

Correlação

Construção do teste de associação entre variáveis

Rafael Nunes Magalhães

Depto. de Ciência Política - USP

Métodos III

Novembro 2012

Em ciências sociais, é frequente termos interesse em saber se dois fenômenos estão associados:

- Quando um país reduz barreiras comerciais, aumenta sua taxa de crescimento?
- Candidatos que investem mais em campanha se elegem com mais frequência?
- Países que produzem mais riqueza têm menor taxa de mortalidade infantil?

Introdução

Associação entre fenômenos

Definimos que duas variáveis estão associadas quando, ao mudar o valor de uma delas, encontramos variação também na segunda. Podemos, então, pensar na correlação como *covariância*.

O termo correlação significa relação em dois sentidos (co + relação), e é usado em estatística para designar a força que mantém unidos dois conjuntos de valores. A verificação da existência e do grau de relação entre as variáveis é o objeto de estudo da correlação.

Relembrando a fórmula da variância

$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{N}$$

É o mesmo que

$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x}) * (x_i - \bar{x})}{N}$$

Para encontrarmos a covariância entre duas variáveis, incluimos o segundo fenômeno de interesse

Covariância

$$\text{cov}(x, y) = \sum_{i=1}^n \frac{(x_i - \bar{x}) * (y_i - \bar{y})}{N}$$

O índice de covariância que calculamos tem um problema: não está padronizado. Assim, não podemos comparar covariâncias entre variáveis com medidas muito desiguais.

Para normalizarmos a covariância, utilizamos o *coeficiente r de Pearson*. Basta dividir a covariância pelo desvio-padrão das duas variáveis.

r de Pearson

$$r = \frac{\sum_{i=1}^n \text{cov}(x, y)}{\sigma_X \sigma_Y}$$

Coeficiente de Correlação

Características

O coeficiente de correlação linear r varia de -1 a $+1$ e sua interpretação dependerá do valor numérico e do sinal:

- $r = 1 \longrightarrow$ correlação positiva perfeita
- $0 < r < 1 \longrightarrow$ correlação positiva
- $r = 0 \longrightarrow$ não há correlação (variáveis independentes)
- $-1 < r < 0 \longrightarrow$ correlação negativa
- $r = -1 \longrightarrow$ correlação negativa perfeita

Se elevarmos o coeficiente r ao quadrado, teremos o valor percentual da variação conjunta entre as duas variáveis. Mais precisamente, podemos dizer que a variação de X permite prever $n\%$ da variação de Y

Teste da Correlação

Fazendo inferências sobre a associação entre variáveis

Quando estamos trabalhando com uma amostra, é interessantes testar a associação entre as variáveis na população. Vamos calcular se a correlação linear ρ é estatisticamente significativa.

H_0 : Não há correlação populacional ($\rho = 0$)

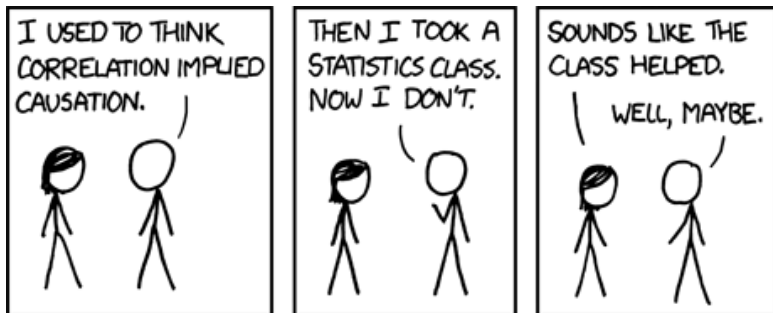
H_1 : Há correlação populacional ($\rho \neq 0$)

Estatística de teste

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

Correlação e causalidade

A lição mais importante da estatística



Correlação e causalidade

A lição mais importante da estatística

Quando encontramos associação entre duas variáveis, não conseguimos distinguir se:

- 1 X influencia Y
- 2 Y influencia X
- 3 X e Y se influenciam mutuamente
- 4 Z influencia X e Y (variável omitida)