

Pillola 5 («scatola degli attrezzi»)

Correlazione

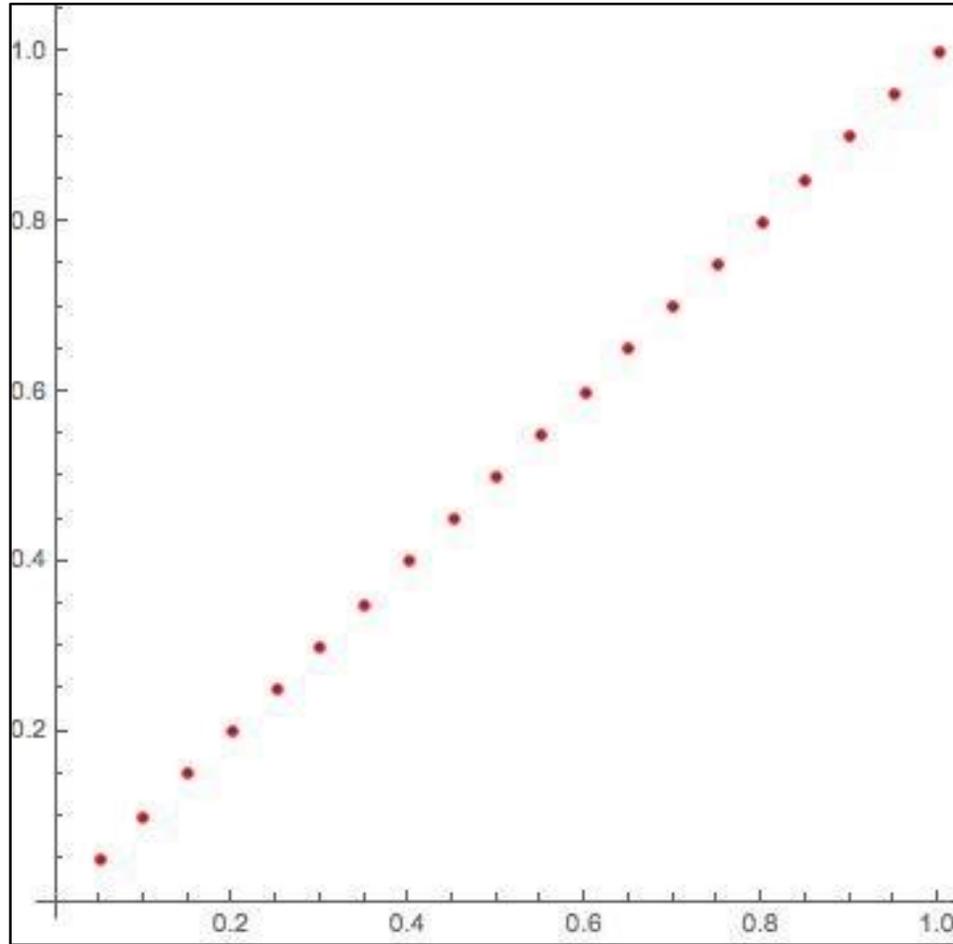


Si ringrazia il collega Prof. Lorenzo Magnea (dispense A.A. 2019-20) per gli spunti

Variabili correlate

- Sia nelle scienze `dure' che in quelle sociali, economiche e mediche spesso misuriamo **coppie di variabili** per verificare se siano legate
- Fatte le misure di solito le confrontiamo con i nostri **modelli teorici**
- Nei casi più semplici osserviamo (o ipotizziamo) che una variabile sia **proporzionale** all'altra (relazione lineare)
- Esiste una semplice **procedura matematica** che ci dice quanto i nostri dati siano **vicini** a una relazione di proporzionalità
- Questa `vicinanza' viene espressa con un numero: il **coefficiente di correlazione r**
 - **$r = 1$** indica **proporzionalità** (perfetta relazione lineare)
 - **$r = 0$** indica **assenza** di correlazione
- Alcuni **esempi** di variabili (potenzialmente) correlate:
 - La distanza percorsa da un'auto, e il tempo trascorso
 - Il numero di armi da fuoco e il numero di decessi che causano
 - Il numero di sigarette fumate e la probabilità di un tumore
 - La probabilità di guarigione e l'assunzione di un medicinale

Esempi di correlazioni

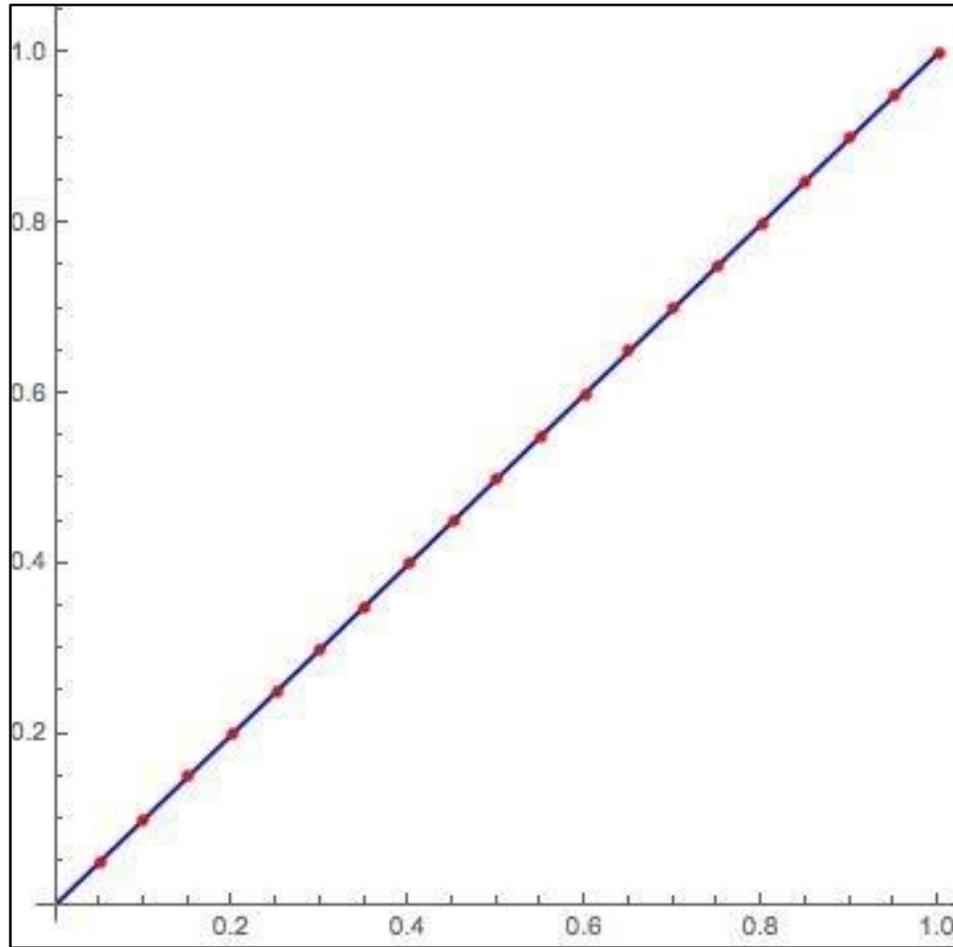


Nei limiti delle misure effettuate, le due variabili sono direttamente proporzionali.

