

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Correla  o

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br



Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

Objetivos da aula

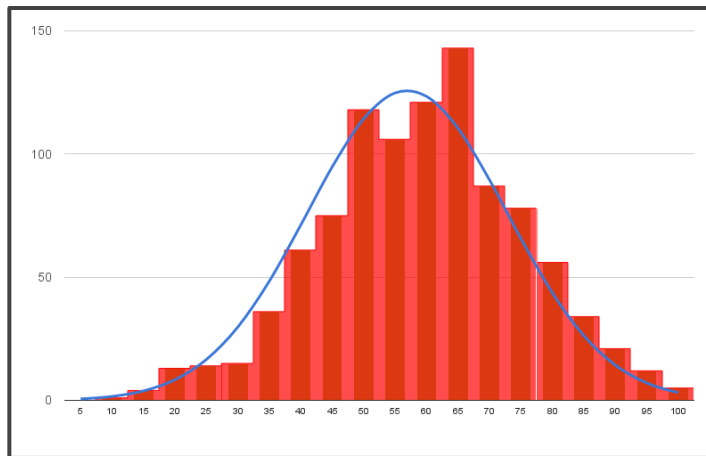
Obter noções básicas de:

- Covariância.
- Correlação.
- Causalidade e variável de confusão.
- Desenho de experimentos.

Até o momento...

Aprendemos a descrever estatisticamente dados que envolve uma variável
(**análise univariada**);

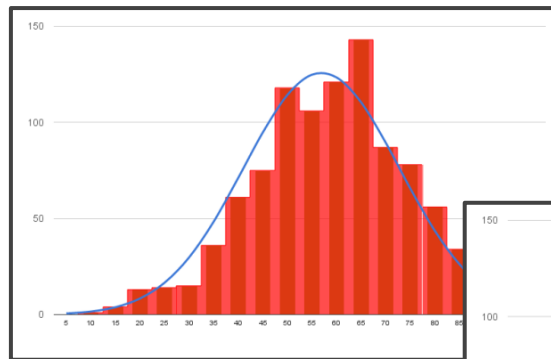
paciente	Força no quadríceps (N)
1	196
2	216
3	216
4	392
...	...



E se...

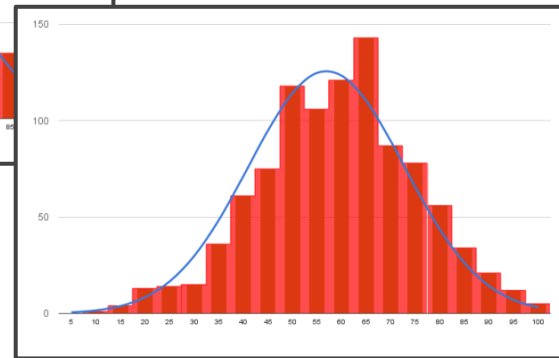
... além de termos os dados da força, temos dados da altura de cada paciente que fez o teste. Que tipos de perguntas podemos fazer?

paciente	Força no quadríceps (N)	Altura
1	196	155
2	216	159
3	216	159
4	392	160
...	...	



