

Inundação de Médias

Armando B. Mendes

Açoriano Oriental (Bits & Bytes) de 30 de Dezembro de 2006.

As médias aritméticas são certamente a estatística mais usada com o objectivo de resumir dados. Infelizmente a sua utilização é frequentemente incorrecta, indiciando literacia ou pelo menos nesciência relativamente à estatística como ciência, possibilitando o seu uso para “enganar com números”.

As médias inserem-se no conjunto das estatísticas de tendência central, de que são outros exemplos a moda e a mediana. O objectivo destas expressões matemáticas é resumir um conjunto de dados, de forma a extrair a informação relevante sem ter de apresentar todos os valores individualmente. A escolha da média sobre as outras estatísticas é compreensível já que resume bem uma quantidade numérica com uma distribuição de valores “bem comportados”.

O problema está quando os valores não se distribuem de forma regular em torno da média ou quando existem valores extremos muito distintos da média, conhecidos por atípicos ou influentes. Este tipo de distorção pode ser encontrada por exemplo na avaliação das exportações nacionais, muito fortemente influenciada pelo elevado valor dos veículos automóveis fabricados na Autoeuropa.

Outro tipo de utilização incorrecta prende-se com a variabilidade dos valores. Por exemplo, foram recentemente apresentados valores à volta de 18 meses para o tempo médio que um licenciado pode despende até conseguir um emprego compatível com a sua formação. Para ser verdadeiramente útil, a informação anterior tem de ser mais desagregada separando por exemplo os valores nulos para algumas ciências da saúde, como enfermagem ou medicina, e os valores muito elevados nas licenciaturas de ensino e nas áreas sociais e humanas.

As médias podem, assim, ser usadas deliberadamente para esconder variabilidades grandes ou valores atípicos.