Al crescere di una delle due, anche l'altra cresce.

Correlazione lineare esatta (positiva)

Esempi di correlazioni



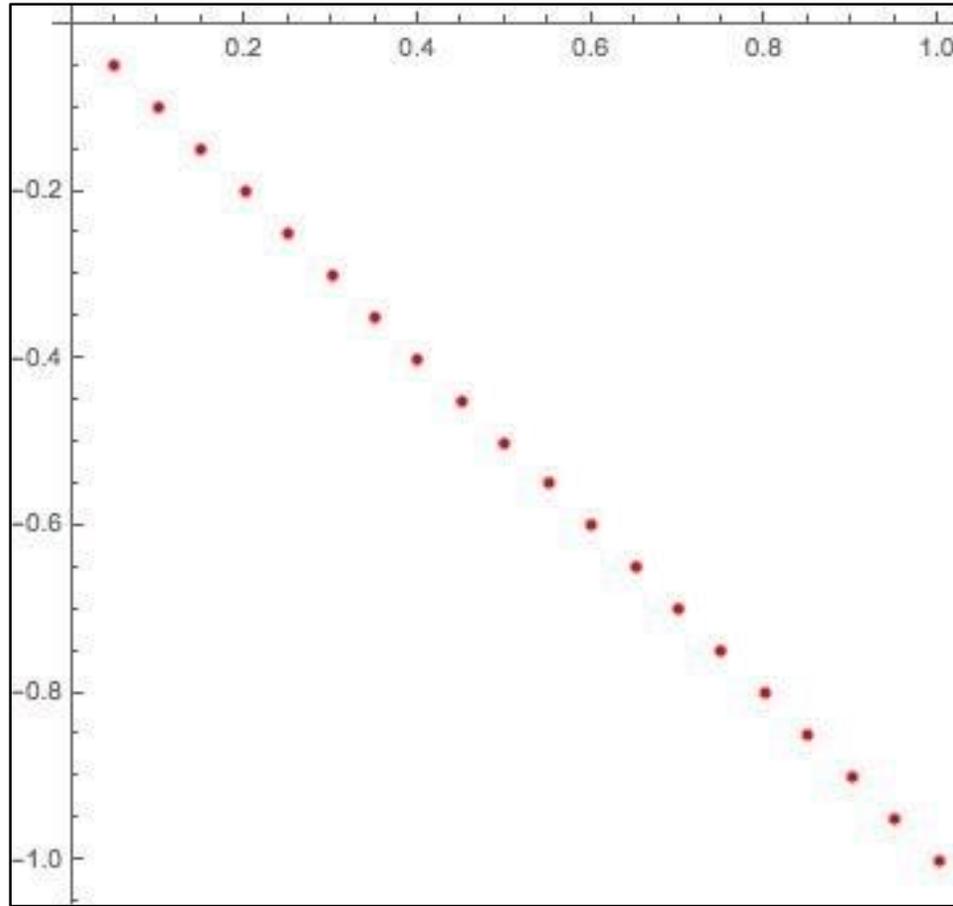
Nei limiti delle misure effettuate, le due variabili sono direttamente proporzionali.

Al crescere di una delle due, anche l'altra cresce.

I punti sono interpolati molto precisamente da una linea retta.

Correlazione lineare esatta (positiva)

Esempi di correlazioni

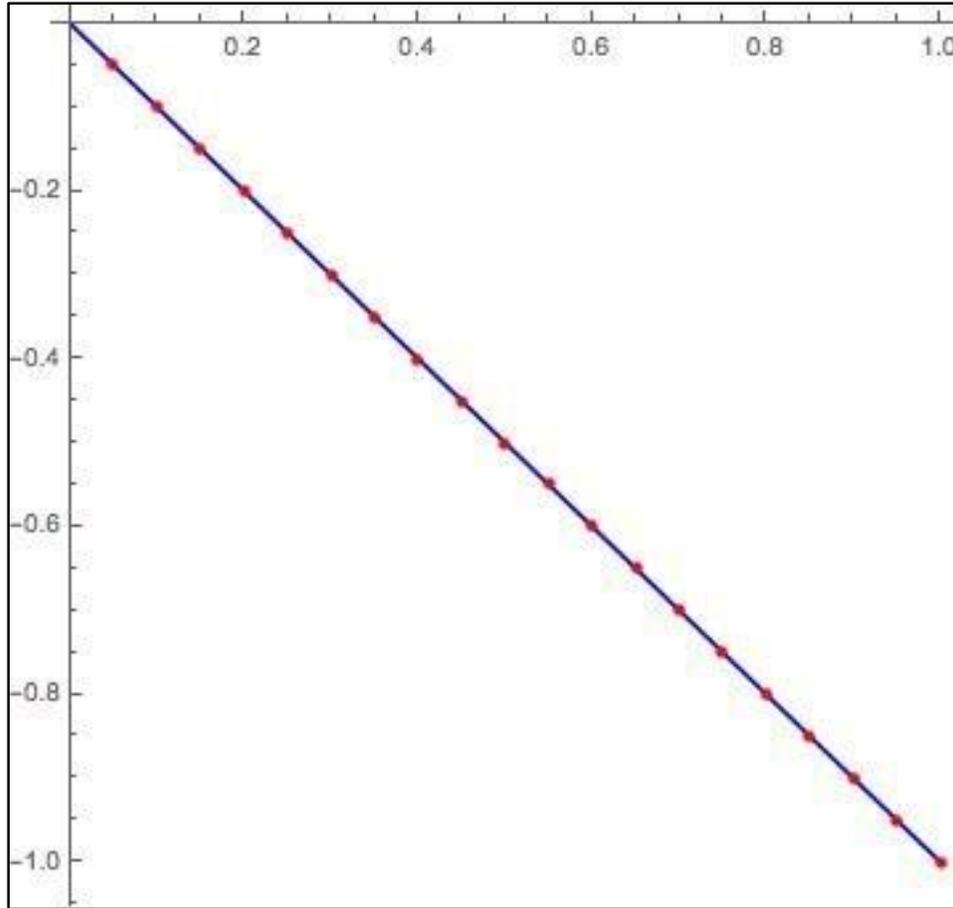


Nei limiti delle misure effettuate, le due variabili sono direttamente proporzionali.