Força

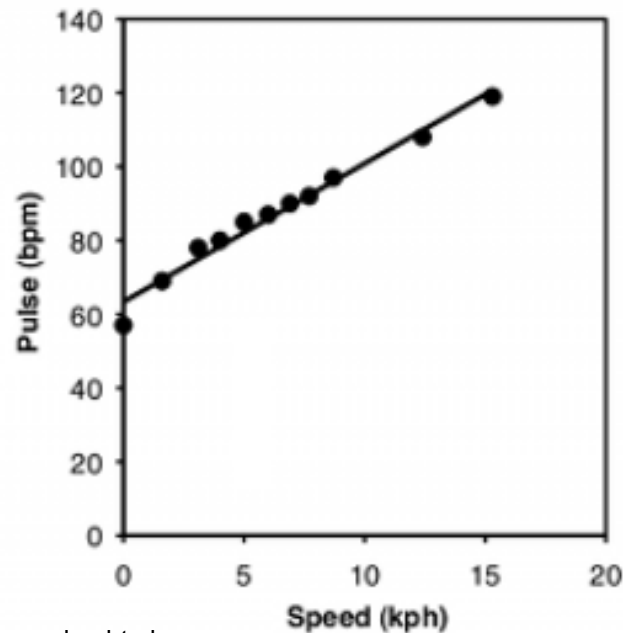
Altura



Relações bivariadas

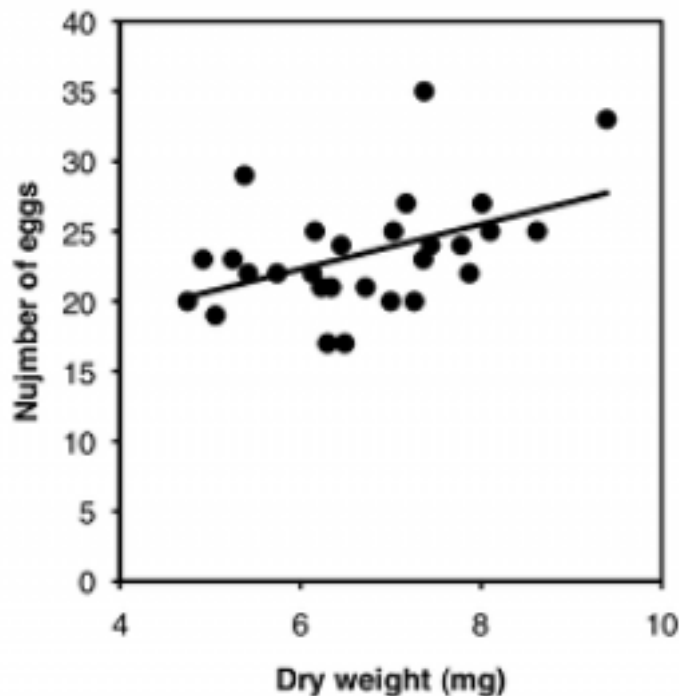
- Pulsação vs. velocidade em uma máquina elíptica

Speed, kph	Pulse, bpm
0	57
1.6	69
3.1	78
4	80
5	85
6	87
6.9	90
7.7	92
8.7	97
12.4	108
15.3	119



Relações bivariadas

- Número de ovos vs.
peso seco no anfípode
Plathoschestia
platensis



Weight, mg	Eggs
5.38	29
7.36	23
6.13	22
4.75	20
8.10	25
8.62	25
6.30	17
7.44	24
7.26	20
7.17	27
7.78	24
6.23	21
5.42	22
7.87	22
5.25	23
7.37	35
8.01	27
4.92	23
7.03	25
6.45	24
5.06	19
6.72	21
7.00	20
9.39	33
6.49	17
6.34	21
6.16	25
5.74	22

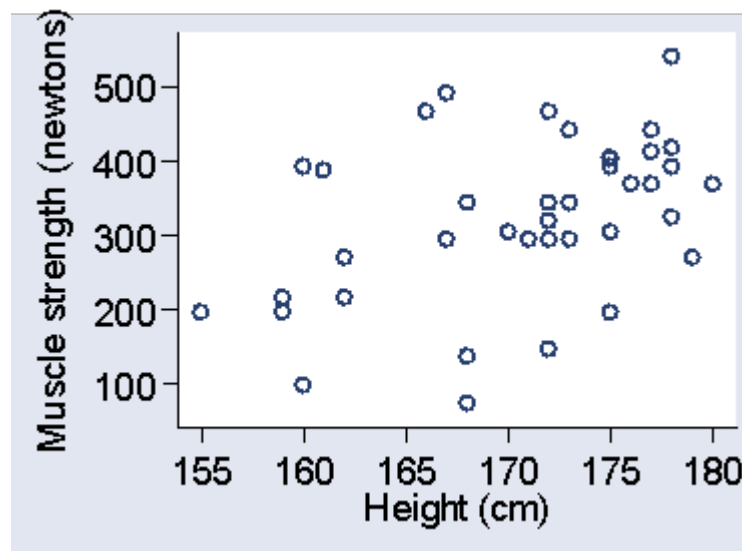
Relações bivariadas

- Quando ambas variáveis são numéricas:
 - Variável resposta (dependente, eixo y).
 - Variável explanatória (independente, preditora, eixo x) **que pode** estar relacionada à variável resposta.
- Função (linear, não-linear)
- Direção (positiva, negativa)
- Magnitude (dispersão)
- *Outliers*

