que os elementos que a integram constituam um conjunto com características distintas das da população. Diversos fatores determinam a dimensão da amostra, entre os quais, se encontram o número de grupos e subgrupos dentro da amostra, o valor da informação que é recolhida no estudo, a precisão pretendida dos resultados, o custo da amostra e a variabilidade da população.

O processo de investigação científica é efetuado, geralmente, de uma forma ordenada, abrangendo três fases ou etapas principais: i) a fase conceptual; ii) a fase metodológica; e iii) a fase empírica.

A fase conceptual abrange a escolha e a formulação do problema de investigação, a revisão da literatura (estado da arte) sobre o tema, a identificação clara do objetivo do estudo e a formulação das questões e/ou das hipóteses de investigação. A fase metodológica envolve o desenho da investigação, que inclui a definição da população e da amostra, a definição das variáveis e a escolha dos métodos de recolha (incluindo a escolha dos instrumentos mais apropriados, tais como entrevistas, questionários, grelhas de observação, escalas de medida) e da natureza da abordagem (qualitativa, quantitativa, mista) de análise dos dados, antevendo-se nomeadamente os métodos estatísticos a serem utilizados aquando da análise dos dados recolhidos, a ter lugar na etapa seguinte. Por último, a fase empírica refere-se à recolha de dados, à análise dos dados e à interpretação e disseminação dos principais resultados da investigação. Nesta última etapa, têm uma especial relevância a utilização de métodos estatísticos no âmbito da Estatística descritiva e da

Estatística Inferencial. Embora as técnicas de inferência estatística pressuponham que as amostras utilizadas sejam probabilísticas, frequentemente deparamo-nos com bases de dados em que a seleção da amostra foi efetuada com base em métodos de amostragem não probabilísticos. No entanto, o bom-senso poderá indicar quando o processo de amostragem, embora não sendo probabilístico, pode ser para efeitos práticos considerado como se o fosse, o que amplia consideravelmente as possibilidades de utilização do método estatístico em geral, embora com alguma cautela relativa à generalização dos resultados obtidos.

As diversas decisões metodológicas que têm de ser tomadas na fase metodológica da investigação são essenciais para assegurar a fiabilidade e a qualidades dos resultados obtidos no decurso de uma investigação científica. Por exemplo, é imprescindível assegurar a validade e a fiabilidade dos instrumentos de recolha de dados e a representatividade da amostra.

Numa amostragem ou sondagem não são observados todos os elementos da população, pelo que os resultados obtidos por essa via comportam um erro amostral. A qualidade das estimativas obtidas mediante a realização de sondagens depende das escolhas efetuadas no desenho da amostra, que condicionam a representatividade da amostra, e da utilização apropriada dos estimadores. Na realidade, a obtenção de soluções adequadas para o problema de amostragem exige, em geral, muito bom senso e experiência (garantir a representatividade da amostra deve ser uma prioridade). Dada as interações entre as três etapas principais da investigação, é fácil compreender que o papel do estatístico deve começar logo no início do processo de investigação e não na última etapa. Além disso, o trabalho do estatístico deve ser complementado, sempre que possível, pelo de um especialista da área a que os dados se reportam. Só com dados de qualidade e com a aplicação de métodos de análise de dados apropriados é que é possível obter resultados fidedignos!

Foto: DR