Al crescere di una delle due, l'altra decresce.

Correlazione lineare esatta (negativa)

Esempi di correlazioni



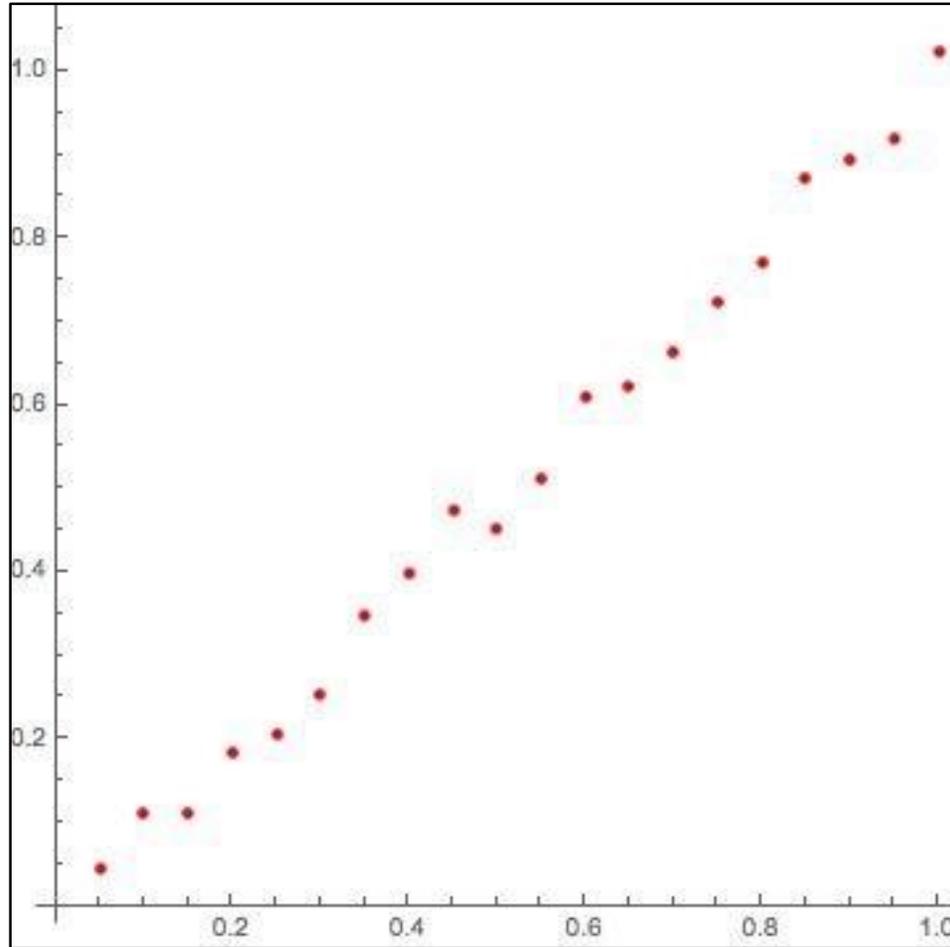
Nei limiti delle misure effettuate, le due variabili sono direttamente proporzionali.

Al crescere di una delle due, l'altra decresce.

I punti sono interpolati molto precisamente da una linea retta.

Correlazione lineare esatta (negativa)

Esempi di correlazioni

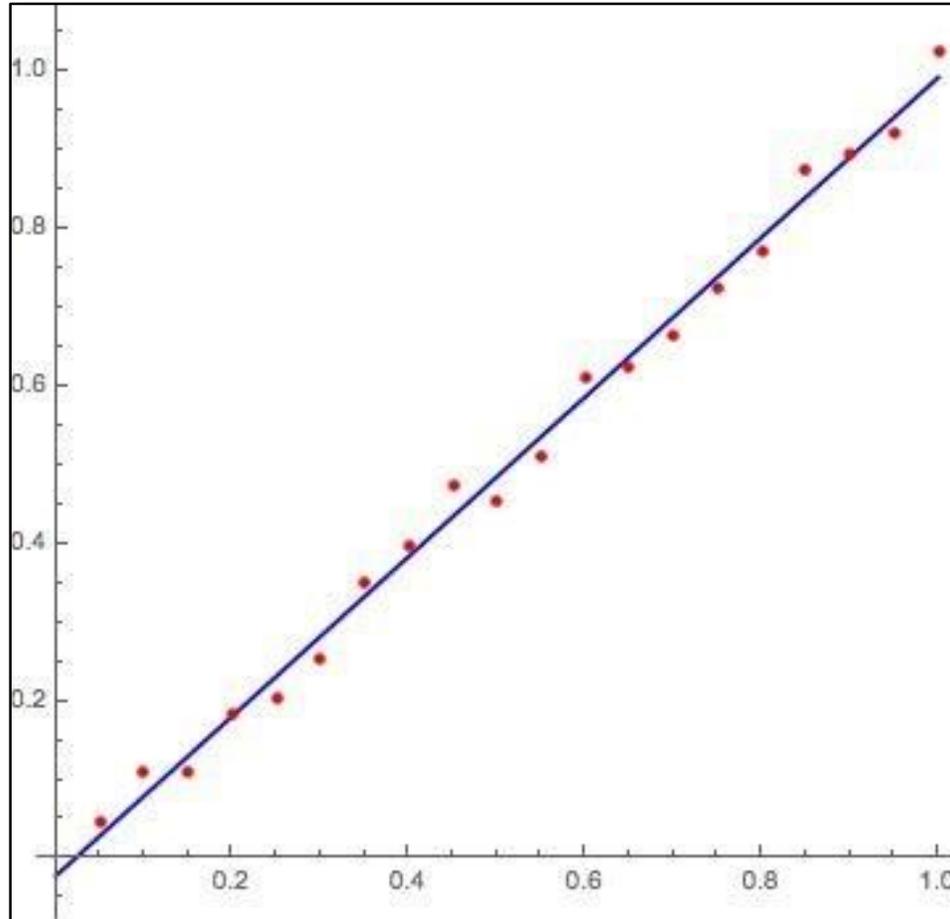


Le misure effettuate suggeriscono una relazione di proporzionalità tra le due variabili .

Al crescere di una delle due, anche l'altra cresce.

