



**Educational Resources / Recursos Educativos**

# **Fundamentos de Causalidade: notações, complicações e soluções**

**Luís Cavique**  
DCeT, Univ. Aberta  
Luis.Cavique@uab.pt

**Lisboa, abril 2025**



Este trabalho está licenciado com uma Licença Creative Commons  
Attribution-NonCommercial-ShareAlike CC BY-NC-SA

## **Resumo**

Este documento pretende complementar a bibliografia da unidade curricular Extração de Conhecimento de Dados (Data Mining) oferecida nos cursos de mestrado MEIW (Mestrado em Engenharia Informática e Tecnologia Web) e MBB (Mestrado em Bioestatística e Biometria).

Para além do debate Rubin versus Pearl que está longe de ser resolvido, a primeira dificuldade em causalidade é a notação.

Quando iniciamos o estudo da causalidade são muitos os novos problemas com que nos deparamos. Neste trabalho vamos dividi-los em três grupos: os confundidores associados aos DAGs, os paradoxos nas intervenções com estratos e os contrafactuais da inferência causal.

No final propomos duas ferramentas com vista a encontrar soluções em ambientes causais.

### **Conteúdos**

1. Notação
2. Introdução histórica
3. Problemas causais
  - 3.1. Confundidores e mediadores
  - 3.2. Paradoxos
  - 3.3. Contrafactuais
4. Ferramentas

Neste trabalho, depois de introduzir a notação causal, apresentamos um resumo histórico da causalidade. De seguida tratamos vários problemas associados como as variáveis confundidoras, os paradoxos causais e os mundos contrafactuais. Finalmente, introduzimos as ferramentas que irão fornecer soluções, a descoberta causal e a inferência causal.-

## 1. Notação

Os modelos analíticos descobrem padrões interessantes e fazem previsões com base em bons ajustes de dados históricos. Nos modelos baseados em dados, emergem movimentos recentes em favor de modelos explicativos. A diferença entre correlação e causalidade está no centro da controvérsia sobre previsão e causalidade/explicação. Na ciência de dados, duas tarefas devem ser distinguidas: previsão e explicação.

Na previsão, duas variáveis são usadas: a variável independente  $X$  e a variável dependente  $Y$ . Em previsão, os dados originais são divididos em conjuntos de dados de treino (ou de aprendizagem) e de teste para encontrar a função  $Y=f(X)$ , onde  $X$  é uma variável independente e  $Y$  é o resultado.

Quando falamos de causalidade, um novo tipo de variável deve ser incluído: a intervenção/tratamento  $T$ . Neste trabalho iremos usar a notação definida por Rubin [1974].

O resultado  $Y$  com base no tratamento  $T$  é o objeto do estudo, onde resultado  $Y$  é uma variável dependente e  $T$  da variável independente. Para tanto, são utilizados conjuntos de dados de teste e controle para tratamento realizado,  $T=1$ , e não realizado,  $T=0$ . A função explicativa utiliza três variáveis,  $Y=f(T, X)$  onde  $X$  são as variáveis de controlo ou covariáveis cujo efeito se procura neutralizar com vista a impedir que interfira entre as variáveis independentes e dependentes. Utiliza-se frequentemente a expressão  $Y = f(T|X)$  onde a ideia é isolar o efeito de  $T$  sobre  $Y$  mantendo  $X$  constante ('ceteris paribus').

Seja a relação de causa ( $T$ ) e efeito ( $Y$ ) representa por  $T \rightarrow Y$ . A covariável  $X$  ao influenciar as duas variáveis  $T$  e  $Y$  representada no DAG ('direct acyclic graph') da Figura 1. perturba a relação  $T \rightarrow Y$ . O que se pretende é controlar  $X$  e estudar o efeito de  $T$  sobre  $Y$ .

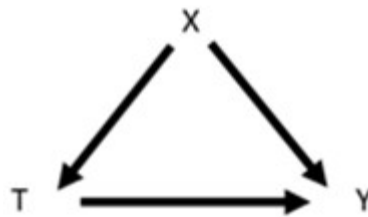


Figura 1. DAG do triplo  $(T, Y, X)$

A intervenção corresponde a um ensaio experimental com grupos de teste e controlo escolhidos aleatoriamente (RCT, 'random controlled trial'). A intervenção responde a perguntas como:

- 'qual é o efeito de  $T=1$  sobre  $Y$ ?' (what if  $T=1$ ?)
- 'porque que é que ocorre  $Y=1$ ?' (why  $Y=1$  occur?)

