

21 Variabilità, correlazione, note finali

21.1 Indici di dispersione ovvero variabilità

[...] *seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:
e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che ne magna due.* [Trilussa]

Gli indici dispersione ovvero variabilità cercano di quantificare con 1 numero la non omogeneità di dati numerici.

Sono statistiche di sintesi $f(x_1, \dots, x_n)$ da \mathbb{R}^n in \mathbb{R} , con n a priori indeterminato, proprio come le varie medie, viste in precedenza. È ovvio che dataset diversi possono avere diversi gradi di "disomogeneità", che vogliamo quantificare, anche se hanno la stessa media e/o mediana. Per esempio i 2 dataset, qua già ordinati,

$$2, 4, 6, 8, 10, 12, 14 \quad 5, 6, 7, 8, 9, 10, 11$$

hanno uguale media, e perfino mediana, sempre 8, ma è chiaro che i valori del secondo dataset variano meno, ovvero sono più addensati intorno alla media. Si pensi a dei redditi per esempio.

Campo di variazione, o *range*.

$$\text{range}(x_1, \dots, x_n) := \max(x_1, \dots, x_n) - \min(x_1, \dots, x_n).$$

Che vale 12 e rispettivamente 6 nei 2 dataset considerati.

Il range è molto sensibile agli eventuali outlier.

Il range come sopra definito è un numero, ma altri Autori chiamano range l'intervallo $[\min, \max]$.

Differenza interquartile. Ben poco sensibile agli outlier:

$$\text{iqr}(x_1, \dots, x_n) := q_{3/4} - q_{1/4}.$$

Varianza, deviazione standard, coefficiente di variazione.

Detta \bar{x} la media aritmetica del dataset x_1, \dots, x_n si definiscono

$$\text{population variance} \quad \sigma_X^2 = \text{Var}(x_1, \dots, x_n) := \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 \quad (68)$$

population standard deviation

$$\sigma_X = \text{sd}(x_1, \dots, x_n) := \sqrt{\text{Var}(x_1, \dots, x_n)} \quad (69)$$

$$\text{ovvero} = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n}}$$

coefficiente di variazione ovvero deviazione standard relativa

$$\text{se } \bar{x} \neq 0: \quad \sigma_X^* = \text{cv}(x_1, \dots, x_n) := \frac{\sigma_X}{\bar{x}} \quad (70)$$

e dell'ultima si può avere la forma percentuale moltiplicando per 100%, ottenendo la *deviazione standard percentuale*.

Alcuni Autori dividono per $|\bar{x}|$ invece che \bar{x} .

Esempio. Sul portale web di Our World in Data, generalmente affidabile (in particolare per aggiornamenti statistici sulla pandemia del covid) e consigliato al lettore, leggiamo⁽¹¹⁶⁾

As an aggregate of the regions with available data – Europe, North America, Australia, and East Asia – they found the mean male height to be 178.4 centimeters (cm) in the most recent cohort (born between 1980 and 1994). The standard deviation was 7.59 cm. (...) Women were smaller on average, with a mean height of 164.7 cm, and standard deviation of 7.07 cm.

e la maggior variabilità dell'altezza maschile (7.59 vs 7.07) si rivela solo apparente perchè dividendo per le altezze medie si ottiene, salvo decimali, lo stesso coefficiente di variazione del 4.3%.

(E lo scrivente si permette di ritenere che la piccola differenza sarebbe stata ancora più piccola se tutti i calcoli fossero stati fatti usando i logaritmi delle altezze – ovviamente privati delle unità di misura – arrivando a qualcosa di forse meno intelligibile ma più intrinseco della cosa).

Si trova scritto sia σ che sd che SD ; e sia σ_X^* che cv che CV .

¹¹⁶<https://ourworldindata.org/human-height>. Letto il 15 dicembre 2022.

Questi indici sono tutti sensibili agli outlier.

La varianza ha un'unità di misura diversa da quella del dataset.

La deviazione standard, di gran lunga preferita nelle Scienze Applicate alla Varianza che invece si preferisce in Matematica, ha la stessa unità di misura del dataset, per esempio litri invece che... litri quadrati.