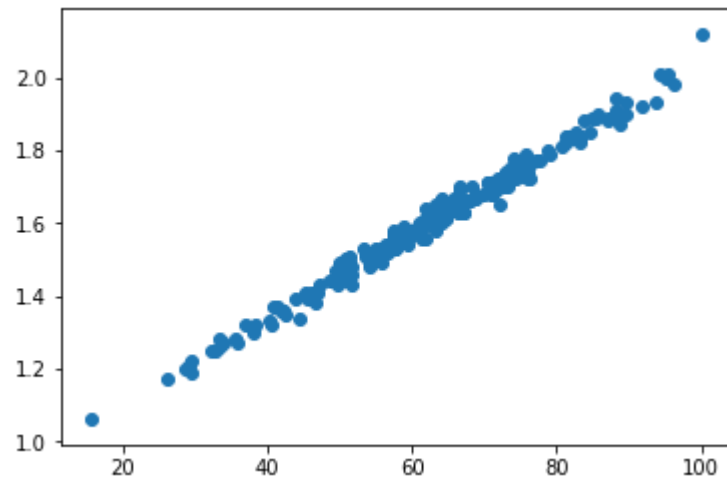
Será que estas variáveis
estão associadas?

Será que estas variáveis se relacionam?

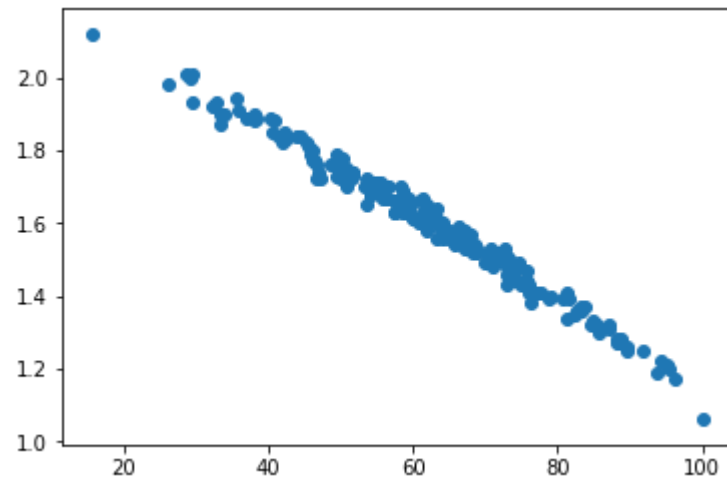
paciente	Força no quadríceps (N)	Altura
1	196	155
2	216	159
3	216	159
4	392	160
...	...	



Como se comportariam duas variáveis que estão associadas?

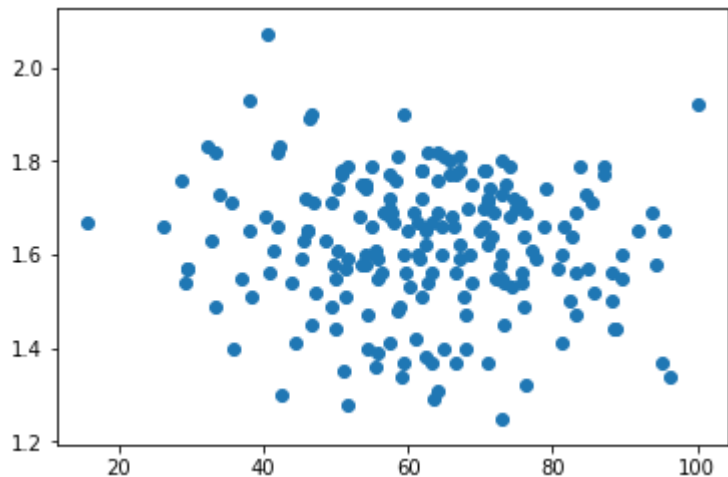


Associação positiva



Associação negativa

Como se comportariam duas variáveis que não estão associadas?



