



DCP098

Fundamentos para Avaliação Quantitativa de Políticas Públicas

Requisitos para inferência causal

Correlação e espuriedade

Aula 03
06 de abril de 2022

Ana Paula Karruz

Como podemos apurar causalidade/ causação?

- Uma resposta: somos **convencidos pela evidência** (i.e., informação que outros podem verificar)
- Qual a base da evidência? **Em alguns casos, “vemos”** causas se desdobrarem em efeitos (e.g., vela tombar e começar um incêndio)
- Em Política e Políticas Públicas, frequentemente precisamos **rastrear uma cadeia causal**, o que pode ser complicado
- Quando não há como observar diretamente causa e efeito, procuramos **evidência de associação causal** em dados

Segundo a lógica da explicação nomotética*, há três requisitos para inferência causal

Para que um resultado seja considerado consequência de uma suposta causa, todos os três critérios devem ser atendidos



Critério	Descrição	Ilustração
Anterioridade	<ul style="list-style-type: none">▪ Pretensa causa preceder efeito	$\Delta X \xrightarrow{\text{tempo}} \Delta Y$
Associação (a.k.a. correlação)	<ul style="list-style-type: none">▪ Existe uma associação empírica entre a suposta causa e o resultado observado (situação inicial + efeito)	$\Delta X \xrightarrow{a} \Delta Y$ <div>Correlação é uma forma de se apurar associação entre duas variáveis. Frequentemente, toma-se a parte pelo todo, usando-se o termo “correlação” como sinônimo de “associação”.</div>
Não espuriedade (da associação)	<ul style="list-style-type: none">▪ Efeito não pode ser atribuído a outra causa (e.g., a outro programa concomitante) <div>Como podemos separar o efeito de X sobre Y do efeito de Z sobre Y?</div>	$\Delta X \xrightarrow{a} \Delta Y$ $\Delta Z \xrightarrow{c} \Delta X$ $\Delta Z \xrightarrow{b} \Delta Y$

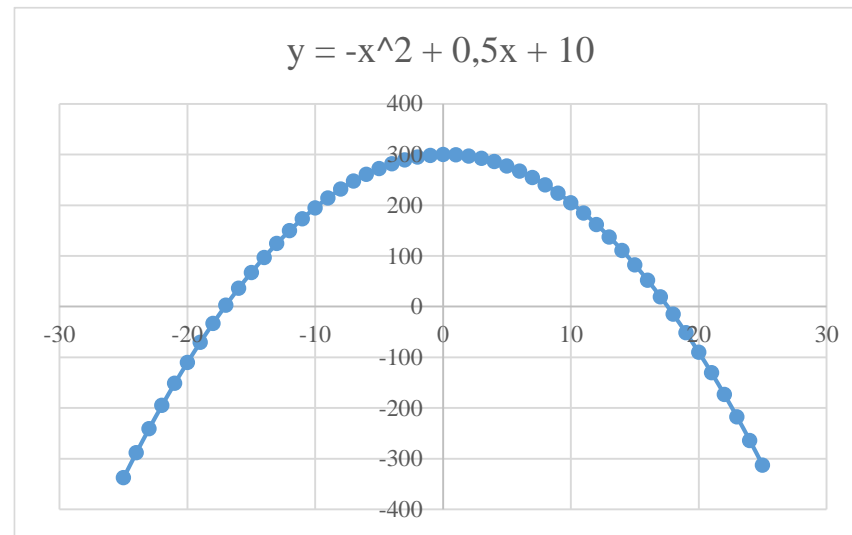
* A explicação nomotética (vs. idiográfica) busca identificar alguns poucos fatores que em geral impactam uma classe de condições ou eventos. Assume causação probabilística (Babbie, 2004).

Duas variáveis são associadas se elas “se movimentam juntas”

- **Associação positiva:** valores altos (baixos) de uma variável são mais frequentes quando a outra variável apresenta valores altos (baixos)
 - Relação/ associação “direta”
- **Associação negativa:** valores altos de uma variável são mais frequentes quando a outra variável apresenta valores baixos
 - Relação/ associação “inversa”