Si noti il fattore $\frac{1}{n}$ nella varianza, e allora poi negli altri indici da essa dipendenti. Certi testi e software danno $\frac{1}{n-1}$. La questione [sarà approfondita](#) in Statistica Inferenziale, ma anticipiamo la questione, peraltro numericamente poco influente per n grande:

If the values instead were a random sample drawn from some large parent population (for example, they were 8 marks randomly and independently chosen from a class of 2 million), then one often divides by 7 (which is $n - 1$) instead of 8 (which is n) in the denominator of the last formula. In that case the result of the original formula would be called the sample standard deviation. Dividing by $n - 1$ rather than by n gives an unbiased estimate of the variance of the larger parent population. This is known as Bessel's correction.⁽¹¹⁷⁾

Nota. Grande varianza ovvero deviazione standard: valori dispersi.

Piccola varianza ovvero deviazione standard: valori addensati presso la media.

La deviazione standard può essere considerata piccola oppure grande relazionandola alla media.

Per la varianza, esistono varianze più piccole (valori più addensati) o più grandi di altre (valori più dispersi) ma non ha molto senso considerare grande o piccola una singola varianza. Similmente la deviazione standard.

Esempio 1.

¹¹⁷da Wikipedia (in inglese), l'enciclopedia libera, alla voce Standard deviation, in https://en.wikipedia.org/wiki/Standard_deviation, il 24 ottobre 2019

the average height for adult men in the United States is about 70 inches (177.8 cm), with a standard deviation of around 3 inches (7.62 cm).⁽¹¹⁸⁾

Esempio 2. In un articolo scientifico⁽¹¹⁹⁾ hanno confrontato, mediante le deviazioni standard, la variabilità dell'ora dell'andare a dormire con l'ora dell'alzarsi in popolazioni di tipo primitivo, trovando maggior variabilità nel primo orario:

The SD of sleep onset times exceeded the SD of sleep offset times

Il confronto fra due deviazioni standard di grandezze omogenee è ben sensato.

- Per i dataset visti in precedenza si calcolino i 5 indici sopradetti.

21.2 Covarianza; correlazione; retta di regressione

Dati 2 dataset numerici di uguale numerosità

$$X : x_1, \dots, x_n \quad Y : y_1, \dots, y_n$$

si definisce la loro *covarianza* (ovvero *covarianza di 2 n-uple di numeri*, ovvero *covarianza osservata*)

$$Cov(x_1, \dots, x_n; y_1, \dots, y_n) := \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

dove \bar{x} e \bar{y} sono le medie dei 2 dataset. Si trova denotata anche anche $\sigma_{X,Y}$, oppure con *cov*(... minuscolo. Si definisce anche l'*indice di correlazione di [Bravais-]Pearson*, o *indice di correlazione lineare* (e al posto della parola *indice* altri scrivono *coefficiente*)

$$\rho_{X,Y} := \frac{Cov(x_1, \dots, x_n; y_1, \dots, y_n)}{\sqrt{Var(x_1, \dots, x_n)Var(y_1, \dots, y_n)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

¹¹⁸da Wikipedia (in inglese), l'enciclopedia libera, alla voce Standard deviation, in https://en.wikipedia.org/wiki/Standard_deviation, il 24 ottobre 2019

¹¹⁹:"Natural sleep and its seasonal variations in three pre-industrial societies", by Yetish G, Kaplan H, Gurven M, Wood B, Pontzer H, Manger PR, Wilson C, McGregor R, Siegel JM8. Curr Biol. 2015 Nov 2;25(21):2862-2868. doi:<https://it.overleaf.com/project/6179b2b3c248087503c542f9> 10.1016/j.cub.2015.09.046. Epub 2015 Oct 17.

e questo dà un'idea di come varino le y_j relativamente alle x_i : un indice prossimo a 1 suggerisce che al crescere delle x_i crescano linearmente le y_i , un indice prossimo a -1 suggerisce che al crescere delle x_i decrescano linearmente le y_i , e un indice prossimo a 0 indica la non presenza di tali relazioni approssimate. In ogni caso

$$-1 \leq \rho_{X,Y} \leq 1.$$

Coi dataset $X : 2, 7, 4, 3$, e $Y : 3, -1, 0, 2$ si calcoli l'indice di correlazione, e si rappresentino sul piano cartesiano i punti (x_i, y_i) . In un senso non banale, una retta $y = mx + q$ si avvicina meglio ai punti se ha coefficienti

$$m := \frac{\sigma_{X,Y}}{\sigma_X^2} \quad q := \bar{y} - m\bar{x} \quad \text{equivalentemente: } m := \frac{\sigma_Y}{\sigma_X} \rho_{X,Y}$$

e in questo caso si chiama *retta di regressione* (dietro, vi è il *metodo dei minimi quadrati*, che non approfondiamo).