Forte correlazione lineare (positiva)

Esempi di correlazioni



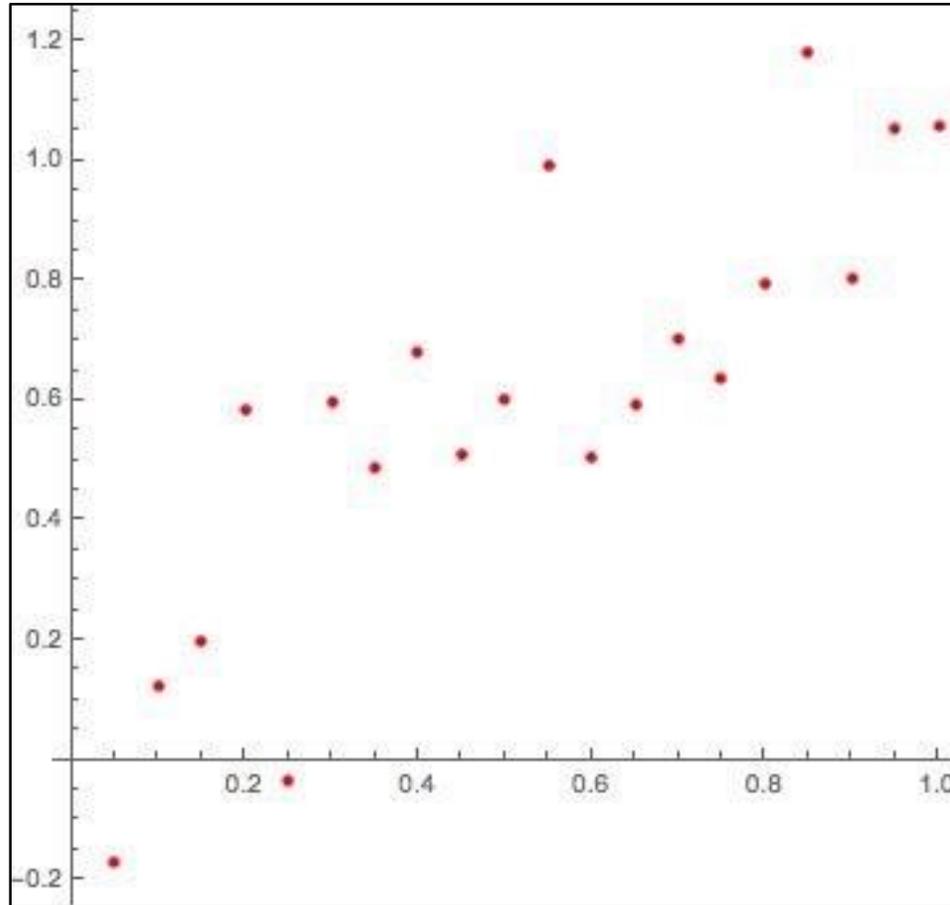
Le misure effettuate suggeriscono una relazione di proporzionalità tra le due variabili .

Al crescere di una delle due, anche l'altra cresce.

I punti possono essere interpolati approssimativamente da una linea retta.

Forte correlazione lineare (positiva)

Esempi di correlazioni

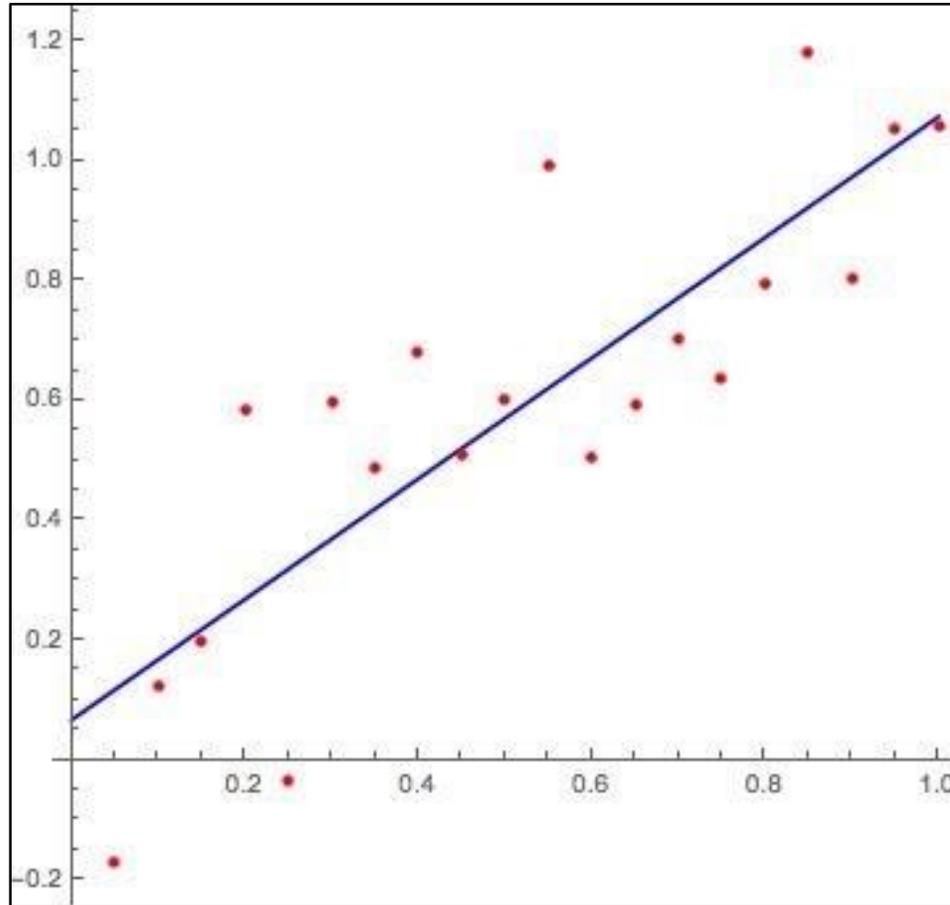


Il risultato è vago
ma potrebbe essere consistente
con una relazione lineare.

Il risultato suggerisce che
al crescere di una delle due variabili
anche l'altra cresca.

