EDUCAZIONE CIVICA digitale

Dalla correlazione alla causalità

Applicazione al caso del global warming

1. Correlazione e causalità

Tra due variabili empiriche esiste **correlazione** quando si constata la tendenza di una di esse a variare (con un'approssimazione più o meno grande) in funzione dell'altra. Si parla di correlazione **positiva** se al crescere di una variabile anche l'altra variabile tende a crescere. Si parla di correlazione **negativa** se al crescere di una variabile l'altra variabile tende a decrescere.

La correlazione è una proprietà statistica di due o più variabili; per capire se è presente una correlazione tra due variabili non serve conoscere nulla di quelle variabili se non il loro andamento.

La correlazione è definibile matematicamente, ma in questa scheda ci limiteremo alla definizione data sopra e a vedere come rilevare la presenza di una correlazione osservando un grafico che riporta l'andamento delle due variabili.

Osserviamo per esempio il grafico a destra: esso riporta l'andamento di due variabili. La prima variabile (curva verde) è la media voti di un alunno calcolata di mese in mese, negli ultimi sei mesi di scuola, da gennaio a giugno; per esempio leggiamo dal grafico che, nel mese di febbraio, la media voti è molto vicina a 6. La seconda variabile (curva rossa) è il numero medio di ore di studio giornaliere: questo può essere calcolato sommando tutte le ore di studio svolto in un mese e dividendo per il numero di giorni del mese. Per esempio, osserviamo che nel mese di febbraio l'alunno ha studiato in media un'ora al giorno.

Osserviamo ora il grafico nel suo insieme: la curva rossa e la curva verde hanno due andamenti simili. Rileviamo che tra i mesi di gennaio (1) e febbraio (2) entrambe le curve scendono, mentre nel rimanente periodo, da febbraio a giugno, esse salgono, anche se i due andamenti non sono perfettamente identici. Rifacendoci alla definizione di correlazione data all'inizio di questo paragrafo, possiamo affermare che tra la media dei voti e il numero di ore di studio esiste una **correlazione positiva**.