O efeito do tratamento ( $T$ ) sobre o resultado ( $Y$ ) pode ser expresso como a diferença entre os resultados potenciais quando o tratamento é aplicado ( $Y^{T=1}$  ou  $Y^1$ ) e quando não é

aplicado ( $Y^{T=0}$  ou  $Y^0$ ). A referida diferença é conhecida como Efeito Médio de Tratamento (ATE, 'average treatment effect'), geralmente utilizando a notação de valor esperado:

$$ATE = E[Y^1 - Y^0]$$

Nesta introdução pretende-se deixar claro que existem diferenças significativas entre a predição e a causalidade, que vão para além da notação.

A predição pode utilizar uma vasta gama de técnicas de aprendizagem automática para otimizar as previsões. A causalidade, por outro lado, tem como base as intervenções (ou tratamentos) realizados em experiências aleatórias controladas também conhecido por testes A/B, ou suportada por dados observacionais.

A escolha entre focar em predição ou causalidade depende do objetivo da análise: prever o que irá acontecer ou encontrar as variáveis para fazer acontecer. Segundo Ackoff [1981] no planeamento ativo, podem ser estabelecidas duas abordagens: a 'pré-ativa', que prevê o futuro, e a proativa, que cria o futuro. A causalidade tem o potencial de antecipação do futuro depois de encontrar as variáveis das intervenções. A causalidade permite encontrar as variáveis acionáveis nas organizações, que são a chave para o Business Experimentation.

## 2. Introdução histórica

A história da causalidade na filosofia e nas ciências inclui várias contribuições de muitos pensadores ao longo dos séculos. Podemos destacar as contribuições de David Hume, John Stuart Mill, Bertrand Russell, Donald Rubin e Judea Pearl.

David Hume (1711-1776), um filósofo escocês do século XVIII, é frequentemente associado ao ceticismo. David Hume apresenta dois conceitos para a causalidade: a regularidade e a ordenação temporal. Se sempre que ligo o interruptor a lâmpada acende, existe regularidade. Por outro lado, deve existir ordenação temporal: primeiro ligo o interruptor e de seguida a lâmpada acende. Na nossa experiência mostra que as causas vêm primeiro e os efeitos depois. Essa ideia é conhecida como o princípio da "conexão necessária", mas Hume argumentou que essa conexão é mais uma questão de hábito mental do que uma verdade objetiva observável.

John Stuart Mill (1806-1873), um filósofo e economista inglês do século XIX, trata a causalidade com base nos seus famosos métodos da Lógica de Mill. Os quatro principais métodos são: a concordância, ou causa necessária do efeito; a diferença, ou causa suficiente do efeito; a concordância e diferença, ou causa necessária e suficiente do efeito e a variação concomitante, com uma variação da causa obtém-se o uma variação no efeito. O método de diferença afirma que, se num fenómeno os casos são idênticos em todos os aspetos, exceto em um, o aspeto pelo qual os casos diferem é a causa do fenómeno. Esta abordagem forneceu uma forma mais sistemática e empírica de identificar relações causais.

Avançando para o início do século XX, Bertrand Russell (1872-1970), conhecido pelos trabalhos em lógica moderna, abandonou o conceito de causalidade. No ensaio "on the notion of cause", escrito entre 1912 e 1913, refere que em ciência avançada, tal como a física, a palavra 'causa' nunca ocorre.

Em contrapartida, Donald Rubin (1943-) contribuiu significativamente para a teoria estatística da causalidade. Ele desenvolveu o "Modelo de Rubin de Causalidade" nos anos de 1970, que usa um conjunto de variáveis e pressupostos para avaliar o efeito causal de tratamentos. Esse modelo tornou-se fundamental na estatística, permitindo a análise mais rigorosa de dados experimentais e observacionais para inferir causalidade.

