

Pillole di R

Lezione 2



Causa ed effetto



Ciao !

Sono ***Jacopo Cerri***

Borsista di ricerca presso l'Università di Torino

Per dubbi, insulti etc... jacopocerri@gmail.com

1

Un concetto complesso

Di cosa parliamo quando parliamo di causa ed effetto?





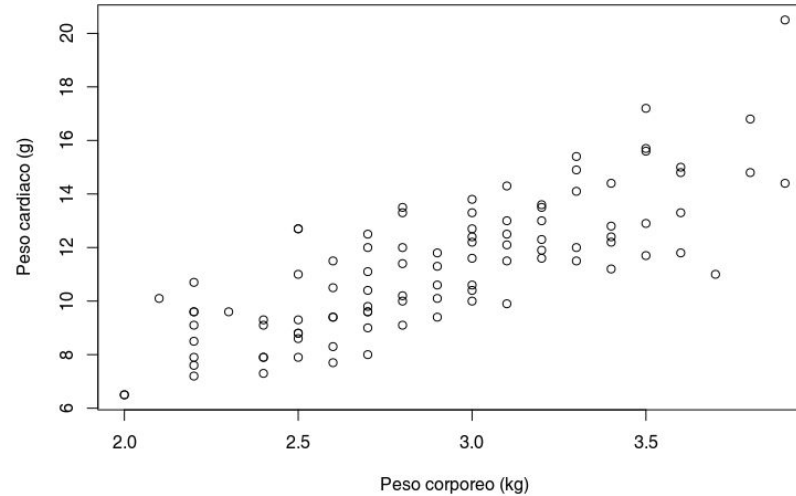
La **correlazione** di Pearson

- Forza dell'associazione
- Variabili continue X ed Y
- Da -1 a +1
- Lineare

È il rapporto tra la covarianza e il prodotto delle deviazioni standard delle due variabili.



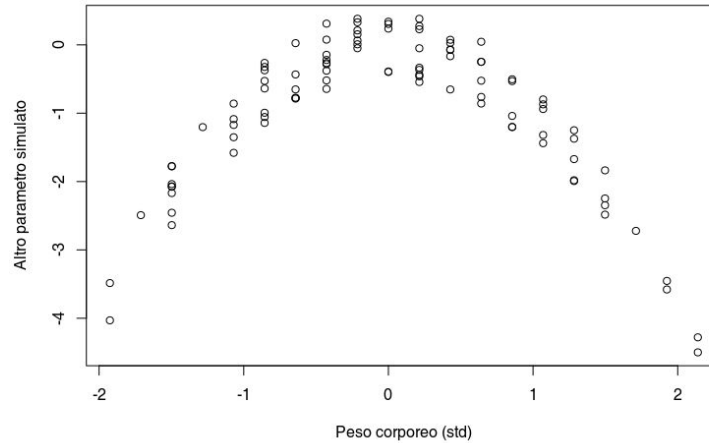
La correlazione di Pearson



Nel caso dei nostri gatti il coefficiente è 0.79, quindi “abbastanza”.



La correlazione di Pearson

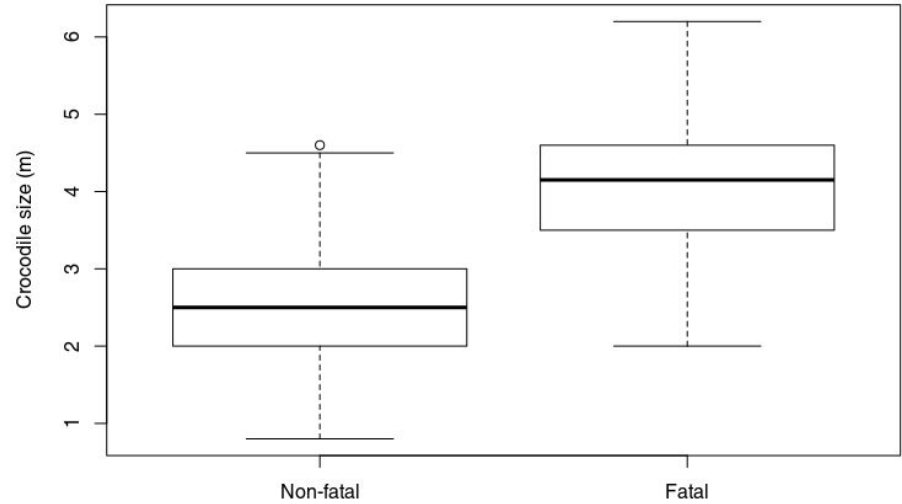


Non va bene per relazioni non lineari (-0.12)



La correlazione di Pearson

Nel caso dei coccodrilli, una correlazione di 0.64 ci porta a pensare che la taglia dell'animale influenzi la sopravvivenza.