Debole correlazione lineare (positiva)

Esempi di correlazioni



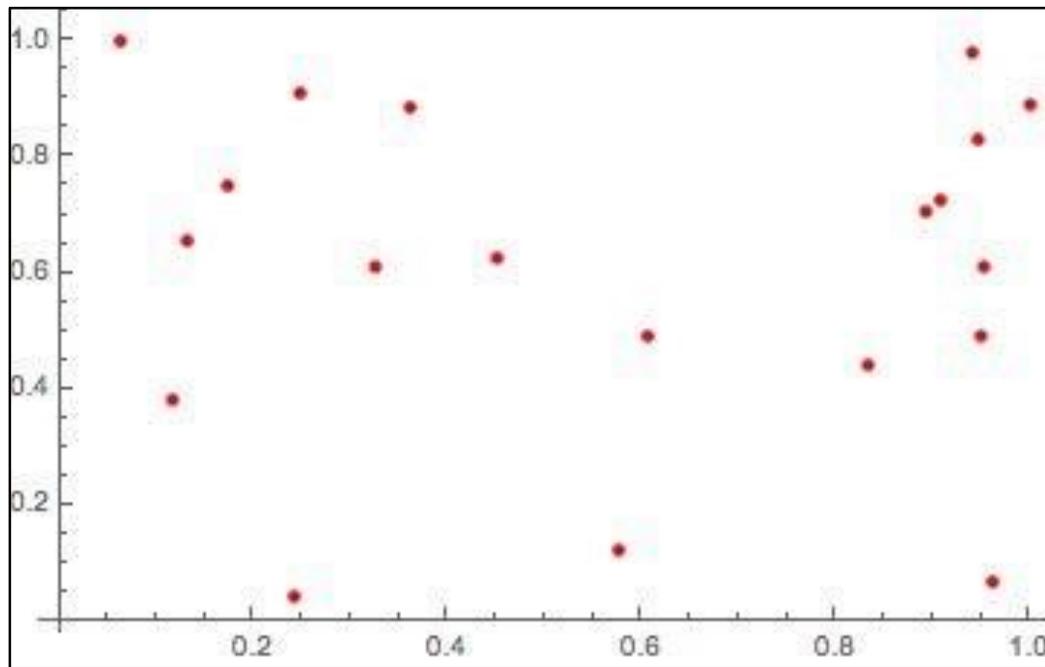
Debole correlazione lineare (positiva)

Il risultato è vago
ma potrebbe essere consistente
con una relazione lineare.

Il risultato suggerisce che
al crescere di una delle due variabili
anche l'altra cresca.

Una procedura standard permette
di costruire la retta che fornisce la
migliore approssimazione.