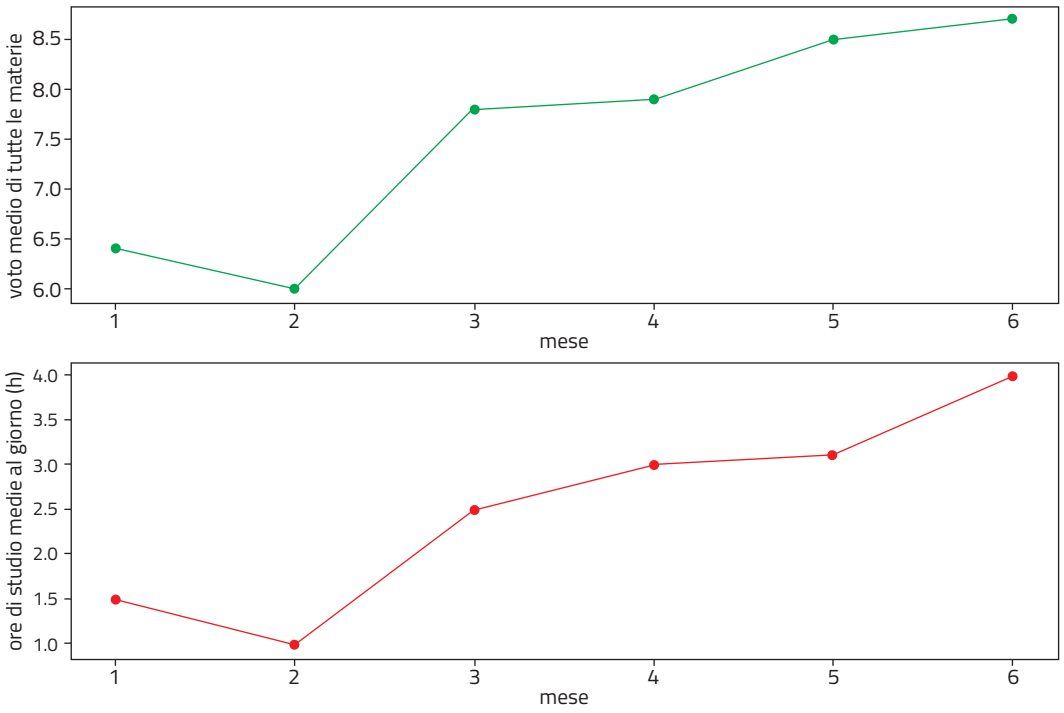


Figura 1 Media dei voti e ore di studio giornaliere medie nel periodo Gennaio-Giugno.

Ma esiste un legame reale tra le due variabili?

Ci risulta evidente che la risposta a questa domanda sia affermativa: voti alti in tutte le materie arrivano soltanto con una buona dose di tempo passato a studiare. Non sempre però il legame tra due variabili è supportato da evidenze così semplici.

Per questo motivo introduciamo il concetto di **causalità**, inteso come il rapporto che lega la causa con l'effetto. Si parla di causalità tra due variabili, se la variazione di una variabile causa una variazione dell'altra variabile (positiva o negativa). La causalità non è una semplice proprietà statistica di due variabili. Per capire se è presente causalità serve conoscere in dettaglio i fenomeni e i comportamenti che sono alla base delle variazioni delle variabili (nel caso delle ore di studio, ogni studente sa che

per ottenere un buon voto occorre un adeguato numero di ore di studio).
Nella maggior parte delle situazioni, invece, per individuare causalità occorrono competenze specifiche sui fenomeni che stiamo analizzando; se non possediamo tali competenze dobbiamo affidarci alla conoscenza degli esperti.
Per rilevare causalità possiamo affidarci alla correlazione? La risposta è no!
Può accadere che due fenomeni siano correlati senza che esista un legame reale tra di essi: in questi casi di parla di correlazione spuria. Il sito web http://www.tylervigen.com/spurious-correlations riporta svariati casi di correlazioni spurie, alcuni dei quali sono buffi o inverosimili. Per esempio osserviamo il grafico sotto, tratto dal sito. Esso riporta il numero di lanci nello spazio di missioni non commerciali e il numero di dottorati in sociologia assegnati negli US.
È evidente che i due fenomeni considerati non hanno nulla a che vedere l'uno con l'altro, ma il grafico mostra una correlazione positiva quasi perfetta!
Possiamo concludere che la correlazione è uno strumento potente, che possiamo utilizzare per confrontare andamenti di variabili diverse, ma deve essere maneggiato con cura. Per trarre conclusioni ulteriori in merito alla causalità, nella maggior parte dei casi, dobbiamo affidarci alla competenza degli esperti. Nei casi più studiati possiamo fare ricorso al web ed effettuare una ricerca accurata di fonti affidabili e autorevoli, alle quali affidarci per un'analisi che voglia andare oltre la semplice correlazione.