Finalmente, Judea Pearl (1936-), um cientista da computação do século XX e XXI, tem vindo a enfatizar a importância da causalidade com a reintrodução dos grafos causais, os DAG ('direct acyclic graphs'). Os DAGs permitem aos analistas visualizar e manipular relações causais complexas. Pearl propôs que, com a ajuda de modelos de estrutura causal ('structural causal models') e determinadas regras de cálculo, é possível deduzir relações causais de dados complexos.

Os DAGs são uma ferramenta útil na representação causal, eles descrevem as suposições causais de cada estudo [Pearl, Glymour 2016]. Os vértices correspondem às variáveis (tratamento T, covariáveis X e resultado Y) e as setas são a eventual associação entre os nós. No estudo de causalidade, a descrição dos dados e o DAG devem ser apresentados antes da fase de modelagem. As suposições representadas nos diagramas causais são tiradas antes das conclusões [Hernán 2017].

Judea Pearl contribuiu ainda para uma nova visão da causalidade do ponto de vista dos cientistas da computação, resumida pela escada da causalidade [Pearl, Mackenzie 2018]. Ele argumenta que a inteligência artificial ainda não domina as relações de causa-efeito.

Pearl [2019] propõe três níveis de causalidade: associação, intervenção e contrafactual. A associação não tem consequências causais, contrariamente à intervenção e aos contrafactuais. A associação corresponde à abordagem preditiva  $Y=f(X)$ , usual em Machine Learning. A intervenção é exemplificada pelo teste A/B, onde o tratamento T aparece na equação  $Y=f(T, X)$ . Finalmente, os dados contrafactuais (ou factos em falta) envolvem mundos imaginários e abordagens específicas que comparam tratamentos. Os diferentes níveis estão associados a questões específicas, como segue:

- Associação: qual é a relação entre X e Y?
- Intervenção: e se fizerem o tratamento T?
- Contrafactuais: e se tivessem agido de forma diferente? (T=1 em vez de T=0)

Da mesma forma, o trabalho de Hernán et al. [2019] defende três tarefas em ciência de dados: descrição, previsão e inferência causal, que são aplicadas a um programa de treinamento em ciência de dados. Os alunos aprendem a diferenciar as três tarefas e depois geram e analisam dados para cada tarefa. Os alunos também aprendem a fazer perguntas científicas para cada tarefa.

### 3. Problemas causais

Quando iniciamos o estudo da causalidade são muitos os novos problemas com que nos deparamos. Neste trabalho vamos dividi-los em três grupos: os confundidores associados aos DAGs, os paradoxos nas intervenções com estratos e os contrafactuais da inferência causal.

#### 3.1. Confundidores e mediadores

Quando se inicia o estudo da causalidade é vulgar ouvir que "correlação não implica causalidade". Surgem imediatamente duas perguntas: porque é que correlação não implica causalidade? e o que é então a causalidade?

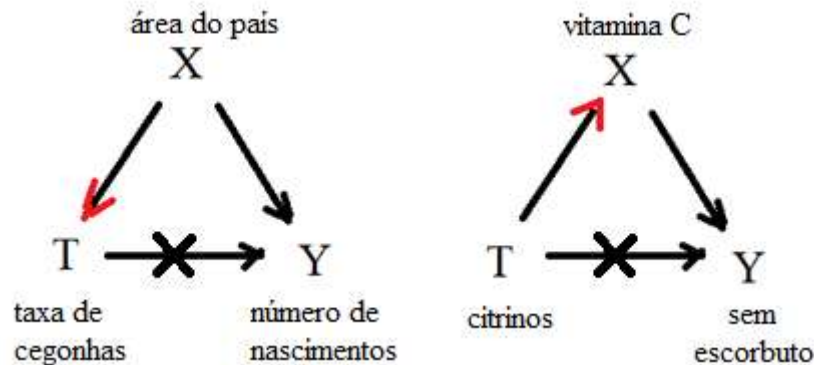
Existe uma forte correlação entre homens casados e uma boa saúde, o que nos leva a inferir uma relação causal de homem\_casado  $\rightarrow$  boa\_saúde. Contudo, dada a correlação existem é também possível inferir o inverso, por ter boa saúde é a causa de o homem ser casado. A correlação não explica o sentido da causalidade, sendo este um dos problemas iniciais da teoria da causa e efeito.

