



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

# O papel da distribuição normal na Estatística

A generalização do recurso a métodos estatísticos como suporte científico é uma mais-valia a nível da qualidade da investigação produzida nas mais diversas áreas da ciência, conferindo-lhe um maior rigor e objetividade. Muitas vezes o tratamento estatístico é efetuado unicamente a nível da exploração e da descrição dos dados recolhidos. Quando, além disso, pretende-se explicar, interpretar, extrapolar para a população (inferência estatística), prever (mediante a aplicação de métodos de previsão) ou decidir torna-se imperativa a aplicação de métodos de análise de dados mais sofisticados, que ajudarão a fundamentar a tomada de decisões em situações de incerteza. No entanto, é importante ter em atenção que sem um razoável conhecimento da teoria da probabilidade não se pode ir muito longe no estudo da Estatística Inferencial, no âmbito da qual se enquadram a estimação (por pontos e por intervalos) e os testes de hipóteses, sendo ainda de sublinhar que o conceito básico que permite quantificar o grau de incerteza é o de probabilidade. Nesse contexto, nunca é de mais lembrar que o modo de seleção de uma amostra a partir de uma população (processo de amostragem) é determinante para a qualidade das inferências realizadas.

As dificuldades com que muitos estudantes e profissionais se deparam a nível da compreensão da grande maioria dos artigos científicos publicados e a nível da seleção e utilização de técnicas estatísticas apropriadas, perante a necessidade de análises de conjuntos de dados torna imperativa uma reflexão sobre as suas causas. Neste contexto, sobressai a ideia de que em muitos cursos do ensino superior os conteúdos a nível da Estatística e o seu nível de aprofundamento estão frequentemente aquém das reais necessidades dos futuros profissionais, havendo assim vários cursos em que a inclusão de mais de uma unidade curricular da área de Estatística poderia contribuir para uma maior aquisição de competências, a esse nível, por parte dos estudantes do ensino superior.

Os atributos observados podem ser quantitativos ou qualitativos. Quando são quantitativos afiguram-se com diferentes intensidades ou valores. Quando são qualitativos comportam várias modalidades ou categorias e podem estar em escala nominal, se a ordem das modalidades não tem significado (e.g., Masculino, Feminino) ou numa escala ordinal, se a ordem das modalidades tem significado (e.g., Discordo totalmente, Discordo, Não concordo nem discordo, Concorde, Concorde totalmente).

As variáveis aleatórias são caracterizadas por parâmetros (e.g., média e desvio padrão da população), os quais têm um papel preponderante na modelação estatística. Uma distribuição estatística é uma função que define uma curva, sendo de sublinhar que a área sob essa curva determina a probabili-

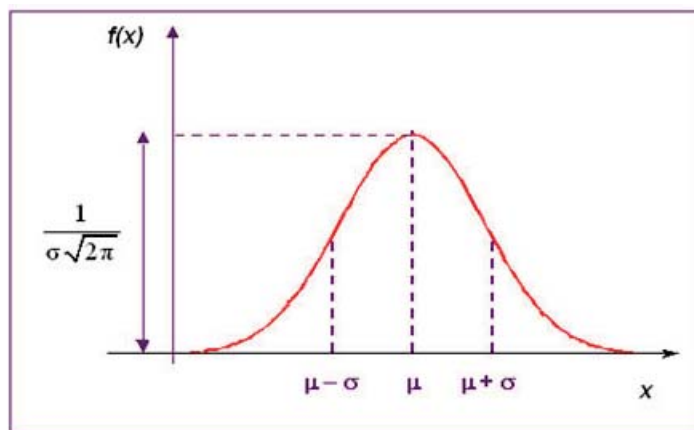
dade de ocorrer um determinado evento. A curva de distribuição mais utilizada é a curva normal ou curva de Gauss, cujas principais características são as seguintes: a área entre a curva e o eixo horizontal é igual a 100%; a curva é unimodal e simétrica em redor do ponto médio e possui a forma de sino.

A distribuição normal, também é designada por distribuição gaussiana, distribuição de Gauss ou distribuição de Laplace–Gauss, em referência aos matemáticos, físicos e astrónomos francêss Pierre–Simon Laplace (1749 – 1827) e alemão Carl Friedrich Gauss (1777 – 1855), é uma das distribuições de probabilidade mais utilizadas a nível da modelação de fenómenos naturais. Esta distribuição foi estudada pela primeira vez por Gauss na abordagem da teoria dos erros de medições repetidas de uma mesma quantidade, o qual, em meados do século XIX, durante os seus estudos sobre eventos da natureza, observou um comportamento padrão entre as amostras que analisou. A função densidade foi construída por Gauss, com o apoio da função exponencial, de modo a garantir as suas principais características: simetria em volta de um ponto (média) em que a função atinge um máximo e convergência rápida para zero à medida que nos afastamos da média. No caso desta distribuição, a média, a mediana e a moda dos valores da variável aleatória considerada possuem o mesmo valor, sendo de salientar que a média e o desvio padrão são os parâmetros que a caracterizam e que a cada par destes valores corresponde uma curva normal, pelo que existem infinitas distribuições normais, dependentes dos valores destes parâmetros. Alterando a média, muda a posição da distribuição. Mudando a variância, muda a dispersão da distribuição.