2. Grafici

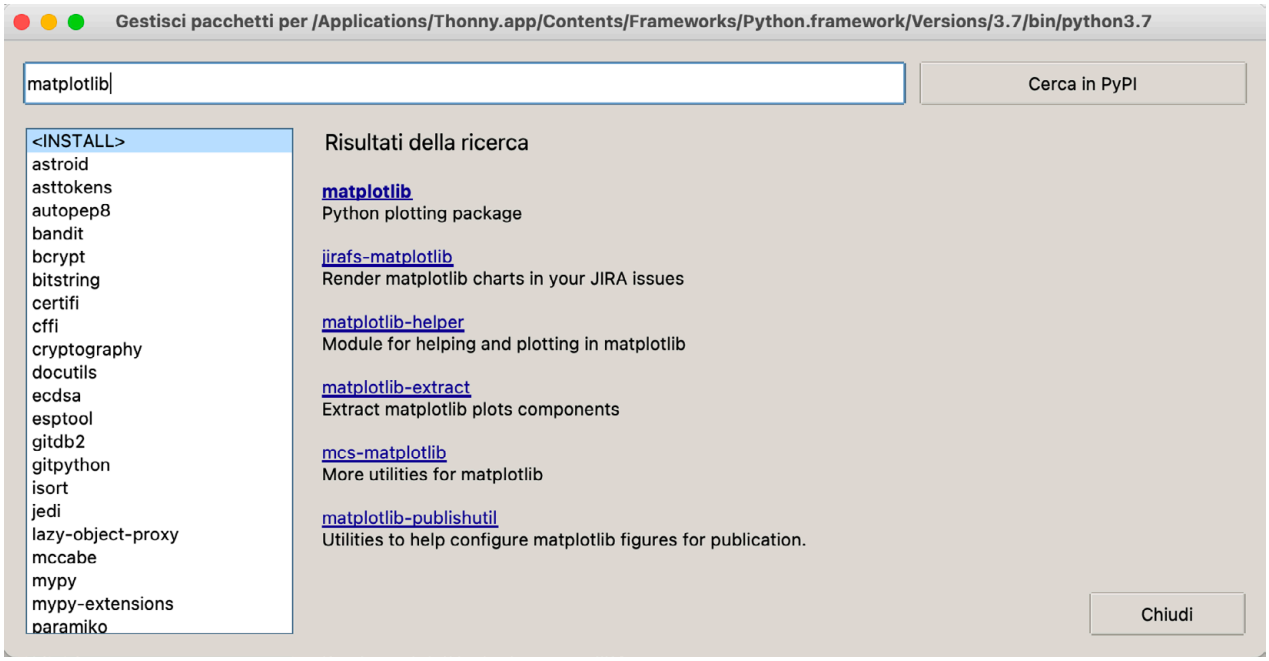
In questo paragrafo mostreremo come creare dei grafici simili ai precedenti, in modo da studiare la presenza di correlazione tra due variabili. Faremo uso del calcolatore e del linguaggio Python, ma in alternativa possiamo utilizzare un foglio di calcolo, oppure carta millimetrata, righello e matite colorate.
Il punto di partenza per tracciare un grafico sono i dati che vogliamo rappresentare: questi possono essere contenuti in una tabella scritta su carta, oppure all'interno di un file. Uno dei formati di file più utilizzati è il **csv** (*comma separate values*) che rappresenta la tabella, riga dopo riga, in formato testuale. I campi della tabella in ogni riga sono separati da virgole (*comma* in inglese). Per esempio il file contenente le valutazioni e le ore di studio del paragrafo precedente (la tabella nella figura sotto a sinistra) è strutturato come nella figura sotto a destra.
La prima riga riporta i nomi delle diverse colonne, mentre le righe successive contengono i dati. Ogni riga riporta nell'ordine il numero del mese, il nome del mese, il numero di ore giornaliere studiate in media quel mese e infine la media dei voti per quel mese.

mese_n	mese	ore_giorno_h	media_voti
1	Gennaio	1.5	6.4
2	Febbraio	1.0	6.1
3	Marzo	2.5	7.8
4	Aprile	3.0	8.1
5	Maggio	3.3	8.4
6	Giugno	3.8	8.5

```
mese_n,mese,ore_giorno_h,media_voti
1,Gennaio,1.5,6.4
2,Febbraio,1.0,6.1
3,Marzo,2.5,7.8
4,Aprile,3.0,8.1
5,Maggio,3.3,8.4
6,Giugno,3.8,8.5
```

Per tracciare il grafico utilizziamo il linguaggio di programmazione Python nell'ambiente di sviluppo integrato Thonny (https://thonny.org). Se non abbiamo confidenza con gli strumenti informatici o con Python, possiamo tracciare il grafico sulla carta millimetrata facendo riferimento ai dati riportati nella tabella della figura sopra (in questo caso, tralasciamo la parte rimanente di questo paragrafo).

Procediamo con Python: all'interno di Thonny occorre installare l'ulteriore modulo **`matplotlib`** utile alla creazione di grafici. Per installare `matplotlib` all'interno di Thonny clicchiamo nel menu *strumenti* e poi sulla voce *Gestisci i pacchetti* che apre una finestra come quella sottostante in cui ricerchiamo `matplotib` e clicchiamo su *installa*.