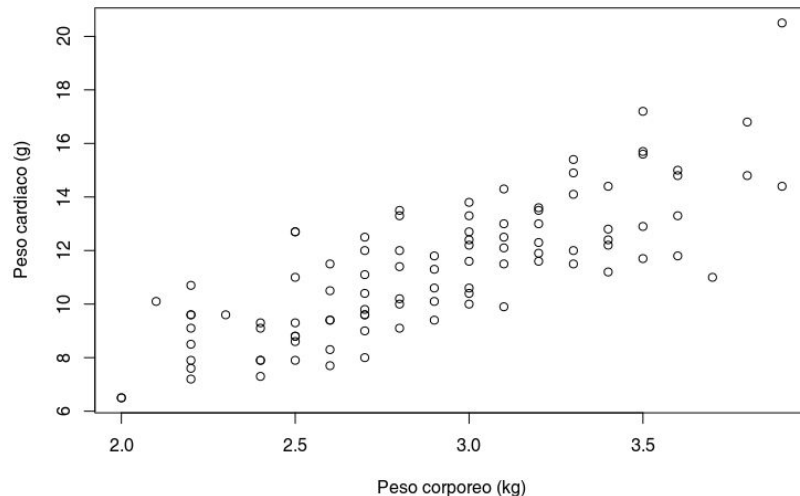




La correlazione di Pearson

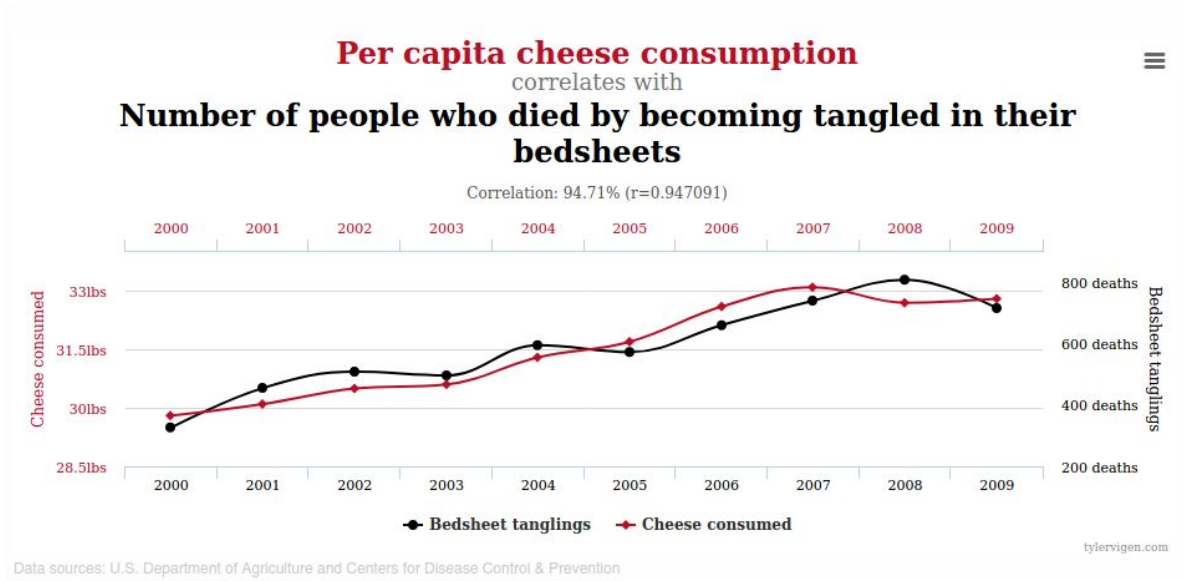
Anche un grosso gatto tende ad avere un cuore più grande.

Ma forse il corpo non influenza direttamente il cuore.