Relembrando o coeficiente de correlação de Pearson

- O coeficiente de correlação de Pearson é uma **medida do grau de associação linear** entre duas variáveis quantitativas



$$\text{Corr}(x,y) = 0,03797225$$

Relembrando o coeficiente de correlação de Pearson

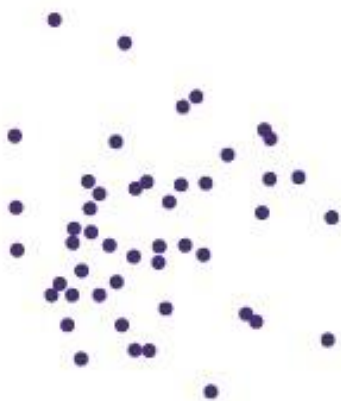
- O coeficiente de correlação de Pearson **varia entre -1 e +1**:
 - -1 indica uma associação linear negativa perfeita
 - +1 indica uma associação linear positiva perfeita
 - 0 significa que não há associação linear
- Quanto **mais próximo** o coeficiente de correlação de Pearson estiver de **+1 ou -1**, **mais forte** é a associação linear entre as duas variáveis
- O coeficiente de correlação de Pearson é normalmente representado pelas letras “r” ou “ ρ ” (letra grega rô)

Relembrando o coeficiente de correlação de Pearson

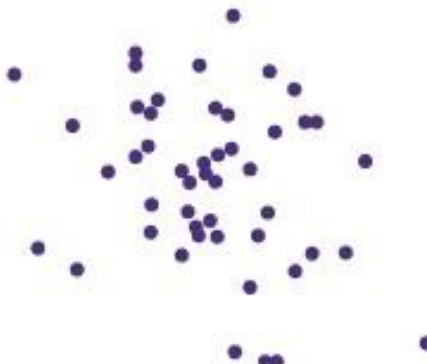
$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}}$$

- **Numerador** é chamado de **soma dos produtos cruzados**
- Conceitualmente, numerador é uma medida de quanto os valores dos **pares ordenados (Xi, Yi)** são **associados**
- Denominador “**padroniza**” o **numerador**, removendo os efeitos das unidades de medida; assim, é possível comparar correlações, independentemente da escala das variáveis

Alguns exemplos de correlação



Correlation $r = 0$



Correlation $r = -0.3$



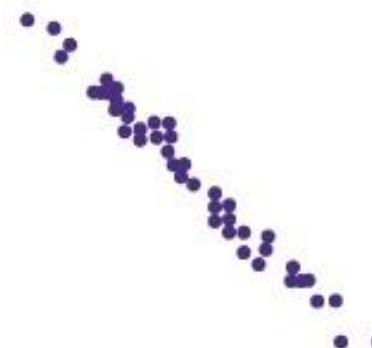
Correlation $r = 0.5$



Correlation $r = -0.7$



Correlation $r = 0.9$



Correlation $r = -0.99$

Observe a direção da associação (sinal) e a dispersão na nuvem de pontos; não confunda com inclinação. Got time? <http://guessthecorrelation.com>

Qual a correlação entre consumo de donuts e massa (peso)?

