

Correlação X Causa

Rodrigo R. Pescim

29 de junho de 2020

ciência

Quando Nicolas Cage faz filmes, mortes são evitadas; veja outras correlações estatísticas que mentem

RICARDO MIOTO
EDITOR-ADJUNTO DE "COTIDIANO"

09/05/2015 © 02h05



Compartilhar



< 0



OUVIR O TEXTO



Mais opções

Quanto mais as pessoas se divorciam no Estado americano do Maine, maior fica o consumo de margarina naquele país. Por outro lado, quanto mais Nicolas Cage lança filmes em um ano, menos pessoas morrem em acidentes de helicóptero.

Quem diz isso são os números. A correlação é muito forte, não há o que se questionar. Certo? Errado, óbvio.

O doutorando em direito em Harvard Tyler Vigen criou [um site](#) para mostrar com exemplos engraçados algo que os estatísticos já sabem há muito tempo: correlação não significa causalidade. Mesmo quanto a correlação é muito forte.

PUBLICIDADE

Correlação NÃO implica em causa

Editoria de Arte/Folhapre

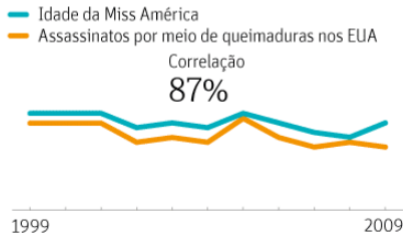
Abra o jornal. Veja a frequência com que argumentos estatísticos questionáveis são utilizados por políticos ou analistas. Após determinada lei, o número de homicídios caiu. Tal governo fez o desemprego ser mais baixo.

Vigen lembra que o mundo é muito mais complexo. "Coincidências existem", diz. "Para estabelecer relações entre duas variáveis, temos de analisar muito bem quais são as relações causais entre elas."

NÚMEROS CERTOS, CONCLUSÕES ERRADAS

Pautando-se apenas pelas estatísticas, é possível chegar a muitas conclusões bizarras

Quanto mais velha a Miss América em um determinado ano, mais gente é assassinada queimada nos EUA



Correlação NÃO implica em causa

Ele lembra que estatísticos têm várias ferramentas à disposição para analisar correlações. A mais clássica, que ele usa, mede só o quanto duas variáveis caminham juntas. Se a correlação é 1, então elas se relacionam perfeitamente: se uma cresce, a outra faz igual na mesma proporção.

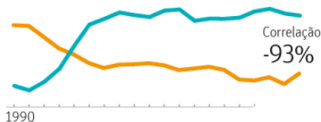
Se esse valor é -1, então a correlação é perfeitamente inversa: toda vez que uma variável cresce, a outra diminui na mesma razão. Por fim, se a correlação é 0, os dados são absolutamente independentes.

Sabe qual a correlação entre o número de americanos que se afogam por ano depois de caírem de um barco quando pescavam e a taxa de casamentos em Kentucky?

Quanto menos se produz e vende mel nos mercados, mais jovens são apreendidos por porte de maconha nos EUA



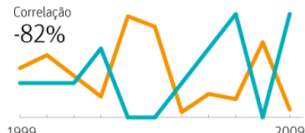
- Número de colônias de abelhas produtoras de mel
- Apreensão de menores por porte de maconha



Quanto menos filmes Nicolas Cage faz em um ano, mais gente morre em acidentes de helicóptero nos EUA



- Número de filmes feitos por Nicolas Cage por ano
- Acidentes com helicópteros matando seus ocupantes



Correlação NÃO implica em causa

Impressionantes 0,95.

E entre a idade da Miss América e o número de assassinatos por queimadura no país? É 0,87 –quanto mais velha a moça, mais você deve suspeitar de quem chega perto com água fervente. (Nos EUA, eles têm mesmo estatísticas para tudo.)

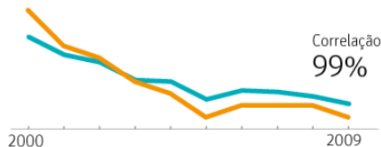
Modos mais sofisticados de análise estão à disposição. Um conceito importantíssimo é o de significância estatística –com ele, é possível calcular a chance de um resultado ser fruto de mero acaso.

O problema que Vigen tenta apontar é que todo esse arsenal mais robusto da estatística é complicado de compreender e muitas vezes acaba ignorado pelo público.

A redução nacional no consumo de margarina leva à diminuição nas taxas de divórcio no Estado do Maine (EUA)



— Consumo per capita de margarina em libras per capita
— Número de divórcios para cada mil pessoas



Fonte: tylervigen.com

Correlação NÃO implica em causa

MENTIRAS

Vigen é fã do clássico "Como Mentir com Estatísticas", de 1954, por Darrell Huff –por incrível que pareça, ele era jornalista, não matemático.

O estatístico Marcelo Menezes Reis, da Universidade Federal de Santa Catarina, também fã do livro de Huff, escreveu uma [interessante página](#) sobre o tema. Ele quer evitar que "o fascínio por números e percentagens turve o senso crítico das pessoas".

Um exemplo de malabarismo estatístico comum citado por ele é a utilização sem cuidado da média. Imagine uma empresa em que quatro funcionários recebem R\$ 300; outros três, R\$ 400. Um único, porém, ganha R\$ 25.000. A média salarial é de R\$ 3.350 –esse povo não ganha tão mal, alguém poderia dizer.

Em outras palavras, um sujeito com a cabeça no forno e as pernas na geladeira está com a temperatura corporal média ótima. É piada, mas pode convencer. ★★