La correlazione di Pearson



In un rapporto di causa effetto

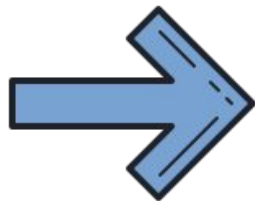
$X \rightarrow Y$

*Modificando il valore di X, ci si
aspetta che il valore di Y cambi.*

“



Causa ed effetto





Alcuni problemi

- Devo immaginarmi ciò che non è successo
- Non sempre si può manipolare
- Verificare e stimare gli effetti causali

Effetto = osservato - controfattuale

2

Il caso facile

Inferenza causale tramite prove randomizzate



Un'idea semplice

- Voglio testare l'effetto di una dieta aggiuntiva sul peso corporeo di pulcini di fagiano di 20 gg
- Prendo n pulcini
- Li divido in due gruppi in modo casuale
- Un gruppo riceve la dieta e uno no

Emergono sostanzialmente due domande.



Due domande

- La dieta ha un effetto causale sul peso corporeo dei pulcini ?
- I pulcini a cui ho dato la dieta erano abbastanza ?

Alla prima domanda si risponde con l'*Average Treatment Effect*.



Due domande

- La dieta ha un effetto causale sul peso corporeo dei pulcini ?
- I pulcini a cui ho dato la dieta erano abbastanza ?

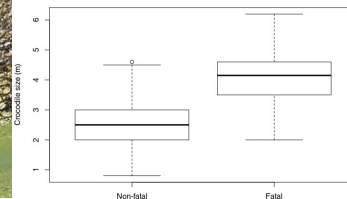
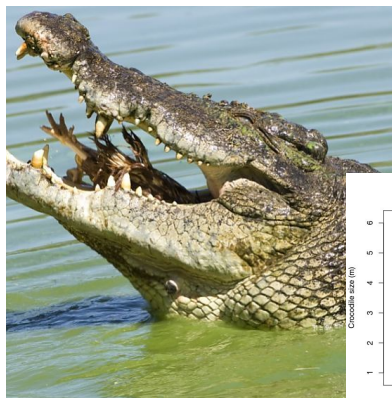
Alla prima domanda si risponde con l'*Average Treatment Effect*.



La randomizzazione

Nelle prove sperimentali, la randomizzazione consente di creare il controfattuale.

Ci sono cose che non si possono manipolare.



3

Il caso difficile

Inferenza causale su dati osservazionali: criterio back-door

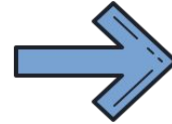
La correlazione
non implica causalità

“



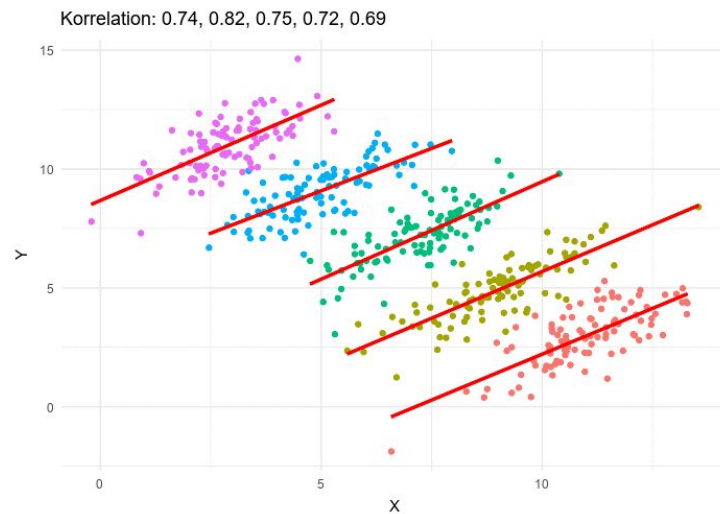
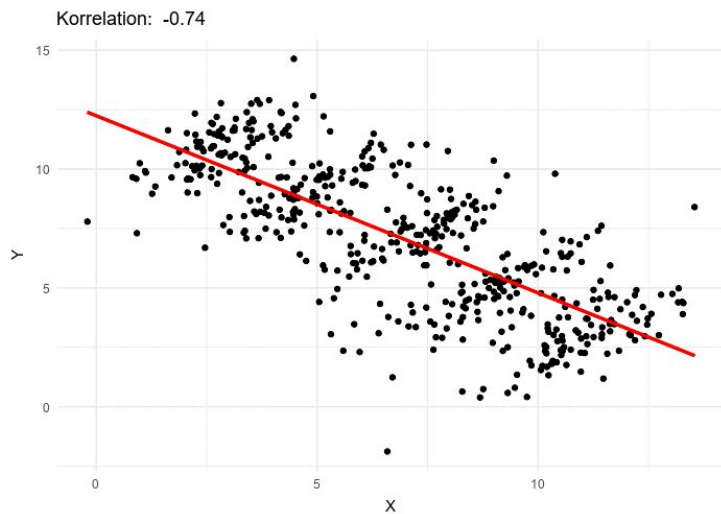
Le associazioni

In particolari condizioni, un'associazione tra due variabili implica una loro correlazione causale.