Per esempio per i punti $(2, 1)$, $(1, 3)$, $(3, -1)$, ovvero per i dataset $X : 2, 1, 3$ e $Y : 1, 3, -1$, si trova $y = -2x + 5$, che passa *esattamente* per i punti. È $\bar{x} = 2$, $\bar{y} = 1$, $\sigma_X^2 = \frac{2}{3}$, $\sigma_{X,Y} = -\frac{4}{3}$. E con $y_2 := 3.14159265$ si trova $y \approx -2.07x + 5.19$. Si facciano i disegni.

Il coefficiente di correlazione e la retta di regressione sono fra i più usati strumenti matematici delle Scienze Biomediche. Si segua questo [LINK->](#)

È la retta che in assoluto meglio riassume uno scatterplot?

No, tant'è vero che facendo la retta di regressione $y = mx + q$ si trova in generale una retta diversa dalla retta di regressione $x = m'y + q'$, eppure lo scatterplot è lo stesso. Bisogna fissare la variabile che riteniamo indipendente, quella in questo paragrafo denotata con x . La questione è sottile e non la approfondiamo.

21.3 L'inimicizia fra progresso e variabilità

Nominate la variabilità a un Ingegnere, in pratica colui che progetta quasi tutte le cose materiali che vedete intorno a voi, e gli

si drizzeranno i capelli in testa: per l'Ingegneria la riduzione della varianza nei processi produttivi è obiettivo prioritario.

Se guardate i flaconi di vetro in un museo di antichità romane, vedete che sono alquanto variabili nella forma: sono fatti a mano uno per uno, di vetro soffiato. Se una farmacia ne comprava una certa quantità per mettervi un certo farmaco, erano tutti diversi. Se guardate in una farmacia odierna i flaconi di un qualunque prodotto, sono tutti uguali, o, per meglio dire, la variabilità – che a livello molecolare comunque esisterà – è ridottissima.

Questo sopradetto è un fatto generalissimo del progresso.

Tutti gli oggetti di uso tecnico e/o scientifico, dai flaconi alle automobili, sono prodotti con una variabilità piccolissima, e la sostanziale inesistenza di variabilità – comunque esistente a un livello microscopico – è uno dei principali indicatori della bontà di un processo produttivo. Ancora all'inizio del '900 auto e aerei erano costruiti uno per uno.

Potete accettare che i le clementine e i limoni, prodotti nei campi, abbiano una certa variabilità nelle dimensioni, ma non certo

le bottiglie d'acqua minerale

le siringhe

i flaconi di un certo farmaco

le compresse o capsule di un certo farmaco

i chip dei computer, eccetera eccetera.

E comunque anche per i limoni, e gli altri prodotti agricoli, col tempo i vari standard legislativi sono sempre più restrittivi riguardo la variabilità: i limoni di *size code 7*, i più piccoli, devono avere diametro fra 45 e 52 mm, e comunque anche senza quel codice non si possono vendere come limoni dei limoni di diametro minore di 45 mm. (Non stiamo qua negando che siano limoni ma non potete venderli come limoni; ve li accetteranno forse come materiale per compost). [LINK ->](#)

È in atto un secolare processo che tende a ridurre sempre più la variabilità degli oggetti prodotti in una determinata serie.

Icasticamente:

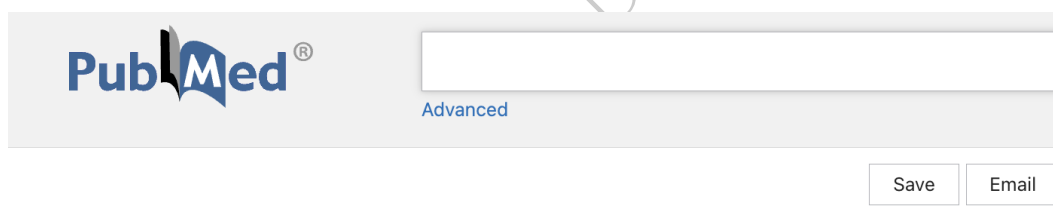
il progresso odia la variabilità

L'uniformità ovvero bassissima variabilità, misurata su qualche parametro (per esempio il diametro dei limoni) consente

- risparmio economico
- risparmio di tempo

In campo farmaceutico, il danno della variabilità è evidente nella necessità di avere dosaggi garantiti uniformi in ogni flacone, in ogni capsula, in ogni compressa, eccetera. È ovvio.