No caso particular em que a média é zero e o desvio padrão é 1, a distribuição normal denomina-se distribuição normal reduzida ou padronizada/estandardizada. As probabilidades associadas à distribuição normal reduzida encontram-se tabeladas.

A curva normal representa satisfatoriamente o comportamento de diversos processos nas empresas (e.g., as medidas de produtos fabricados em série e os erros de medidas, desempenhando assim um papel relevante a nível do controle estatístico do processo) e de muitas variáveis biométricas (e.g., altura ou peso de uma população, a pressão sanguínea ou o nível de colesterol).

A partir da observação gráfica é fácil visualizarmos as principais características intrínsecas à distribuição normal: i) a variável aleatória pode assumir qualquer valor real; ii) o seu gráfico é uma curva em forma de sino, simétrica em torno da média; iii) a área sob a curva é 1, dado que corresponde à probabilidade de a variável aleatória assumir qualquer valor real. Logo a probabilidade de uma observação assumir um valor entre dois

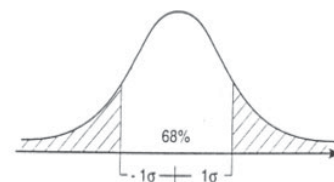


pontos quaisquer corresponde à área compreendida entre esses dois pontos; iv) dada a sua simetria em torno da média, os valores superiores à média e os valores inferiores à média ocorrem com igual probabilidade; v) O aspeto da curva depende de dois parâmetros, a média e a variância.

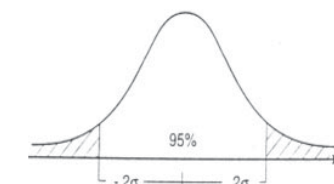
A distribuição normal é frequentemente utilizada quando a variável em estudo apresenta valores concentrados em torno de um valor central, entendendo-se por normalidade a propriedade de uma variável aleatória se distribuir segundo a distribuição normal. No que se refere aos valores de uma variável aleatória com distribuição normal, cerca de 68% estão dentro de mais ou menos um desvio padrão a partir da sua média; cerca de 95% dos valores estão dentro de mais ou menos dois desvios padrão a partir da sua média; e cerca de 99% dos valores estão dentro de mais ou menos três desvios padrão a partir da sua média.

A distribuição normal serve também como base para a inferência estatística clássica. É de referir ainda que a sua importância é relevada pelo Teorema do Limite Central, segundo o qual mesmo que os dados não sigam uma distribuição normal, quando a dimensão da amostra é suficientemente grande a distribuição da média amostral de uma população com variância finita é uma distribuição aproximadamente normal, o que constitui um resultado fundamental em aplicações práticas e teóricas. Graças a este teorema, a distribuição normal aparece como situação limite de muitas distribuições, quer discretas, quer contínuas. Por outro lado, por vezes uma variável que não possui distribuição normal pode ser transformada numa outra com distribuição normal.

À exceção dos testes estatísticos referentes às proporções, um pressuposto geral dos testes paramétricos é que a variável em estudo seja quantitativa e tenha distribuição normal (ou aproximadamente normal) na população, pelo que a aplicação de testes de



$$P(\mu - \sigma < X < \mu + \sigma) \approx 0,683$$



$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0,954$$

normalidade para aferir o ajustamento de uma distribuição empírica a uma distribuição normal tem um especial interesse no âmbito da análise estatística.

Os grandes avanços a nível de software estatístico e a sua mais fácil e frequente utilização tornam ainda mais premente a necessidade de uma aposta num tipo de ensino capaz de desenvolver competências adequadas no âmbito da Estatística, já que uma seleção incorreta dos métodos de análise e/ou das variáveis e/ou a falta de verificação dos pressupostos necessários para a aplicação de testes de hipóteses (por exemplo, a não verificação do pressuposto de normalidade quando este é requerido) conduz geralmente a resultados incorretos. Assim, importa consciencializar os leitores para o perigo da acessibilidade generalizada a software estatístico, os quais derivam essencialmente da utilização de técnicas estatísticas inapropriadas e da má interpretação de alguns resultados.