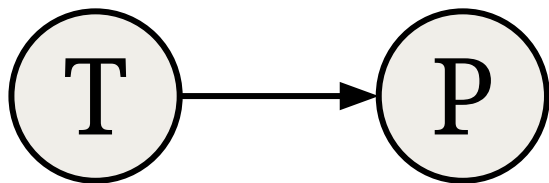
Il paradosso di Simpson





X -> **Y**

Nel caso dei coccodrilli posso ipotizzare che la taglia influenzi la probabilità di sopravvivere all'attacco.



	Vivi	Morti
< 2 metri	10	0
2 - 3 metri	74	18
3 - 4 metri	10	11
> 4 metri	4	51



Criteri di Bradford Hill

Evidenza diretta:

- Magnitudo dell'effetto
- Prossimità spaziale-temporale
- Reattività al dosaggio (reversibile)

Evidenza meccanicistica:

- C'è un meccanismo

Evidenza parallela:

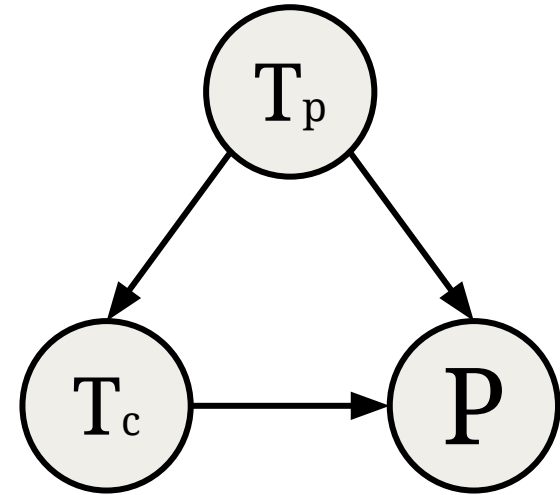
- L'effetto torna con ciò che è noto
- L'effetto si trova in studi replicati o simili



X -> **Y**

Ma la taglia delle persone influenza sia la taglia dei coccodrilli, che la probabilità di sopravvivenza.

È un fenomeno noto come confounding, all'origine del paradosso di Simpson.

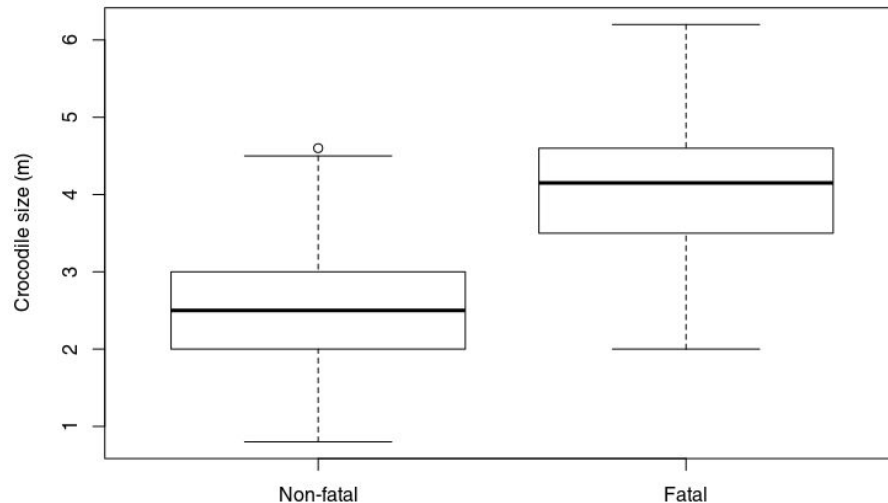




Confounding

Il confounding in questo caso può non rendere comparabili i due gruppi di osservazione.