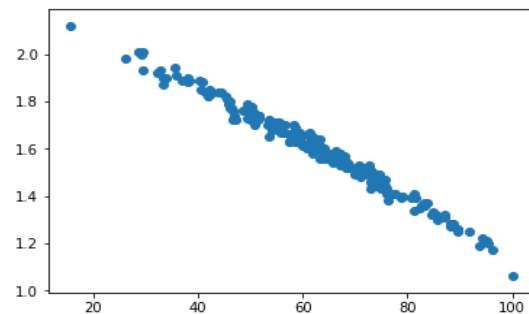
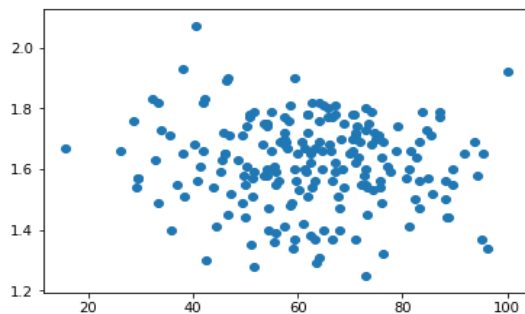
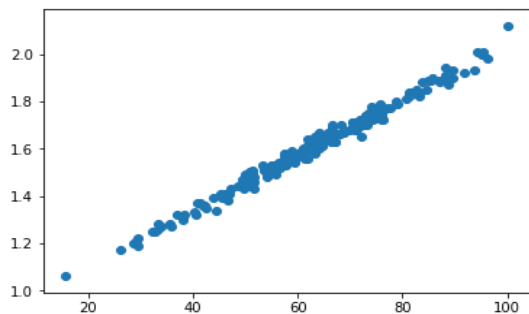
Ter informações sobre uma variável não fornece nenhuma informação sobre qual será o valor correspondente na outra variável!

Quão fortemente estas duas
variáveis estão associadas?

Como medir a relação entre as variáveis?

Covariância: média do produto do desvio das variáveis X e Y.

$$Cov(x, y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{n}$$



Propriedades da Covariância

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{n}$$

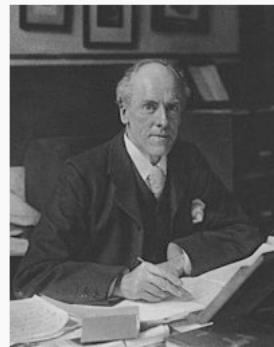
- $-\infty < \text{Cov}(X, Y) < +\infty$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, C) = 0$ se C é uma constante;
- A unidade da covariância seria: (unidade de X) * (unidade de Y);
- Comparar covariância de diferentes pares de variável é difícil, pois se alterarmos a escala, a covariância muda também.
- Como remover a escala dos dados?

Correlação de Pearson (r ou ρ)

- Karl Pearson (1896);
- Mede a intensidade da associação entre duas variáveis quantitativas;
- ρ é a covariância dos dados padronizados de X e Y;
- Adimensional (lida com proporção);
- $-1 < \rho < 1$

$$r = \frac{\text{COV}_{xy}}{S_x \times S_y} \Rightarrow r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \times \sum (y - \bar{y})^2}}$$

Karl Pearson



Nascimento	27 de março de 1857 Londres
Morte	27 de abril de 1936 (79 anos)
Nacionalidade	Britânico
Cidadania	Reino Unido
Filho(s)	Egon Pearson
<i>Alma mater</i>	Universidade de Cambridge, Universidade de Heidelberg
Ocupação	matemático, historiador da matemática, estatístico, historiador, filósofo, biógrafo, psicólogo
Prêmios	Medalha Darwin (1898), Prêmio Memorial Weldon (2012)
Empregador	University College London, Gresham College, King's College de Londres
Religião	Quaker
Orientador(es)	Francis Galton
Orientado(s)	Philip Hall
Campo(s)	Matemática