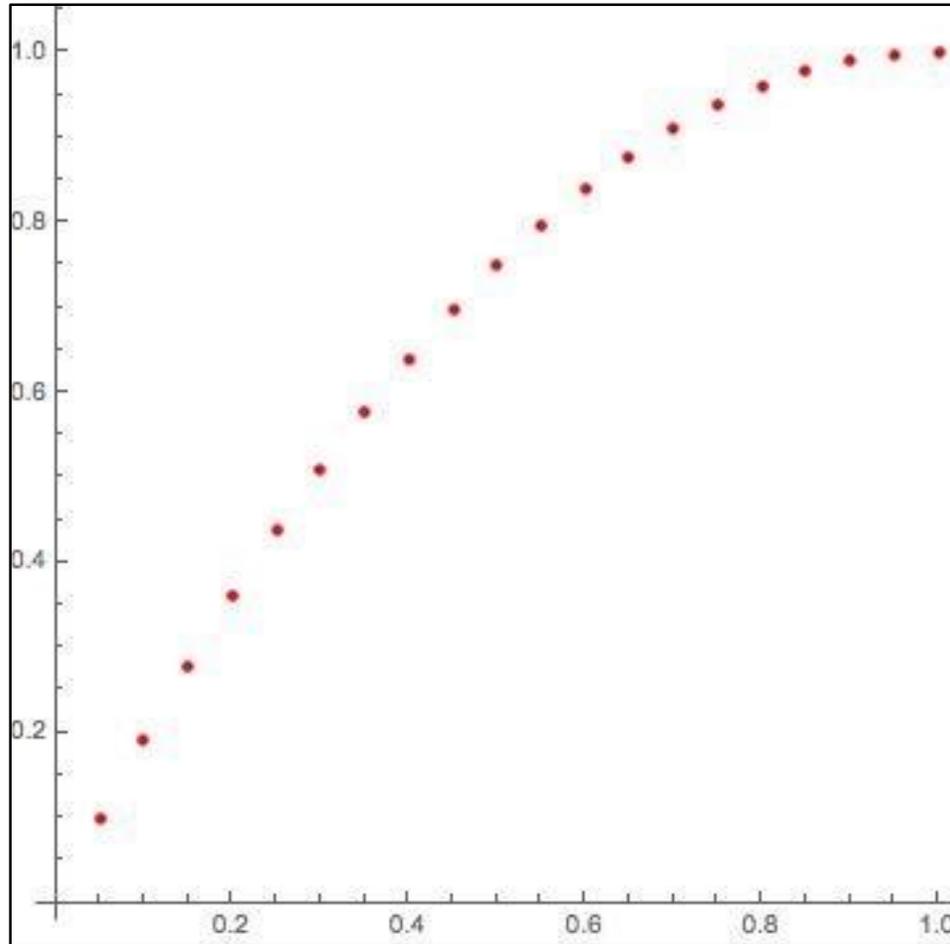
Esempi di correlazioni



Il risultato suggerisce che non esista nessuna relazione tra le due variabili

Nessuna correlazione

Esempi di correlazioni

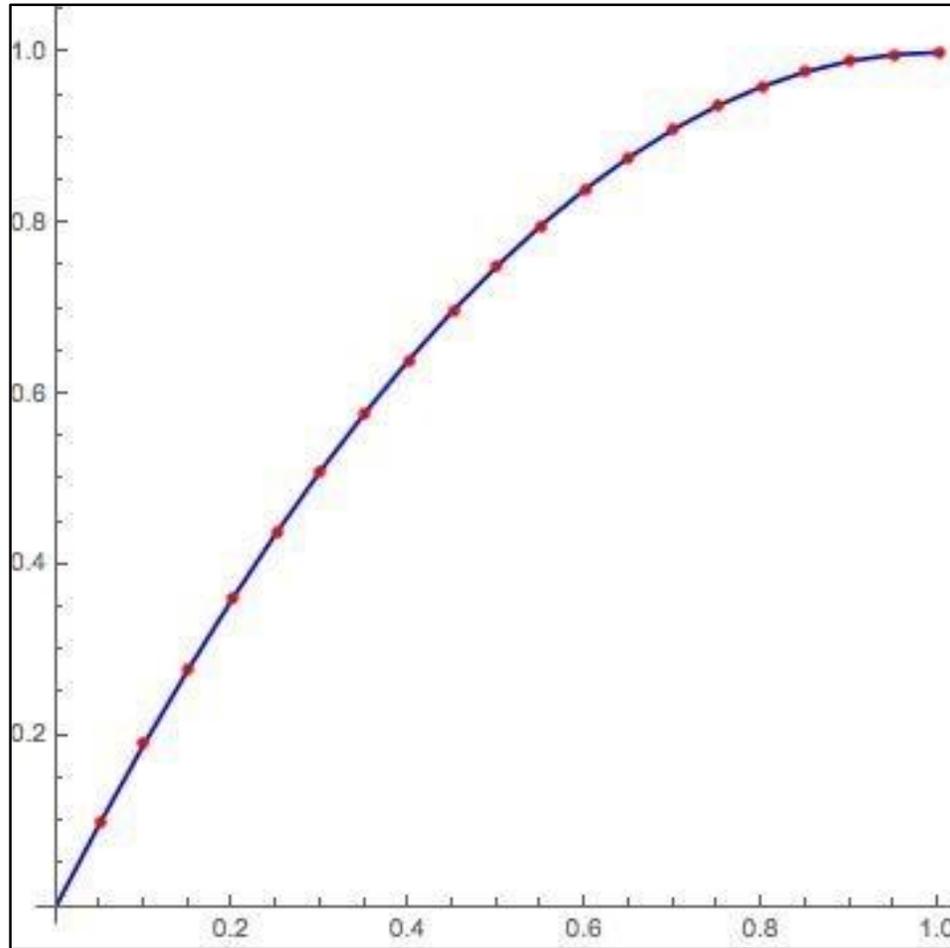


Nei limiti delle misure effettuate,
le due variabili sono legate da
una relazione quadratica.

Non è facile escludere relazioni
funzionali più complicate ...

Correlazione non lineare (parabolica) esatta

Esempi di correlazioni



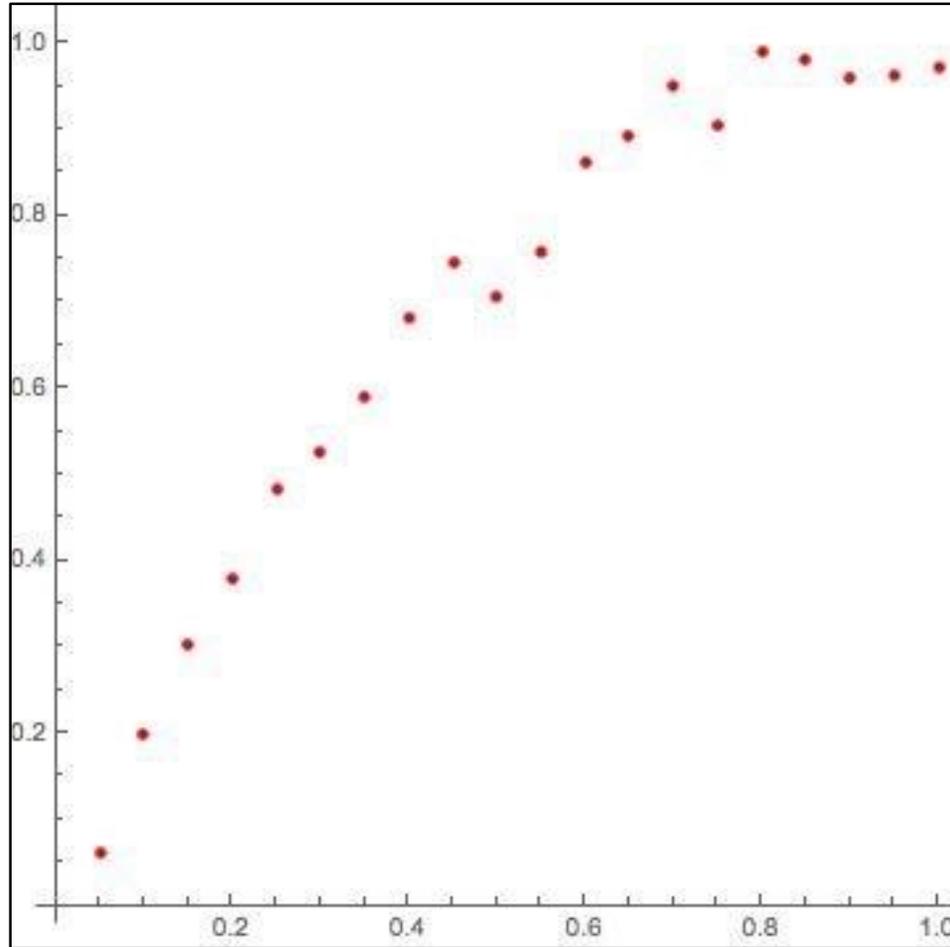
Nei limiti delle misure effettuate,
le due variabili sono legate da
una relazione quadratica.

Non è facile escludere relazioni
funzionali più complicate ...

I punti sono interpolati molto
precisamente da una
parabola.

Correlazione non lineare (parabolica) esatta

Esempi di correlazioni

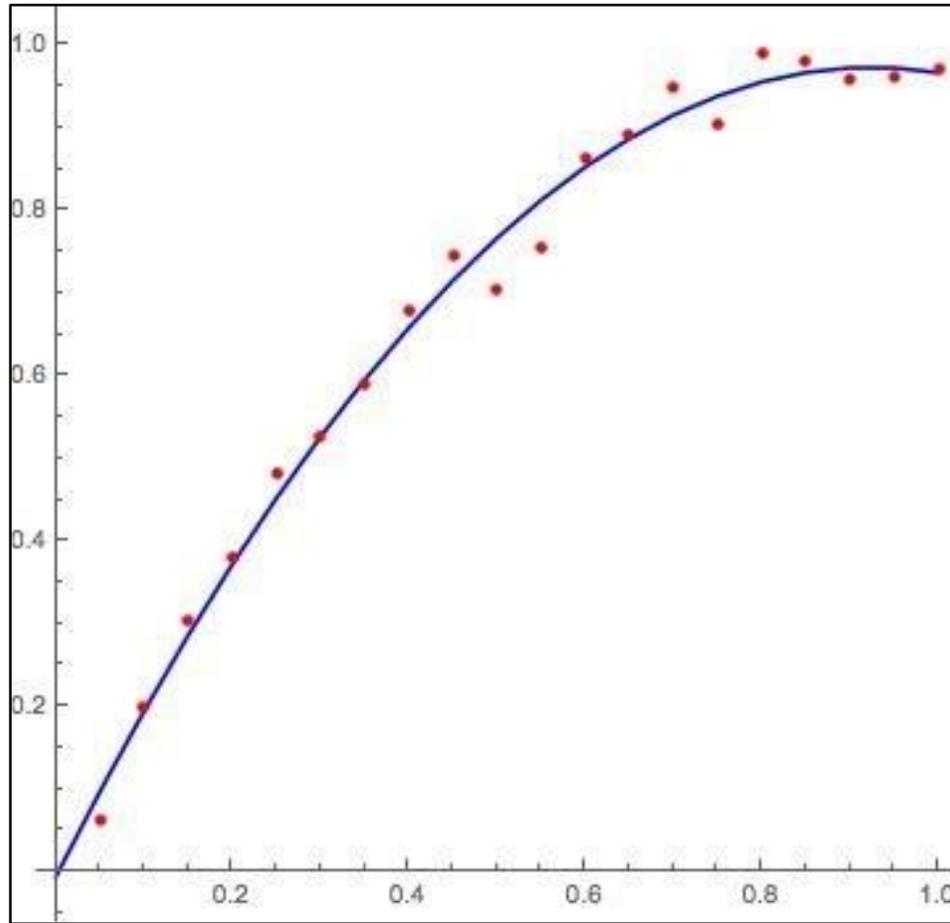


I risultati suggeriscono l'esistenza di una relazione funzionale semplice, probabilmente quadratica.