I due gruppi potrebbero avere persone di dimensione diversa.

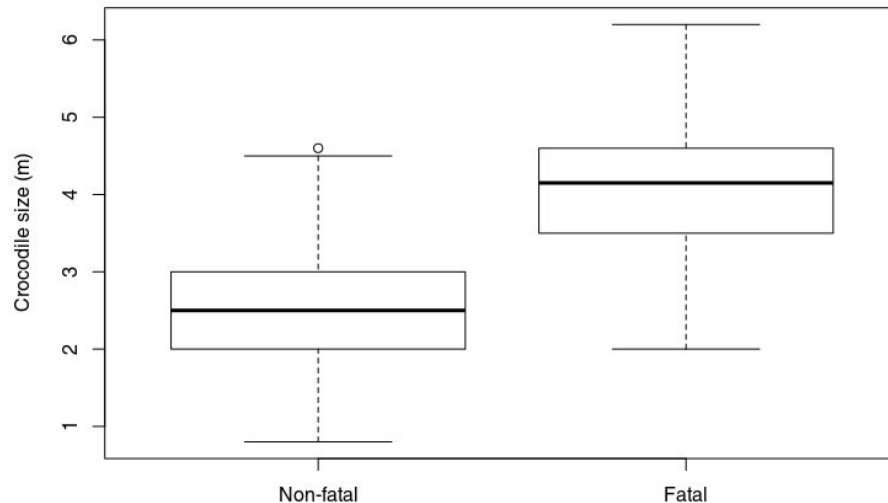




Confounding

Come si può fare?

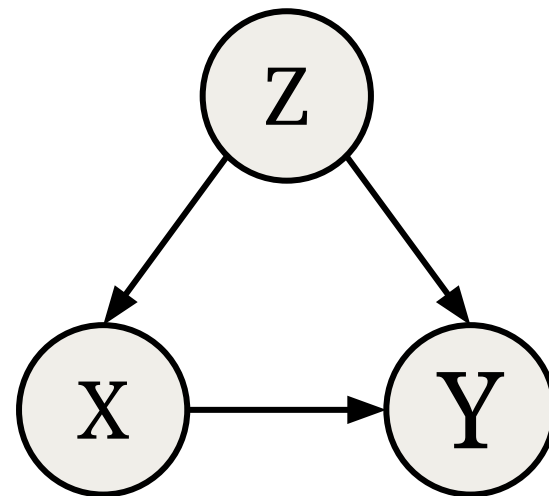
- Disegno dello studio
- Controllo tramite covariate
- Approcci di nicchia
 - Front-door
 - Variabili strumentali





Criterio **back-door**

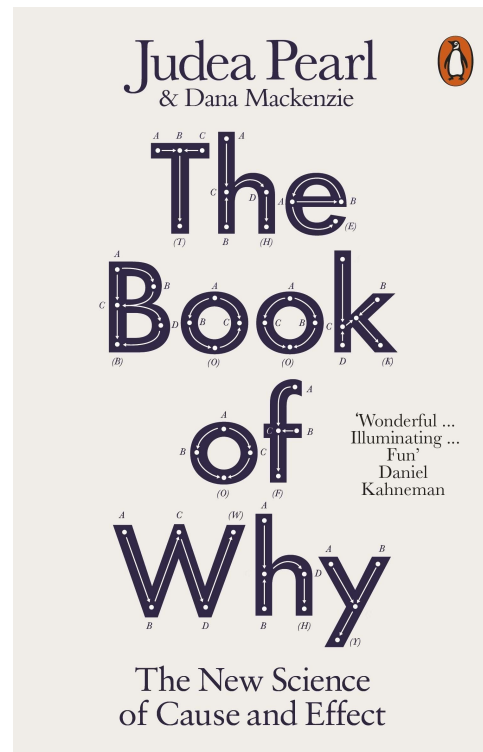
Una relazione $X \rightarrow Y$ riflette un effetto causale (identificata), quando tutti i possibili confounder, **rilevanti**, sono gestiti, tramite stratificazione o inclusione come covariate in un modello.