Correlação de Pearson (r ou ρ)

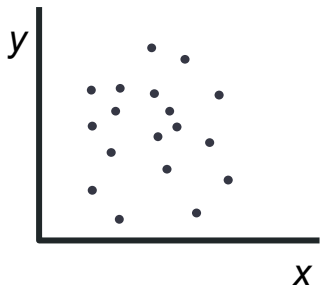
Pressuposições

- As unidades de medida são selecionadas ao acaso.
- Cada observação fornece valores para as duas variáveis.
- As variáveis medidas devem ser independentes.
- Não mede corretamente associações não-lineares (coeficiente de Spearman)

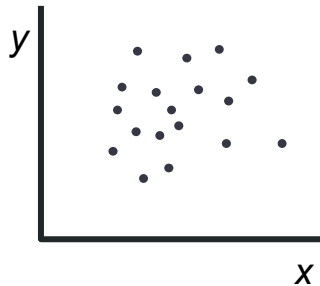
$$r = \frac{\text{COV}_{xy}}{S_x \times S_y}$$

Correlação de Pearson (r ou ρ)

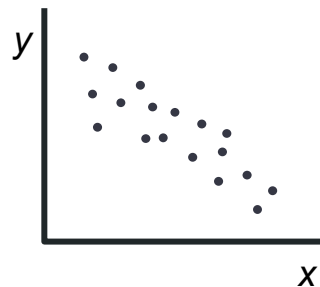
Correlação Inexistente



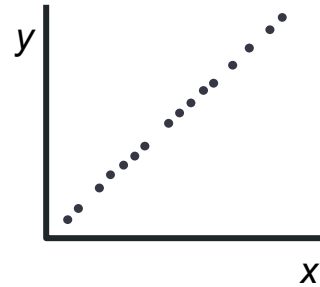
Correlação Positiva Fraca



Correlação Negativa Forte



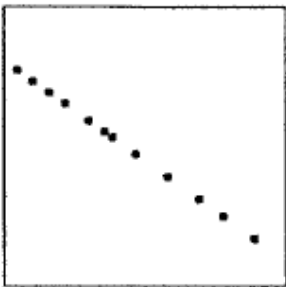
Correlação Positiva Perfeita



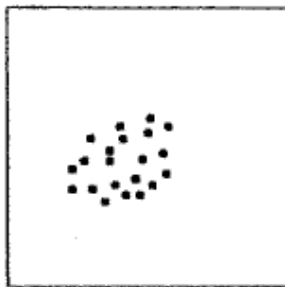
- Correlação nula: $r = 0$
- Correlação fraca:
 $0 < r < 0,3$ ou $-0,3 < r < 0$
- Correlação moderada:
 $0,3 < r < 0,6$ ou $-0,6 < r < -0,3$
- Correlação forte:
 $0,6 < r < 0,9$ ou $-0,9 < r < -0,6$
- Correlação muito forte:
 $0,9 < r < 1,0$ ou $-1,0 < r < -0,9$
- Correlação perfeita: $r = 1$