Table 1.1: Donut Consumption and Weight

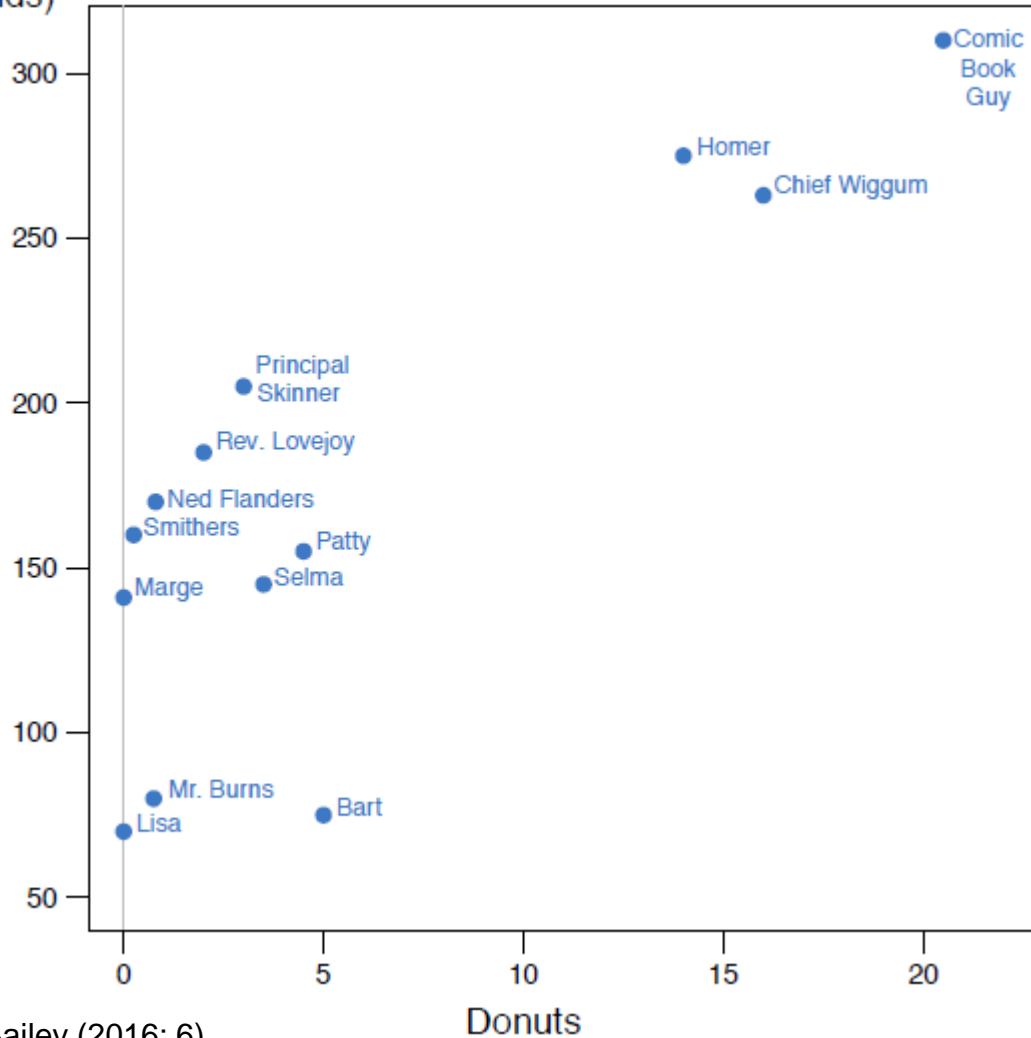
Observation number	Name	Donuts per week	Weight (pounds)
1	Homer	14	275
2	Marge	0	141
3	Lisa	0	70
4	Bart	5	75
5	Comic Book Guy	20	310
6	Mr. Burns	0.75	80
7	Smithers	0.25	160
8	Chief Wiggum	16	263
9	Principal Skinner	3	205
10	Rev. Lovejoy	2	185
11	Ned Flanders	0.8	170
12	Patty	5	155
13	Selma	4	145

Fonte: Bailey (2016: 5).



MATT GROENING

Weight
(in pounds)