Terminata la procedura possiamo scrivere il codice Python che legge il file e crea un grafico come quello visto poc'anzi. Il codice è il seguente.

```
1  import matplotlib.pyplot as plt
2  import csv
3
4  mesi_n = [] #lista per i mesi (numero del mese)
5  ore_studio = [] #lista per le ore di studio
6  voti = [] #lista per la media dei voti
7  data_file = open("./voti_studio.csv")
8  data_reader = csv.reader(data_file, delimiter=',')
9  next(data_reader) #salta la prima riga del file contenente i titoli
10 for row in data_reader:
11     mesi_n.append(int(row[1]))
12     ore_studio.append(float(row[2]))
13     voti.append(float(row[3]))
14 data_file.close()
15
16 fig, (ax1, ax2) = plt.subplots(2, 1) #crea una figura con due grafici
17 fig.suptitle('Media dei voti e ore di studio giornaliere medie nel periodo Gennaio-Giugno')
18
19 #primo grafico (media dei voti)
20 ax1.plot(mesi_n, voti, 'o-g')
21 ax1.set_xlabel('mese')
22 ax1.set_ylabel('voto medio di tutte le materie')
23
24 #secondo grafico (ore di studio)
25 ax2.plot(mesi_n, ore_studio, 'o-r')
26 ax2.set_xlabel('mese')
27 ax2.set_ylabel('ore di studio medie al giorno (h)')
28
29 plt.show()
```

Alla riga #1 importiamo il modulo `matplotlib.pyplot` per tracciare i grafici mentre alla riga #2 importiamo il modulo nativo csv per la lettura del file di dati. Tra le righe #4 e #14 leggiamo il file *voti_studio.csv* e carichiamo i dati letti all'interno di liste. Le righe da #16 a #29 realizzano il grafico. Eseguiamo il programma cliccando sul tasto verde di Thonny e creiamo il nostro grafico dal quale possiamo desumere la correlazione positiva tra media voti e numero di ore di studio.
Questo programma è il punto di partenza per tracciare grafici rappresentativi di altri fenomeni come quelli indicati negli esercizi 1, 2 e 3.

3. Il *global warming*

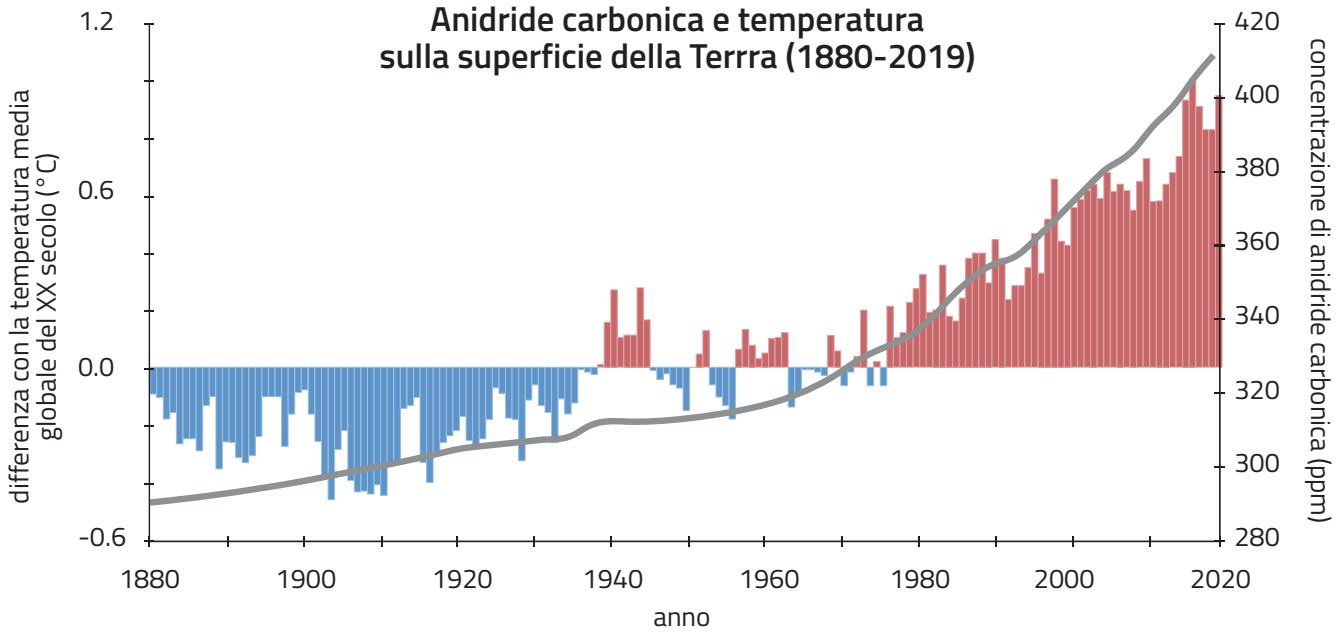
Ora siamo pronti per applicare quanto appreso a un problema che riguarda tutti noi da vicino, tutto il pianeta Terra e i suoi ecosistemi naturali: il riscaldamento globale (*global warming*).