a. $r=1$



b. $r=-1$



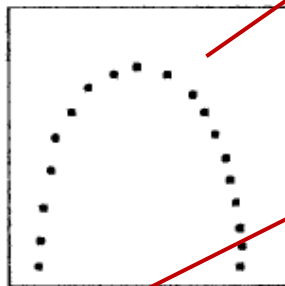
c. $r=0$



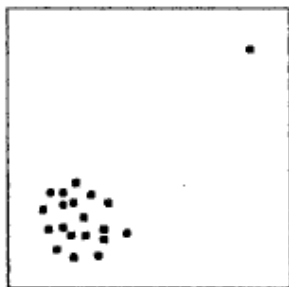
d. $0 < r < 1$



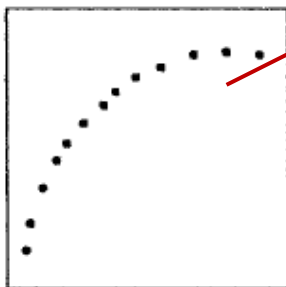
e. $-1 < r < 0$



f. $r=0$



g. $0 < r < 1$



h. $0 < r < 1$



j. $-1 < r < 0$

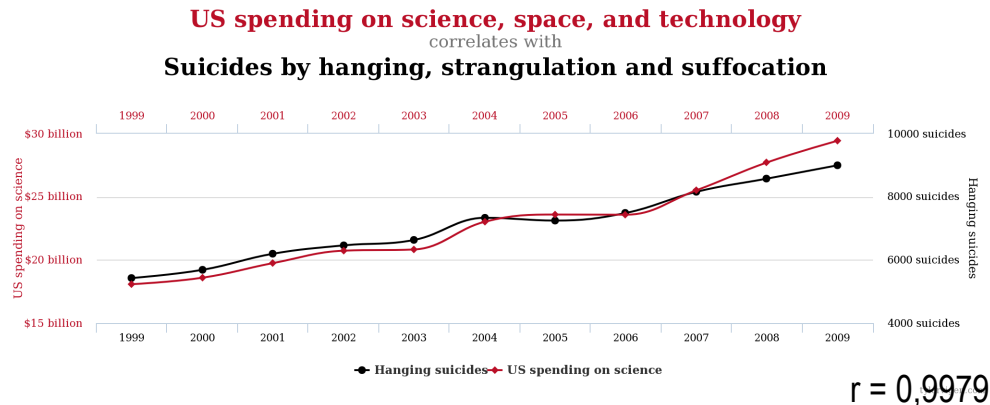
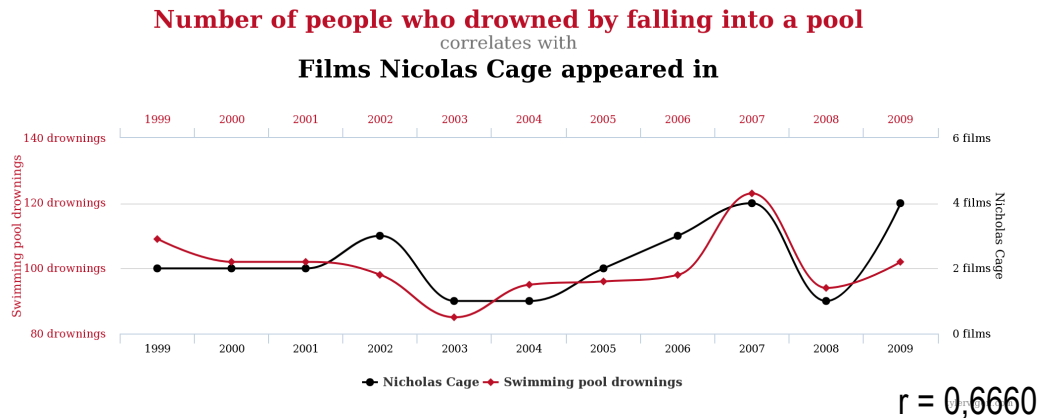
Para pontos cujo eixo principal forma uma curva, o valor de r não mede corretamente a associação entre as variáveis (não-linear).

Solução:

