

Título

Ok, Computer : confluências na informática

Coordenador da Colectânea dos Textos

Luís Mendes Gomes

Autores

Armando Brito Mendes, Bernardo Rodrigues Peixoto,
José Manuel Cascalho, Luís Mendes Gomes,
Gunther Matthias Funk e Rui Sampaio da Silva

Edição

Influir

Direcção Gráfica e Capa

José Júlio Ribeiro

Impressão

Coingra

ISBN

978-989-97107-1-9

Dep. Legal

329360/11

Apoios

INUNDAÇÃO DE MÉDIAS

As médias aritméticas são provavelmente a estatística mais usada em todos os contextos sempre que o objectivo é resumir uma tabela de dados. Infelizmente a sua utilização é frequentemente incorrecta, indiciando iliteracia ou pelo menos nesciência relativamente à estatística como ciência, podendo ser usada 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 (ou estatísticas) é resumir um conjunto de dados, de forma a extrair a informação relevante sem ter de apresentar todos os valores individualmente. Esta funcionalidade das estatísticas de tendência central é muito apreciada numa diversidade de utilizações desde relatórios científicos até artigos de natureza informativa.

A escolha da média sobre outras estatísticas de tendência central é compreensível se se considerar que tem em conta todos os valores originais, o que não acontece com a moda que se refere exclusivamente ao(s) valor(es) mais frequentes, esquecendo os restantes, e resume bem uma quantidade numérica com uma distribuição de valores “bem comportados”.

O problema surge 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. Neste último caso a utilização da média aparada, ou seja, a média após a eliminação de alguns valores extremos, é mais adequada para resumir o grosso dos valores. Este tipo de distorção pode ser encontrado, por exemplo, na avaliação das exportações nacionais, muito fortemente influenciada pelo elevado valor dos veículos automóveis fabricados na Autoeuropa ou pela importância dos produtos petrolíferos refinados fabricados pela Petrogal. Deste modo, uma média das receitas resultantes da exportação por empresa teria pouco significado.

É fácil encontrar utilizações incorrectas da média em vários contextos. Por exemplo a propósito do relatório da OCDE sobre a avaliação do ensino superior, foram apresentados valores à volta de 18 meses como o tempo médio que um licenciado pode levar até conseguir um emprego compatível com o curso que tirou. Esta informação parece útil, mas na verdade, esconde uma grande variabilidade de



valores que tornam a média anterior, no mínimo, enganadora. Para ser verdadeiramente útil a informação anterior tem de ser mais desagregada permitindo concluir, por exemplo, sobre os elevados níveis de empregabilidade em algumas ciências da saúde, como a medicina e os reduzidos nas licenciaturas que formam alunos para o ensino e as das áreas sociais e humanas.

As médias podem, assim, ser usadas deliberadamente para esconder variabilidades grandes ou valores atípicos, enganando mais do que informando. Outras formas de enganar com estatística são, por exemplo, as variações percentuais comparando momentos ou situações desadequadas ou pouco usuais, o uso de amostras enviesadas, o uso de correlações espúrias, e muitas outras, mas deixemos este tema muito rico para próximos artigos.

Recursos

Huff, D. (1954). *How to Lie with Statistics*: Norton.

Steele, J. M. (2005). *Darrel Huff and fifty years of how to lie with statistics*: Statistical Science, 20, pág. 205-209.

Artigo na wikipédia, en.wikipedia.org/wiki/Misuse_of_statistics

A descrição de um exemplo, www.ibiblio.org/rdu/a-liestt.html