Um segundo exemplo mostra que existe uma forte correlação entre a distância Júpiter-Sol e o número de secretárias no Alasca. Não existido qualquer relação causa-efeito entre as duas variáveis baseada na lógica, ou na física, dizemos que se trata de uma correlação espúria. Este e outras centenas de exemplos podem ser encontrados em [http://www.tylervigen.com/spurious-correlations\\_](http://www.tylervigen.com/spurious-correlations_)

Um terceiro exemplo mostra que existe uma forte correlação do número de cegonhas com a taxa de nascimento de bebés num vasto conjunto de países. Será que são as cegonhas que trazem os bebés? Neste caso existe uma terceira variável que afeta a taxa de nascimento e o número de cegonhas, essa variável é a área de cada país. Este exemplo é explicado pelo Princípio da Causa Comum. O Princípio da Causa Comum de Reichenbach [1956] afirma que se duas variáveis, A e B, estão correlacionadas, isso implica que A influencia causalmente B (ou vice-versa) ou há uma variável C que influencia causalmente A e B. A variável C diz-se um "confundidor" da relação entre A e B.

Um quarto caso prático refere-se aos estudos do médico James Lind, no século XVIII, relativamente a uma doença comum nos marinheiros, o escorbuto. O médico experimentou 12 tratamentos nos marinheiros com diferentes ácidos, como vinagre, cidra e citrinos, chegando à conclusão os citrinos melhoravam a saúde dos pacientes. Foi então estabelecida a relação citrinos  $\rightarrow$  anti-escorbuto sem ter uma explicação clara do sucesso da experiência. Só nos anos de 1930 com a descoberta a vitamina C foi encontrada a explicação. A vitamina C é um "mediador" [Baron, Kenny 1986]. entre as variáveis iniciais.

A Figura 2. apresenta dos DAGs dos exemplos três (cegonhas) e quarto (escorbuto). O confundidor e o mediador diferem pela diferença do sentido de uma das arestas.



**Figura 2.** DAG com confundidor (esquerda) DAG com mediador (direita)

Voltando às perguntas iniciais, porque é que correlação não implica causalidade e o que é a causalidade, exemplificámos que correlação não implica causalidade. Contudo, deve-se realçar para existir causalidade deve existir correlação. Já a definição de causalidade é mais difícil de expressar para N variáveis. Para três variáveis correlacionadas a causalidade é explicada pelo Princípio da Causa Comum [Reichenbach 1956] e pela análise da mediação [Baron, Kenny 1986].

### 3.2. Paradoxos

Consideremos o caso de um tratamento T (T=1 tratamento; T=0 placebo), com um resultado potencial Y, que indica se o paciente recuperou (Y=1) ou não (Y=0), e ainda uma condição binária C (C=1 homem; C=0 mulher). Para cada combinação dos valores binários de (T, Y, C) foi encontrado o número de pacientes.

Na Tabela 1. apresenta dos resultados potenciais dos dados agregados e com estratificação. O valor do efeito dos dados agregados tem um valor negativo indicando que a taxa de recuperação aumenta com o placebo. Nos dados estratificados verifica-se o contrário, o efeito é positivo indicando que a taxa de recuperação aumenta com o tratamento. Os dados parecem paradoxais, sendo explicados pelo paradoxo de Simpson.

O paradoxo de Edward Simpson, descrito em 1951, é um fenómeno estatístico no qual uma tendência que aparece em diferentes grupos de dados desaparece quando esses grupos são combinados. O resultado aparentemente contraditório ocorre devido às diferentes dimensões dos grupos, influenciando a análise combinada.

A solução para o paradoxo de Simpson envolve uma análise cuidadosa da estratificada dos dados para garantir que as conclusões sejam válidas sem distorcer os valores quando agrupados. A expressão para os dados estratificados é a seguinte:

$$E(Y^t) = \sum [E(Y^t | C=c) \cdot \Pr(C=c)]$$

Da mesma forma a expressão do do-calculus de Pearl para dados estratificados é a seguinte:  $E[Y | \text{do}(T=t)] = \text{Sum} [E(Y|T=t, C=c) \cdot \Pr(C=c)]$ .

Encontradas as probabilidades de  $P(C=1) = 51,00\%$  e  $P(C=0) = 49,00\%$ , podemos aplicar a expressão para os dados estratificados. Os valores dos resultados potenciais são  $E(Y^1) = 83,25\%$  e  $E(Y^0) = 77,89\%$ , obtendo um efeito positivo de  $E(Y^1 - Y^0) = 5,37\%$ , que está de acordo com os dados estratificados em que 'o tratamento é bom'.