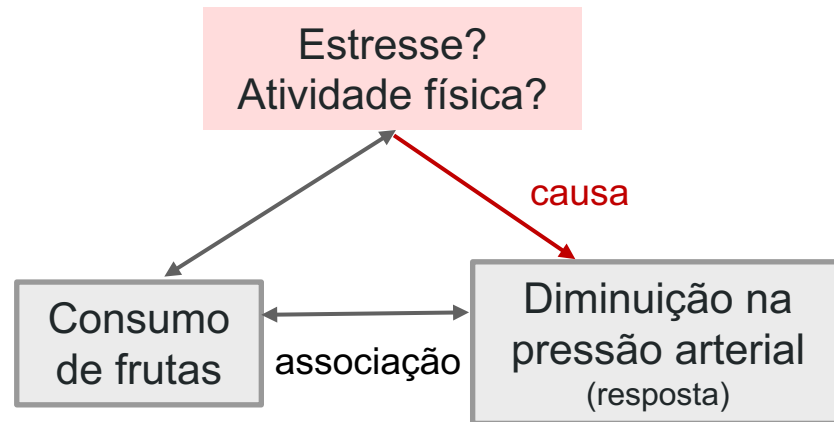
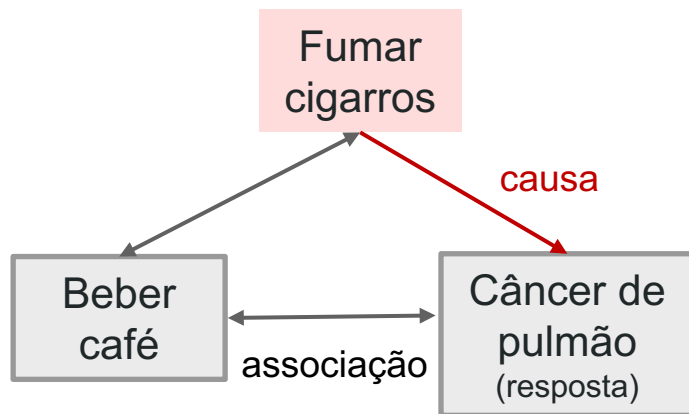
- Transformação dos dados (e.g., escala log-linear, log-log, etc);
- Uso de um coeficiente de correlação não linear (coeficiente de Spearman)

Correlação Espúria

- Fortes medidas de associação podem não ter relação causal.
- **Correlação não implica causalidade!**
- Coincidências acontecem...
- E variáveis de confusão existem!



Variável de confusão



Desenho de Experimentos

O experimento visa estabelecer uma relação causa-efeito.

Qual é o efeito da intervenção (tratamento) na resposta?

- Tratamento: variável explicativa / independente.
- Resposta: resposta / variável dependente.

Qual é o efeito de uma nova droga/vacina no tratamento/prevenção de determinada doença?

- Tratamento: nova droga.
- Resposta: número de indivíduos curados.

Experimentos Controlados

- Os participantes são separados em 2 grupos:
 - Grupo de tratamento: recebe a droga a ser testada
 - Grupo de controle: recebe placebo ou outro medicamento já conhecido.
- Os grupos devem ser comparáveis para que a causa possa ser inferida.
- Se os grupos não forem comparáveis, isso pode levar a confusão (viés)
 - Idade média do grupo de tratamento: 25
 - Idade média do grupo controle: 50
 - A idade é um potencial fator de confusão

Padrão-ouro de Experimentos

Teste controlado e aleatório

- Os grupos devem ser o mais semelhante possível, com exceção do fator tratamento.
- Os participantes são atribuídos aos grupos controle e tratamento aleatoriamente.
- Duplo-cego: o pesquisador que administra o tratamento e analisa os resultados não tem conhecimento de quem recebeu o tratamento verdadeiro ou o placebo.

Placebo

- Assemelha-se ao tratamento, mas não tem efeito.
- Os participantes não saberão em qual grupo estão.
- Em ensaios clínicos, uma pílula de açúcar garante que o efeito da droga é realmente devido à própria droga e não a ideia de receber a droga.

Na 3^a unidade...

Qual é a função / relação quantitativa
que descreve esta associação?...

Referências

- Vieira, S. *Introdução à Bioestatística*. 4ª ed. Rio de Janeiro: Elsevier, 2008.
- McDonald, J.H. 2014. *Handbook of Biological Statistics* (3rd ed.). Sparky House Publishing, Baltimore, Maryland. <http://www.biostathandbook.com/index.html>
- Mangiafico, S.S. 2015. *An R Companion for the Handbook of Biological Statistics*, version 1.3.2. rcompanion.org/rcompanion/.