

1. O que é coeficiente de correlação?

↳ É o grau de correlação entre duas variáveis em uma escala contínua (conhecida também como coeficiente de correlação de Pearson ou " r "), é um número cujo os valores podem variar entre -1.0 e $+1.0$.

Existem três correntes possíveis:

↳ Variáveis "positivamente" correlacionadas, I.E:

↳ Quando os valores mudam de forma coincidente, se uma aumenta a outra aumenta.

↳ Variáveis "negativamente" correlacionadas, I.E:

↳ Quando valores crescentes de uma variável estão associados à redução dos valores da outra, se uma aumenta a outra diminui.

↳ Sem correlação:

↳ Quando não é possível perceber relação entre as variáveis

2. Quando o coeficiente de correlação é útil e quando ele pode ser a ferramenta errada p/ analisar a relação entre duas variáveis?

* O C.C. é a ferramenta errada nos seguintes cenários:

↳ Relação não linear: a relação tem de ser linear.

↳ Quando as amostras não são independentes umas das outras.

↳ Quando as amostras possuem outliers

↳ Quando as amostras possuem subgrupos para uma ou ambas variáveis

↳ Caso o tamanho da amostra seja muito pequeno

↳ Quando uma ou ambas as variáveis foram medidas em uma escala ordinal

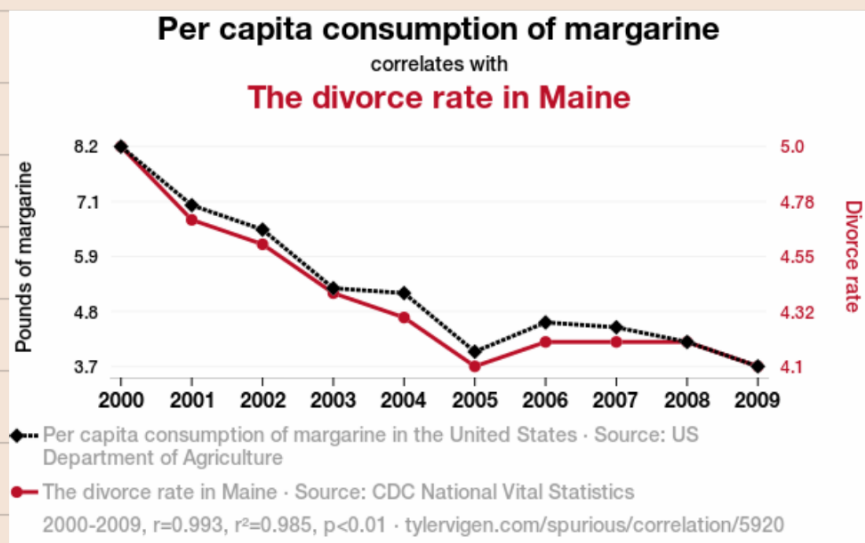
↳ Em caso de "Heteroscedasticidade", I.E: situação em que uma variável tem variabilidade desigual ao longo da faixa de valores da segunda variável

* O C.C. é útil para identificar possíveis relacionamentos entre variáveis em um conjunto de dados

3- Qual a diferença entre causalidade e correlação?

Correlação se refere da relação entre duas variáveis, fala a respeito do comportamento de uma variável perante a mudança da outra.

Causalidade se refere sobre a causa de uma p/la outra, I.E a mudança da variável A causa a mudança da variável B (lembrando que a relação tem de ser linear).



Um exemplo famoso, a taxa de consumo de margarina com a taxa média de divórcios em Maine. Um dado não causa o outro mas a correlação entre os dados aparece.

O mesmo possivelmente também vale para o texto sobre a quantidade de prêmios nobel com a quantidade de consumo de chocolate. Os dados parecem ter muito mais relação com fatores socio ambientais. Além de cair em alguns casos onde não deve ser utilizado. (tamanho pequeno de amostra, outlier, dentre outros).