Criterio **back-door**

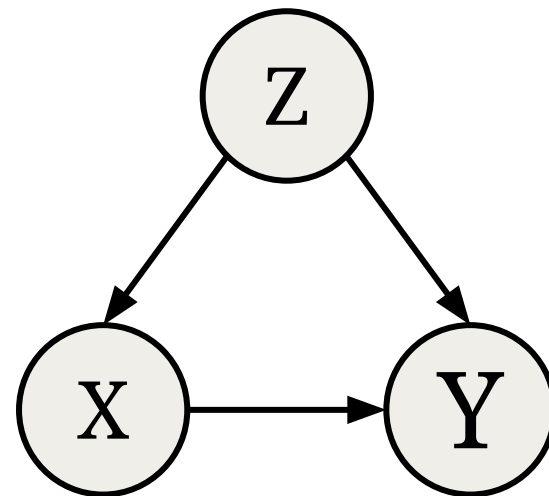
Una relazione $X \rightarrow Y$ riflette un effetto causale (identificata), quando tutti i possibili confounder, **rilevanti**, sono gestiti, tramite stratificazione o inclusione come covariate in un modello.





Criterio **back-door**

Per identificare tutti i confounder che devono essere tenuti sotto controllo, dobbiamo rappresentare la nostra relazione causale come un **grafo diretto aciclico** (DAG)





DAGitty: proviamo

News, sport and

Valigia Blu - Bas

404 - ???404Title???

WebMail PEC

Predator-Prey Im

My Drive - Google

Viola - SlidesCarr

r confounding c

Judea pearl the





DAGitty - drawing an

www.dagitty.net

pearson correlation density wildlife

DAGitty — draw and analyze causal diagrams

DAGitty is a browser-based environment for creating, editing, and analyzing causal diagrams (also known as directed acyclic graphs or causal Bayesian networks). The focus is on the use of causal diagrams for minimizing bias in empirical studies in epidemiology and other disciplines. For background information, see the ["learn"](#) page.

Launch	Download	Learn	Code
 Launch DAGitty online in your browser.	 Download DAGitty's source for offline use.	 Learn more about DAGs and DAGitty.	 The R package "dagitty" is available on CRAN or github.

DAGitty is developed and maintained by [Johannes Textor](#) (Tumor Immunology Lab and Institute for Computing and Information Sciences, Radboud University Nijmegen).

[Tweet #DAGitty](#) [Tweet to @JohannesTextor](#)

The algorithms implemented in DAGitty were developed in close collaboration with [Maciej Liśkiewicz](#) and [Benito van der Zander](#), University of Lübeck, Germany (see literature references below).

DAGitty development happens on [GitHub](#). You can download all source code from there and also get involved.

How can I get help?

If you encounter any problems using DAGitty, or would like to have a certain feature implemented, find me on Twitter or write to ["johannes \[dot\] textor \[at\] gmx \[dot\] de"](#). Your feedback and bug reports are very welcome and contribute to making DAGitty a better experience for everyone. Past contributors are acknowledged in the [manual](#).

Is it free?

Versions

The following versions of DAGitty are available:

- [Development version](#)
Recent development snapshot. May contain new features, but could also contain new bugs.
- [Experimental version](#)
Most recent development snapshot. May not even work.
- 3.0: Released 2019-01-09
- 2.3: Released 2015-08-19
- 2.2: Released 2014-10-30
- 2.1: Released 2014-02-06
- 2.0: Released 2013-02-12
- 1.1: Released 2011-11-29
- 1.0: Released 2011-03-24
- 0.9b: Released 2010-11-24
- 0.9a: Released 2010-09-01

Changelog

2020-01-09

Version 3.0 has been released! Complete reimplementaion of the interface, should work with mobile/touch now.

2018-04-04

Updated the development version and preparing for a long overdue release!

2015-08-19

Version 2.3 has been released! The most notable new feature: instrumental variables.

2014-10-30



Come fare ?

- Capire se la relazione causale ha senso, con la regola di Bradford Hill.
- Identificare tutti i confounder rilevanti
- Scrivere il DAG
- Tenerli sotto controllo come covariate in un modello



Ci sono molti modi ...

Regressioni

Il DAG detta l'inclusione delle covariate e l'effetto marginale della X è $X \rightarrow Y$.

L'esplorazione dei dati e l'analisi di sensitività aiutano.

SEM

Uniscono l'analisi di regressione con una struttura grafica simile ai DAG.

Dicono anche se la struttura ipotizzata è compatibile coi dati.

Reti Bayesiane

Uniscono l'analisi con una struttura grafica, ma si basano sull'inferenza Bayesiana.

Funzionano in termini di indipendenze condizionali.