**Guess the
correlation**

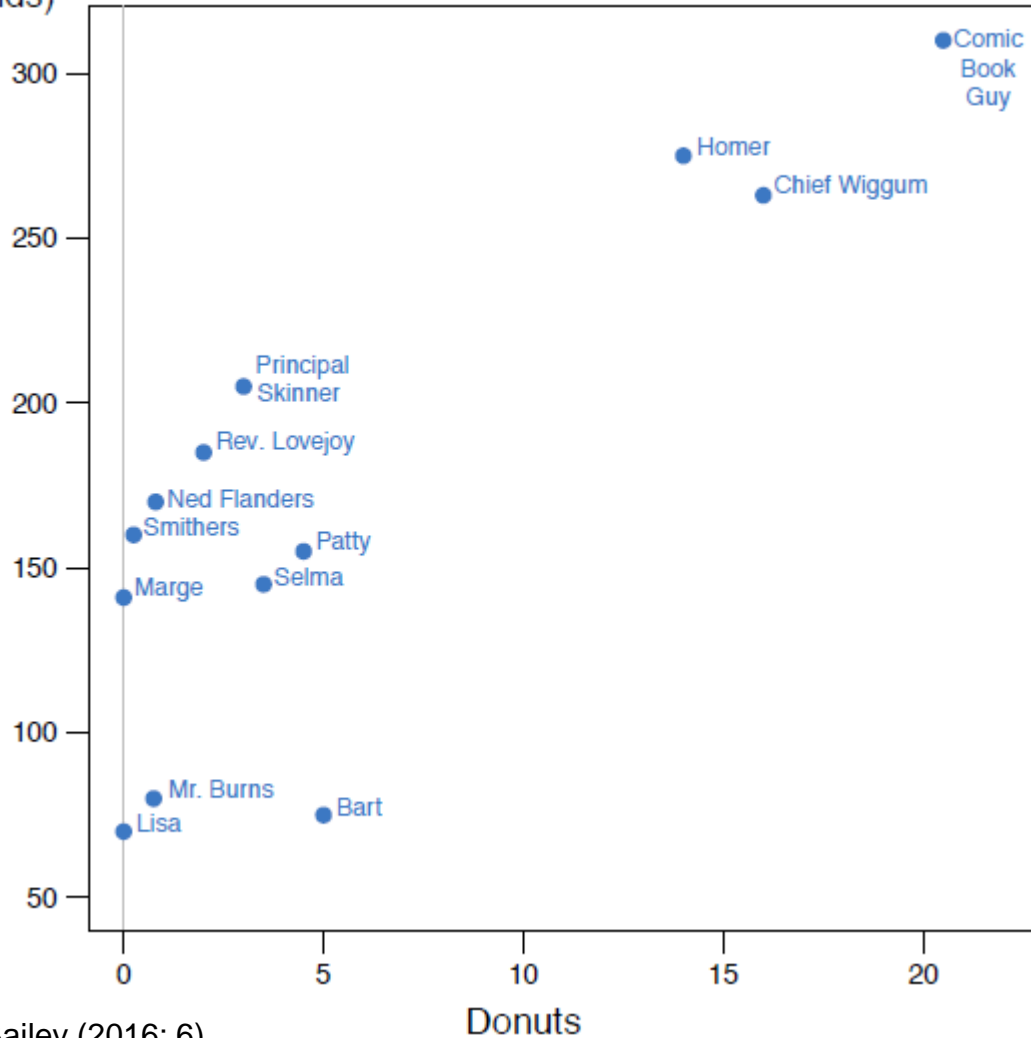


Fonte: Bailey (2016: 6).



MATT GROENING

Weight
(in pounds)

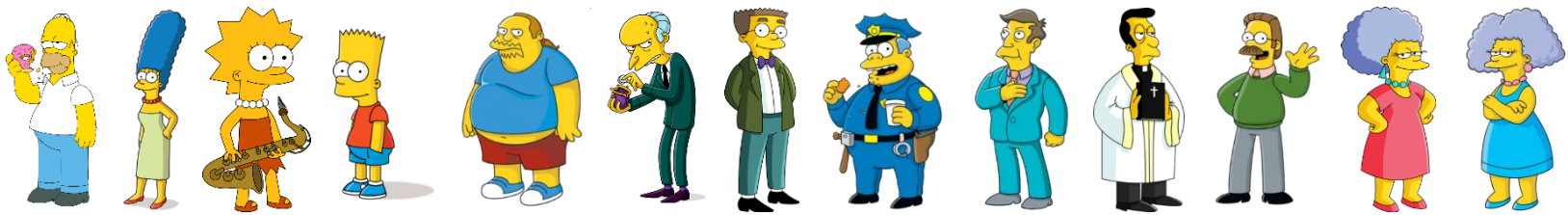


**Guess the
correlation**



$$r = 0,817$$

Fonte: Bailey (2016: 6).



MATT GROENING

Correlação observada é forte o suficiente para acreditarmos que ela existe na população?

Ou, alternativamente, a correlação observada é provavelmente fruto do acaso?

Teste de significância estatística para correlação

Hipóteses

- H_0 : Não há associação entre donuts per week e weight in pounds (correlação na população = 0)
- H_1 : Há associação entre donuts per week e weight in pounds (correlação na população $\neq 0$)

Estatística de teste

$$\text{Test statistic: } t = r \sqrt{\frac{n-2}{1-r^2}}, \quad d.f. = (n-2)$$
$$t = 0.817 * (11/0.333)^{1/2} = 4.699$$

Regra de decisão

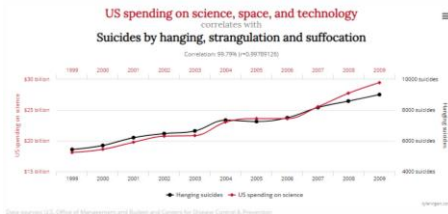
- Compare a estatística de teste (t) com o t -crítico $_{(\alpha, d.f.)}$
- Rejeite H_0 se $|t| \geq t\text{-crítico}_{(\alpha, d.f.)}$

$$t\text{-crítico}_{(\alpha, d.f.)} = t\text{-crítico}_{(5\%, 11)} = 2.201$$

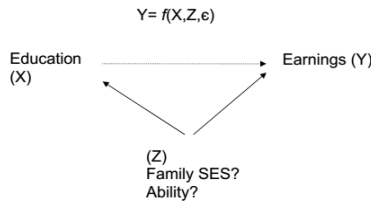
Resultado do teste

- Correlação entre donuts per week e weight in pounds é estatisticamente diferente de zero, a 5% de significância