**Tabela 1.** Resultados potenciais agregados e com estratificação

agregado				
	Y=1	Y=0	soma	
T=1	273	77	350	$E(Y^1) = 78.00\%$
T=0	289	61	350	$E(Y^0) = 82.57\%$
			700	$E(Y^1 - Y^0) = -4.57\%$
				o tratamento é mau

estrato C=1				
	Y=1	Y=0	soma	
T=1	81	6	87	$E(Y^1 C=1) = 93.10\%$
T=0	234	36	270	$E(Y^0 C=1) = 86.67\%$
			357	$E(Y^1 - Y^0 C=1) = 6.44\%$
				o tratamento é bom

estrato C=0				
	Y=1	Y=0	soma	
T=1	192	71	263	$E(Y^1 C=0) = 73.00\%$
T=0	55	25	80	$E(Y^0 C=0) = 68.75\%$
			343	$E(Y^1 - Y^0 C=0) = 4.25\%$
				o tratamento é bom

### 3.3. Contrafactuais

Na origem dos dados para o estudo da causalidade, para além dos ensaios clínicos aleatórios (RCT, 'mice&dice') ou a sua versão simplificada com os testes A/B, existem os estudos observacionais introduzidos por Rubin [1974]. Os estudos observacionais estudam os efeitos das intervenções em configurações naturais sem manipulação por parte do investigador, úteis para questões éticas ou logisticamente complexas onde os RCTs não são possíveis.

O Problema Fundamental da Inferência Causal [Rubin 1974; Holland 1986] afirma que não podemos observar simultaneamente todos os resultados potenciais de uma unidade. Por exemplo, não podemos observar simultaneamente o resultado da saúde de um paciente que recebeu tratamento e a saúde do mesmo paciente que não recebeu o tratamento. Se o paciente recebeu efetivamente o tratamento, o não tratamento é contrafactual e vice-versa. Os contrafactuais ajudam-nos a compreender o efeito de um tratamento, comparando o que aconteceu quando a intervenção foi aplicada com o que teria acontecido se não tivesse sido aplicada. O filósofo David Lewis [1973] propõe que os contrafactuais podem ser analisados por meio de "mundos possíveis", uma ideia conhecida na lógica modal.

### Resultados potenciais

Convém definir resultado factual, contrafactual e potencial. Resultado factual é o resultado observado do paciente que ocorreu para o tratamento observado. Resultado contrafactual é o resultado não observável do paciente, que teria ocorrido se o paciente

tivesse recebido um tratamento diferente. Resultado potencial é o resultado hipotético do paciente que seria observado se um determinado tratamento fosse administrado.

O efeito de um tratamento num indivíduo é a comparação entre os resultados potenciais e uma alternativa  $Y^1$  versus  $Y^0$ . Podemos definir três possibilidades, embora neste trabalho iremos só desenvolver a primeira delas:

- Diferença entre os resultados potenciais:  $Y^1 - Y^0$
- Rácio dos resultados potenciais:  $Y^1 / Y^0$
- Diferença relativa dos resultados potenciais:  $(Y^1 - Y^0) / Y^0$

Para exemplificar as métricas dos efeitos em ambientes contrafactuais vamos utilizar os dados e a abordagem referida em Cunningham [2021]. Seja a variável Y o número de anos de vida do paciente, e o tratamento  $T=1$  ou  $T=0$ , respetivamente cirurgia ou quimioterapia.

Na Tabela 2. apresenta os dados disponíveis relativos a 10 pacientes. No conjunto de dados observados existem vários valores em falta ('missing values'). Como foi referido, pelo Princípio Fundamental da Inferência Causal, para o mesmo paciente não podemos observar os resultados dos dois tratamentos.