Si tratta di uno dei maggiori problemi che affligge il futuro del pianeta: da oltre un decennio è sotto gli occhi di tutti e al centro delle attenzioni della stampa.

Dalla prima rivoluzione industriale in avanti, l'uso dei combustibili fossili è aumentato di pari passo con il consumo di energia, infatti la loro combustione costituisce ancora oggi una fonte primaria di energia. La combustione di carbone, petrolio e suoi derivati produce anidride carbonica (CO_2) che viene immessa nell'atmosfera della Terra e si aggiunge a quella già presente naturalmente.

La curva grigia del grafico sotto rappresenta la concentrazione dell'anidride carbonica nell'atmosfera in parti per milione (ppm) dal 1880 al 2020. Le barre blu e rosse rappresentano invece l'*anomalia di temperatura* del pianeta in gradi Celsius, anno per anno. Con *anomalia di temperatura* si intende la differenza tra la temperatura media globale dell'anno in esame e la temperatura media globale di tutto il ventesimo secolo. Quindi, per esempio, le barre rosse corrispondono ad anni con anomalie di temperatura elevate, cioè ad anni in cui la temperatura media del pianeta è più alta della temperatura media di tutto il ventesimo secolo. Si può notare che l'anomalia degli ultimi cinque anni è vicina a 1°C.

È evidente la correlazione tra i due fenomeni: a meno di isolate eccezioni, notiamo che l'andamento



crescente della concentrazione dell'anidride carbonica va di pari passo con l'anomalia di temperatura del pianeta Terra. Si può dire che esiste causalità tra emissioni e aumento della temperatura globale? La risposta è affermativa. Il legame tra i due fenomeni va sotto il nome di **effetto serra**. L'effetto fu ipotizzato e quantificato per la prima volta nel 1896 dallo scienziato svedese Svante Arrhenius, anche se il termine serra fu introdotto nel 1901 dal meteorologo Nils Gustaf Ekholm. L'effetto serra è un fenomeno naturale: alcuni gas che compongono l'atmosfera sono in grado di trattenere il calore che la Terra emette dopo essere stata riscaldata dal Sole. Nella storia antica della Terra, l'effetto serra ha reso il pianeta idoneo alla vita come; senza effetto serra il pianeta sarebbe gelido e in massima parte ghiacciato.

Negli ultimi decenni tuttavia le emissioni di anidride carbonica dovute alle attività umane hanno acquisito un'entità tale da incrementare notevolmente l'effetto serra. Oggi la grande maggioranza della comunità scientifica ritiene che ci sia un rapporto di causalità tra le emissioni antropiche di gas serra e l'aumento della temperatura globale, in virtù dell'effetto serra. Da decenni gli scienziati studiano il problema e il rapporto di causalità è ormai consolidato. Gli studi scientifici non si limitano soltanto al surriscaldamento planetario, ma anche alle sue conseguenze quali i gravi effetti sugli ecosistemi, l'estremizzazione dei fenomeni meteorologici, l'innalzamento dei livello dei mari e l'impatto negativo sulle attività umane. Approfondiremo questi aspetti nell'esercizio 4.