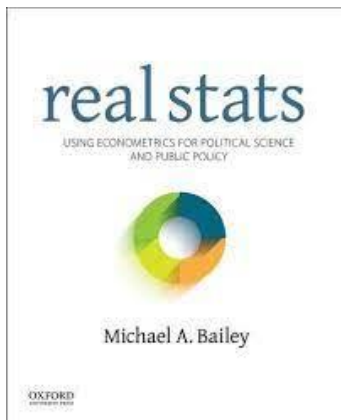
Nem toda associação é causal: tipos de associação espúria



- Associações podem ser **fortuitas** (i.e., fruto do acaso). Exemplos:
 - <http://www.tylervigen.com/spurious-correlations> (apesar do nome, essas correlações não me parecem propriamente espúrias)

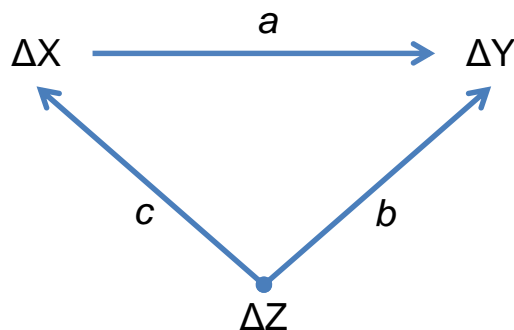


- Associações podem ser **explicadas** pela relação de ambos X e Y com uma terceira variável, Z. Exemplos:
 - Consumo de sorvete e mortes por afogamento
 - Música country e suicídio (Bailey, 2016, p. 25-28)
 - Golfada e engorda de bebês (Figueiredo Filho e Silva Júnior, 2009, p. 131-132)



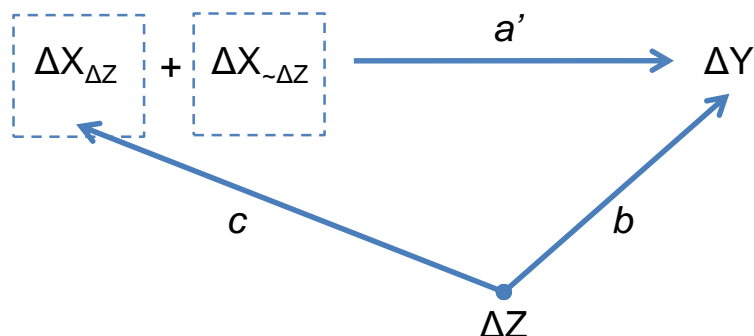
Zooming in na associação espúria

a é associação espúria; a' pode não ser espúria



Associação a é espúria (i.e., não é uma medida válida da associação causal entre X e Y) porque reporta uma combinação de dois efeitos: efeito de X sobre Y e efeito de Z sobre Y.

- Para que a fosse não espúria, para cada potencial fonte de espuriedade Z, as associações b e c não poderiam existir concomitante (ou seja, pelo menos uma delas teria de ser nula)
- **Atribuição aleatória ao tratamento** (desenho experimental) e **análise por estrato de Z eliminam variações de Z**; se Z não varia, não se associa com X ou Y na amostra
- Análise por estrato de Z é uma solução inferior à atribuição aleatória ao tratamento; estratificação deve ocorrer para todo potencial Z, e pode render estratos com poucas observações



- Se associação a' for não nula para todas as potenciais fontes de espuriedade Z, então a proposição de que variações em X causam variações em Y é válida.
- **O controle estatístico** (i.e., regressão múltipla em que se estima o efeito de X sobre Y controlando-se por Z) e o uso de **variável instrumental** são formas de se **identificar a associação a'**
- Porém, ambas são soluções inferiores à atribuição aleatória ao tratamento



DCP098

Fundamentos para Avaliação Quantitativa de Políticas Públicas

Requisitos para inferência causal

Correlação e espuriedade

Aula 03
06 de abril de 2022

Ana Paula Karruz