O ATE ('average treatment effect') é dado por  $ATE = E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$ , contudo o que temos são os valores de  $E(Y^1|T=1)$  e  $E(Y^0|T=0)$ , pelo só nos é possível calcular um valor aproximado, o NATE ('naïve average treatment effect'), dado por:

$$NATE = E(Y^1|T=1) - E(Y^0|T=0)$$

**Tabela 2.** Dados dos pacientes e cálculo do NATE

paciente	Y	T	$Y^1 T=1$	$Y^0 T=0$	$Y^1 - Y^0$
1	7	1	7		
2	6	0		6	
3	5	1	5		
4	8	0		8	
5	4	1	4		
6	10	1	10		
7	10	0		10	
8	6	0		6	
9	7	0		7	
10	9	1	9		
média			7	7.4	-0.4
			$E(Y^1 T=1)$	$E(Y^0 T=0)$	NATE

Na Tabela 3. apresenta os dados dos pacientes e cálculo do ATE, com base em valores contrafactuais, com fundo colorido. Note-se que o valor de NATE difere do valor de ATE, tendo mesmo sinais opostos.

São ainda apresentados o ATT ('average treatment effect of the treated') e o ATU ('average treatment effect of the untreated') dados por:  $ATT = E(Y^1 - Y^0|T=1)$  e  $ATU = E(Y^1 - Y^0|T=0)$ .

**Tabela 3.** Dados dos pacientes e cálculo do ATE, ATT, ATU

paciente	Y <sup>1</sup>	Y <sup>0</sup>	Y <sup>1</sup> -Y <sup>0</sup>	Y <sup>1</sup> -Y <sup>0</sup>  T=1	Y <sup>1</sup> -Y <sup>0</sup>  T=0
1	7	1	6	6	0
2	5	6	-1	0	-1
3	5	1	4	4	0
4	7	8	-1	0	-1
5	4	2	2	2	0
6	10	1	9	9	0
7	1	10	-9	0	-9
8	5	6	-1	0	-1
9	3	7	-4	0	-4
10	9	8	1	1	0
média	5.6	5	0.6	4.4	-3.2
	E(Y <sup>1</sup> )	E(Y <sup>0</sup> )	E(Y <sup>1</sup> -Y <sup>0</sup> )	E(Y <sup>1</sup> -Y <sup>0</sup>  T=1)	E(Y <sup>1</sup> -Y <sup>0</sup>  T=0)
			ATE	ATT	ATU

Finalmente, vamos explorar uma situação ideal que emprega o NATE para estimar o ATE utilizando simulação. O método de Monte Carlo é aplicado 2.000.000 de vezes, onde em cada iteração é calculado o NATE médio sob a condição de independência, i.e. garantida pela escolha aleatória de 5 números dos 10 existentes. Na simulação realizada recorrendo a um programa em C, o NATE foi, em média, de 0,595013, bem próximo do valor de ATE igual a 0,6.

### Pressupostos da inferência causal

Definimos quatro pressupostos necessários para estimar os efeitos causais: positividade, consistência, permutabilidade e não-interferência.

**Positividade:** Se por absurdo, o grupo de teste, T=1, ou o grupo de controlo, T=0, tem toda a população não é possível calcular a média do tratamento. Cada paciente tem probabilidade maior que zero de receber, ou não receber, um tratamento. Isso também é chamado de pressuposto da sobreposição.

- $P(T=1|X=x) > 0$
- $P(T=0|X=x) > 0$

**Consistência:** O resultado potencial é igual ao resultado observado do paciente no tratamento selecionado, o que implica que os resultados potenciais são claramente definidos e observáveis. Este pressuposto garante que não existe variação oculta.

- se T=0 então  $Y^0=Y$
- se T=1 então  $Y^1=Y$

**Permutabilidade:** Dadas as covariáveis observadas X, a atribuição do tratamento T é independente dos resultados potenciais Y. Isto é posso permutar o grupo de teste com o grupo de controlo, mantendo os mesmos resultados potenciais. Essa suposição também é chamada de Inconfundibilidade ou Ignorabilidade.

- $\{Y^0, Y^1\} \perp T$

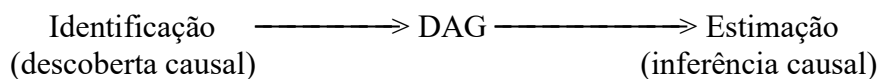
Não-interferência: O resultado para qualquer paciente não depende da atribuição de tratamento de outros pacientes.