EDUCAZIONE CIVICA digitale

Dalla correlazione alla causalità

Applicazione al caso del global warming

METTIAMOCI ALLA PROVA

1 Causalità e correlazione

Crea un file csv, analogo a *voti_studio.csv*, contenente la media dei tuoi voti mensili e il numero medio giornaliero di ore che passi a studiare. Crea i grafici delle due variabili: esiste una correlazione? Confronta i tuoi grafici con quelli dei tuoi compagni. Come può essere spiegata l'assenza di correlazione tra la media dei voti e il numero di ore di studio che eventualmente si riscontra? Discutine in classe.

2 Causalità e correlazione

1. Ricerca sul web le temperature medie climatiche per ogni mese dell'anno del tuo comune di residenza. (variabile A)
2. Per ogni mese dell'anno stima il numero di giorni in cui per uscire di casa indossi una giacca. (variabile B)
3. Per ogni mese dell'anno stima il numero di giorni in cui vai a scuola. (variabile C)
4. Per ogni mese dell'anno stima il numero di giorni in cui giochi ai videogame. (variabile D)
5. Utilizza un editor di testo per creare un file csv in cui ogni riga contenga un mese e i dati A, B, C e D riferiti a quel mese
6. Crea 6 grafici che ti consentano di individuare se c'è correlazione tra A e B, tra A e C, tra A e D, tra B e C, tra B e D, tra C e D.
7. Sei in grado di individuare se c'è causalità tra le diverse coppie di variabili (per esempio A causa B, oppure B causa A)?

3 I gas serra e il *global warming*: correlazione

Il file *anomalia_emissioni.csv* contiene tre colonne:

- Anno (dal 1880 al 2010)
- Anomalia di temperatura globale in gradi Celsius
- Emissioni totali di anidride carbonica di origine antropica in milioni di tonnellate.

Crea il grafico che mostri gli andamenti dell'anomalia globale di temperatura e delle emissioni totali di anidride carbonica. Rilevi correlazione?

Fonti dei dati:
https://www.climate.gov/maps-data/dataset/global-temperature-anomalies-graphing-tool
http://cdiac.ess-dive.lbl.gov/ftp/ndp030/CSV-FILES/global.1751_2014.csv

4 I gas serra e il global warming: causalità

Ricerca sul web cinque diverse fonti autorevoli e affidabili che trattino il legame tra la concentrazione in atmosfera di gas serra come l'anidride carbonica e l'aumento della temperatura media globale. Sulla base delle fonti che hai individuato, produci una presentazione da mostrare alla classe. Nella presentazione approfondisci almeno un grave effetto del *global warming* per la Terra, gli ecosistemi o l'umanità.

5 Una visualizzazione alternativa della temperatura che aumenta

Il professor Ed Hawkins dell'University of Reading ha ideato una visualizzazione alternativa (https://showyourstripes.info/s/globe) per rappresentare come è variata l'anomalia di temperatura del pianeta Terra nel corso degli anni. Lo scopo di questa visualizzazione è illustrare la gravità e la velocità del surriscaldamento climatico. Come vediamo nella figura sotto, il grafico, detto **warming stripes**, rappresenta ciascun anno mediante una striscia che viene colorata di un colore che è tanto più rosso intenso quanto più l'anomalia di temperatura è elevata. Il blu intenso corrisponde invece ad anomalie di temperatura negative. Utilizza i valori dell'anomalia di temperatura globale in gradi Celsius presente nel file *anomalia_emissioni.csv* per disegnare le tue warming stripes con il metodo che preferisci (Python, foglio di calcolo, oppure carta e colori).

Utilizza le immagini prodotte per sensibilizzare altri alunni della tua scuola sul problema del *global warming.*