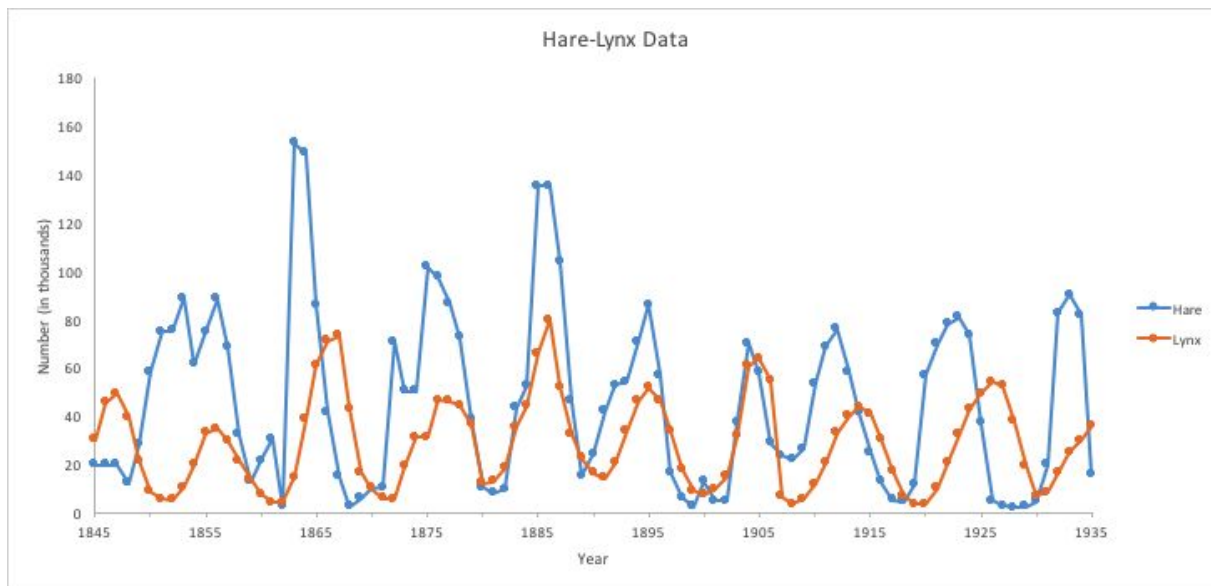
4

Il caso strano

Causalità e dati longitudinali



Lepri e linci

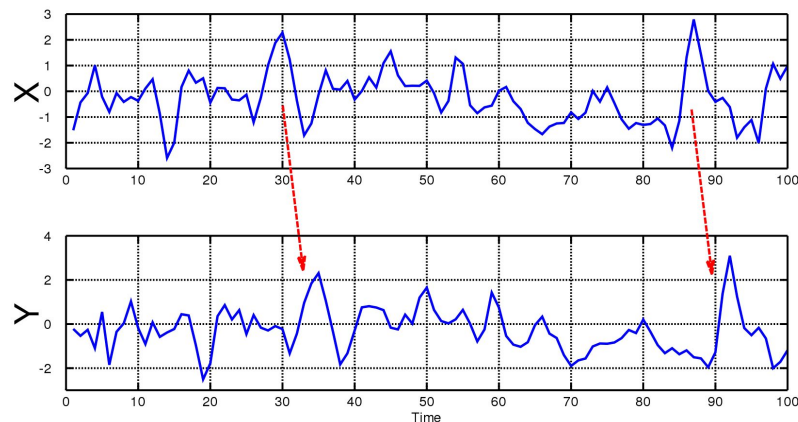




Causalità di Granger

In economia un concetto diffuso è la causalità di Granger.

L'idea è che una serie storica al tempo t possa essere predittiva di un'altra al tempo $t+k$