Non è facile escludere relazioni funzionali più complicate.

Forte correlazione non lineare

Esempi di correlazioni

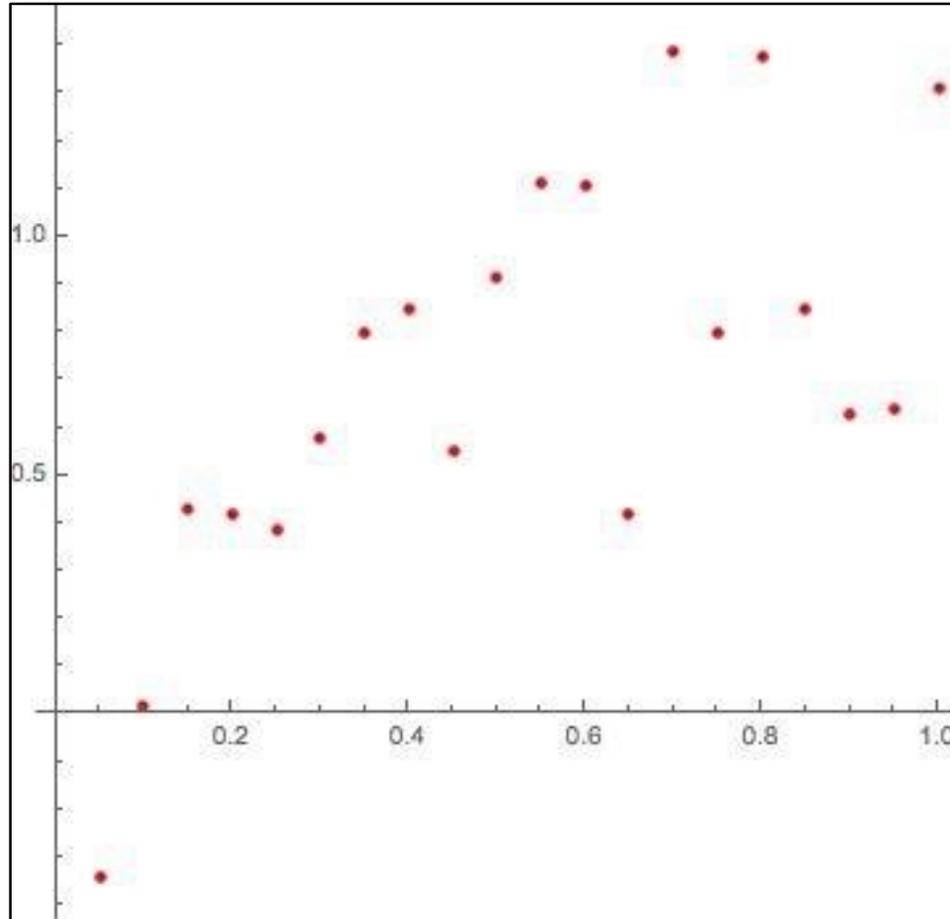


I risultati suggeriscono l'esistenza di una relazione funzionale semplice, probabilmente quadratica.

Non è facile escludere relazioni funzionali più complicate.

Forte correlazione non lineare

Esempi di correlazioni

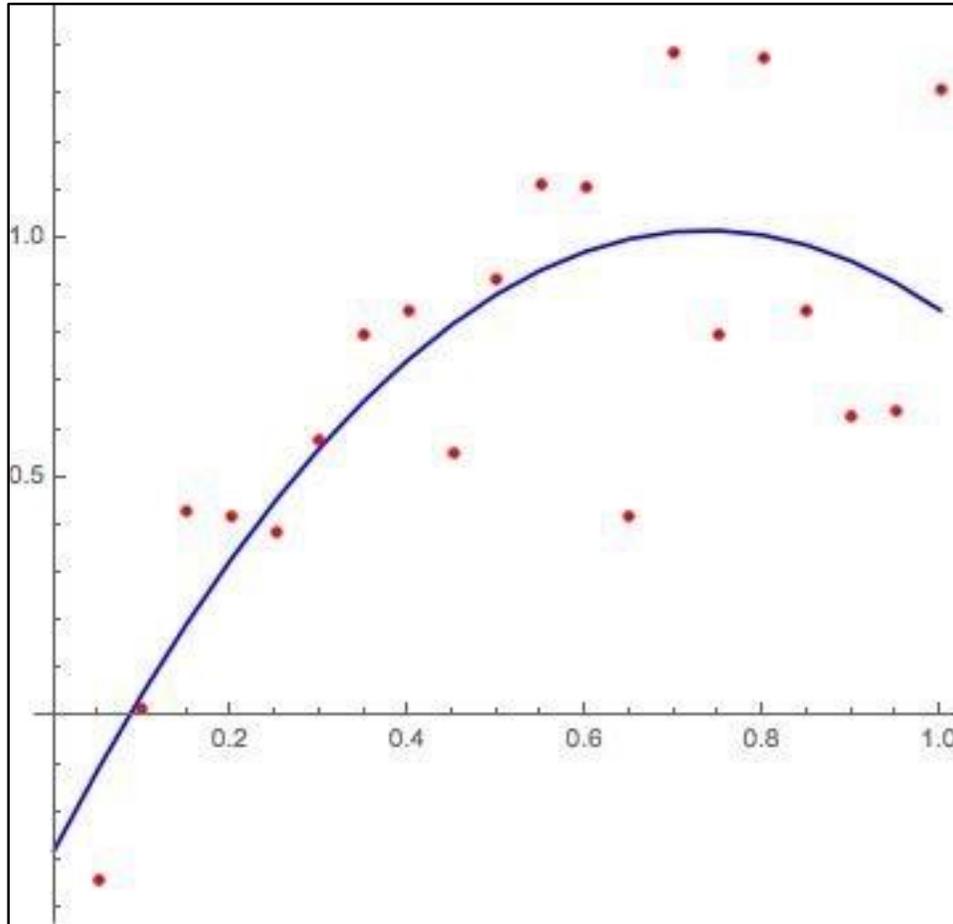


Il risultato è vago e l'eventuale relazione funzionale tra le variabili non è facile da individuare.