SUTVA ('Stable Treatment Value Assumption') combina os pressupostos de não-interferência e consistência. O resultado para qualquer paciente não depende da atribuição de tratamento de outros pacientes, e não há variação oculta no efeito do tratamento em diferentes ambientes ou populações.

#### 4. Ferramentas

Depois de apresentar as novas notações e várias complicações (confundidores, paradoxos e contrafactuais) esperam-nos algumas soluções. Nesta subsecção abordamos duas ferramentas fundamentais para lidar com a causalidade, a descoberta causal e a inferência causal.

Vamos seguir a dicotomia de identificação seguida da estimação, desenvolvida para a Análise de Caminhos ('path analysis') de Heise [1975]. A identificação das relações causais é encontrada por técnicas de descoberta causal e a estimação dos efeitos é calculada usando a inferência causal. Esquemáticamente teremos:



A descoberta causal refere-se à identificação de relações causais, ou dependências entre variáveis, a partir de dados observacionais sem conhecimento prévio da estrutura causal subjacente. O seu objetivo é descobrir as ligações causais subjacentes entre variáveis com base em padrões estatísticos e dependências observadas nos dados. A arquitetura das relações causais é geralmente representada por um DAG. Ver mais em [Cavique 2024].

A inferência causal refere, por outro lado, envolve a utilização de conhecimentos prévio sobre relações causais para estimar os efeitos causais de tratamentos. Centra-se na estimativa do efeito causal da variável de tratamento T no resultado potencial Y, considerando a estrutura causal que já é conhecida previamente. Ver mais em [Cavique 2023].

A descoberta causal e a inferência causal funcionam sequencialmente. As técnicas de descoberta causal podem ajudar a identificar possíveis relações causais entre variáveis, que podem ser usadas como entradas para modelos de inferência causal. A inferência causal, por outro lado, depende do DAG e dos dados experimentais para estimar os efeitos causais.

## REFERÊNCIAS

Ackoff R. (1981), *Creating the Corporate Future: Plan or be Planned*, John Wiley, New York.

Baron R.M., D.A. Kenny (1986), The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology*, vol. 51 (6), pp. 1173–1182, <https://doi.org/10.1037/0022-3514.51.6.117>.

Cavique L. (2023), Causality: The Next Step in Artificial Intelligence, in *Philosophy of Artificial Intelligence and Its Place in Society*, Eds. L. Moutinho, L. Cavique, E. Bigné, IGI-Global, (ISSN: 2328-1316; eISSN: 2328-1324), pp.1-17.

Cavique L. (2024), *Causal discovery: an introduction*, publicação Repositório Aberto.

Cunningham S. (2021), *Causal Inference: The Mixtape*, Yale University Press, ISBN-13: 978-0300251685, <https://mixtape.scunning.com/>.

Heise D.R (1975), *Casual Analysis*, John Wiley & Sons Inc, ISBN-13: 978-0471368984.

Hernán M.A. (2017), *Causal Diagrams: Draw Your Assumptions Before Your Conclusions*, Available: <https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your>. [Accessed: 12-Nov-2021].

Hernán M.A., J. Hsu, B. Healy, (2019), A second chance to get causal inference right: a classification of data science tasks, *Chance*, vol. 32(1), pp. 42–49, <https://doi.org/10.1080/09332480.2019.1579578>

Holland P.W. (1986), Statistics and Causal Inference, *Journal of the American Statistical Association*, vol. 81 (396), pp. 945–960.

Lewis D.K. (1973), *Counterfactuals*, Blackwell, Malden, Massachusetts, DOI 10.2307/2273738.

Pearl J. (2000), *Causality: models, reasoning, and inference*, Cambridge University Press.

Pearl J. (2019), The seven tools of causal inference, with reflections on machine learning, *Communications of the ACM*, March, vol. 62 n. 3, pp. 54-60.

Pearl J., D. Mackenzie (2018), *The Book of Why: The New Science of Cause and Effect*, Basic Books, NY, United States, ISBN-13: 978-0465097609.

Pearl J., M. Glymour, N.P. Jewell (2016), *Causal Inference in Statistics: A Primer*, Wiley, ISBN-13: 978-1119186847.

Reichenbach H. (1956), *The Direction of Time*, University of California Press, Los Angeles.

Rubin D.B. (1974), Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, vol. 66(5), pp. 688–701, <https://doi.org/10.1037/h0037350>.