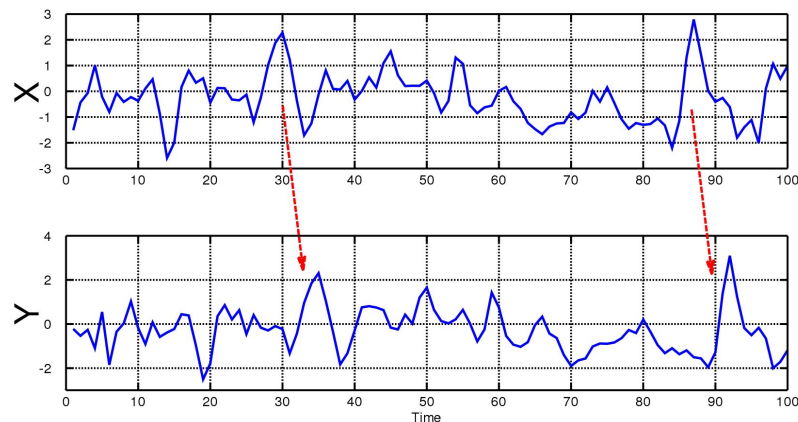




Causalità di Granger

Alla base devono comunque essere rispettati gli assunti di Bradford Hill

Ci sono test statistici apposta, per testare la causalità di Granger.

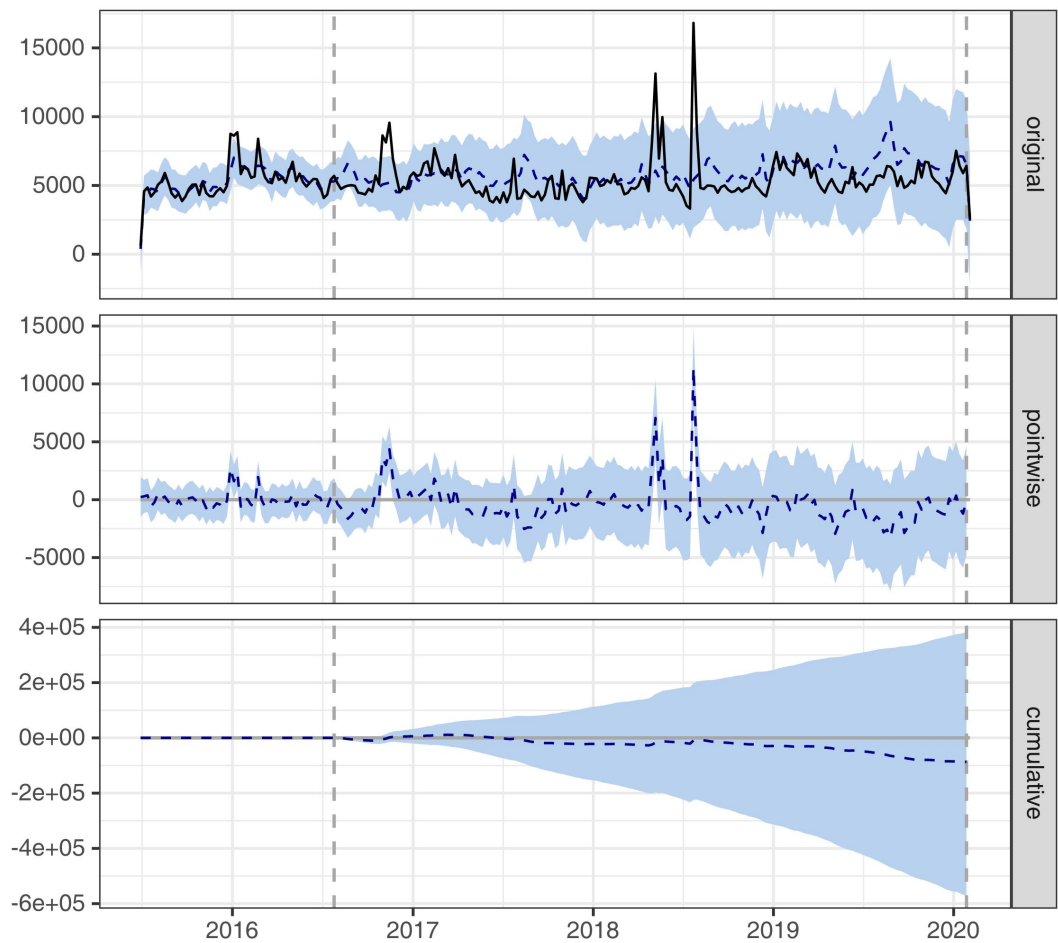




I controlli sintetici

Nei controlli sintetici, l'effetto di un trattamento, ricevuto da una serie storica al tempo t , viene comparato con una (o più) serie storiche di controllo che non hanno ricevuto il trattamento. Purché:

- predittive nel pre-trattamento
- nessuno *spillover*





Grazie!

Alla prossima !

Scrivete a jacopocerri@gmail.com, per dubbi o curiosità.