Sono necessarie ipotesi ulteriori, o modelli per il comportamento delle variabili.

Debole correlazione non lineare

Esempi di correlazioni



Il risultato è vago e l'eventuale relazione funzionale tra le variabili non è facile da individuare.

Sono necessarie ipotesi ulteriori, o modelli per il comportamento delle variabili.

Fatta un'ipotesi (per esempio un legame quadratico), una procedura standard permette di costruire la parabola che fornisce la migliore approssimazione.

Debole correlazione non lineare

Correlazione e causalità

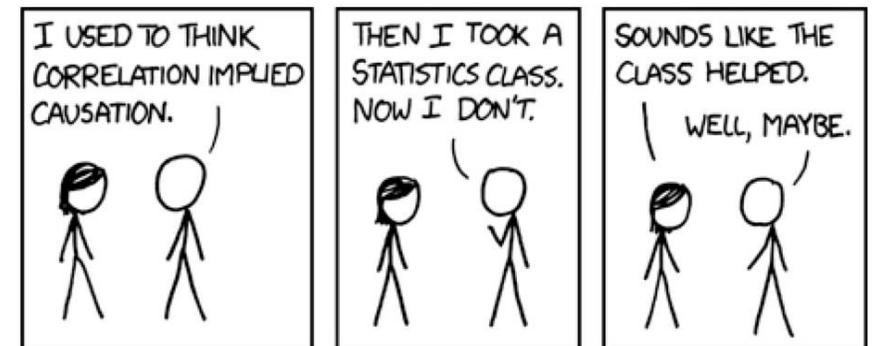
L'EVENTUALE correlazione tra due variabili

NON

implica **NECESSARIAMENTE** un rapporto di causalità

La correlazione non implica la causalità!

- Spesso l'informazione contenuta in un grafico è molto limitata (per esempio, molte variabili aumentano **lentamente in funzione del tempo**)
- Se A e B sono correlate, potrebbe essere che:
 - A sia causa di B
 - oppure che B sia causa di A
 - oppure A e B potrebbero avere una causa comune, C, non considerata
- Se si esaminano un numero molto grande di coppie di variabili, è molto probabile che si trovino coppie correlate per pura coincidenza
- Una forte correlazione suggerisce di iniziare a studiare con altri mezzi se esista un nesso causale, e in quale direzione
- Per esempio potremmo chiederci:
 - Il tumore al polmone potrebbe causare l'abitudine al fumo?
 - Il numero di vittime delle armi da fuoco potrebbe causarne la diffusione?
 - Se i sintomi A e B compaiono spesso insieme, curare A curerà anche B?
 - Il testosterone abbrevia la durata della vita?



Variabili correlate? Siamo proprio sicuri?

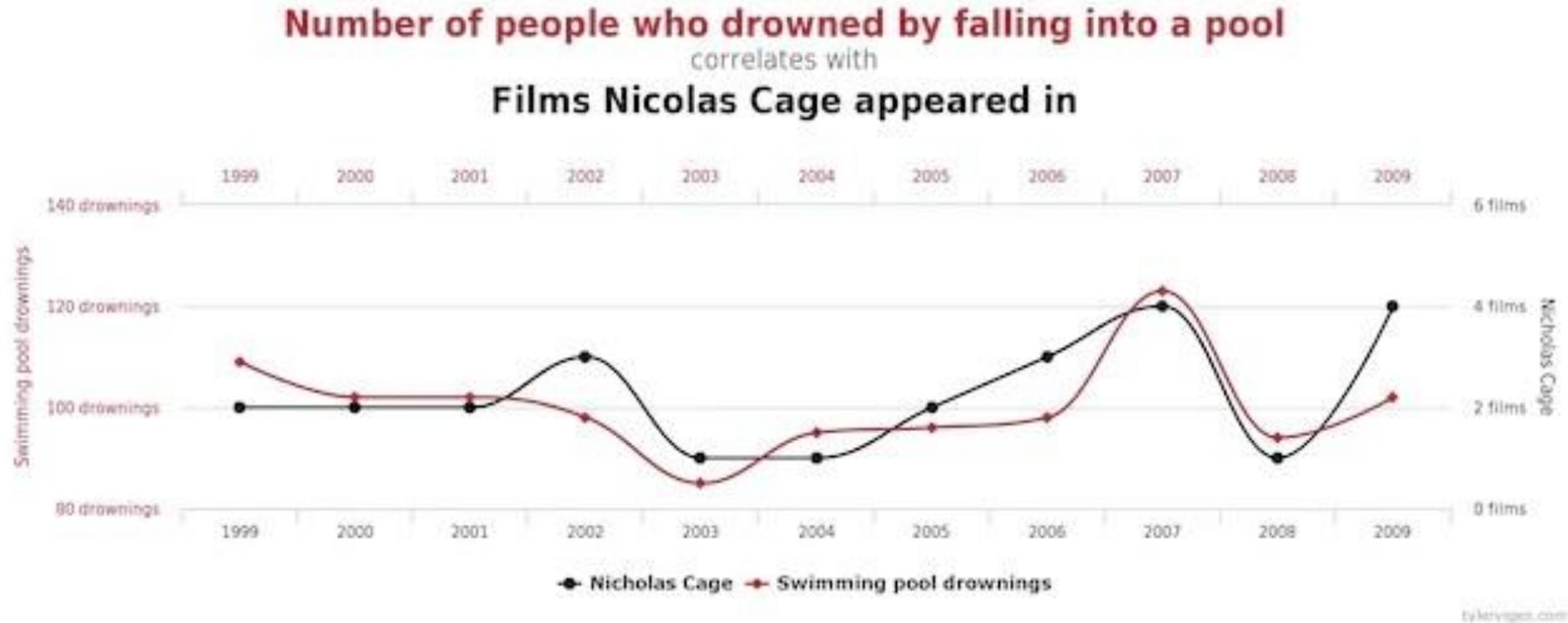


Chart courtesy of Tyler Vigen Chart courtesy of Tyler Vigen

C'è una correlazione tra il numero di persone annegate dopo essere cadute in una piscina e ...
... il numero di film in cui ha recitato Nicolas Cage?

Altri esempi di correlazioni improbabili: <http://www.tylervigen.com/spurious-correlations>