Fra 2 compresse di aspirina di una stessa marca c'è da aspettarsi una variabilità infinitesimale.



> [Vaccine](#). 2021 Oct 29;39(45):6565-6569. doi: 10.1016/j.vaccine.2021.09.065. Epub 2021 Oct 6.

All vials are not the same: Potential role of vaccine quality in vaccine adverse reactions

Figure 28: Anche il farmaco più miracoloso che l'Uomo avesse mai prodotto potrebbe talvolta fare male se – per ipotesi assurda – il prodotto che arriva all'utenza non è esattamente uguale a quello previsto.

21.4 Note finali sulla Statistica Descrittiva

Ad un livello basico, nella Statistica Descrittiva **si vuole riassumere molti valori (*dataset*) con 1 valore o pochi valori, e/o con un diagramma.**

Quest'opera di sintesi non è banale.

Viene fatta per comprendere i dati, divulgarli, confrontarli. Alcuni lo fanno perfino per presentarli manipolativamente; allora noi dovremo attrezzarci conoscitivamente per non essere tratti in inganno.

Questo è centrale in Farmacia. Basta ridefinire 1 soglia, e qualche miliardata di persone da sane potrebbero trovarsi riclassificate come malate, verosimilmente bisognose di farmaci.

Leggiamo su un articolo⁽¹²⁰⁾ del British Medical Journal:

There's a lot of money to be made from telling healthy people they're sick. Some forms of medicalising ordinary life may now be better described as disease mongering: widening the boundaries of treatable illness in order to expand markets for those who sell and deliver treatments. Pharmaceutical companies are actively involved in sponsoring the definition of diseases and promoting them to both prescribers and consumers. The social construction of illness is being replaced by the corporate construction of disease.

21.5 Bugie, bugie cattive, e statistica

Da più di un secolo si dice "[lies, damned lies, and statistics](#)". Ai numeri si può far dire quel che si vuole o quasi.

¹²⁰Moynihan R, Heath I, Henry D. Selling sickness: the pharmaceutical industry and disease mongering. *BMJ*. 2002 Apr 13;324(7342):886-91. doi: 10.1136/bmj.324.7342.886. PMID: 11950740; PMCID: PMC1122833.

Attenzione!

PARTE I – DATI VERI

21.6 La scelta dei parametri. Cosa misurare?

In questa trattazione elementare siamo sempre partiti da dati numerici, ma a monte c'è il problema di quali numeri considerare in una data situazione.

Se consideriamo un quadrato di lato 2 ovvero area 4, e 2 quadrati di lato 5 ovvero area 25, e volessimo rappresentare un "quadrato medio", facendo la media dei lati abbiamo 4 (con area 16), invece facendo la media delle aree abbiamo 18 (con lato $3\sqrt{2} \approx 4.24$).

Qual è il senso di quelle tabelle che mettono ai primi posti il peperoncino per contenuto di vitamina C? Il peperoncino si mangia, eventualmente, a grammi, mica a etti come certa frutta, che, certo, contiene meno vitamina C *all'etto*, ma ne contiene di più *per porzione* (*ragionevole*, per quanto ciò possa significare). La scelta di un diverso parametro cambia completamente l'impostazione della questione.

"Il Molise ultimo per decessi covid"

sì ne aveva solo 31, ad una certa data, meno di ogni altra regione, ma ha anche pochi abitanti. Considerando i decessi per 100mila abitanti ne aveva 10, più della Basilicata che ne aveva 8, pur avendo un numero maggiore di decessi. Non è più ultimo.

Se leggiamo

La regione X prima per tumori

potremmo farci l'idea che la regione X sia più inquinata, o meno assistita medicalmente, magari peggio governata da quel noto partito che ha la maggioranza nella regione X – associazioni d'idee che magari sono il motivo reale per cui è stato pubblicato quell'articolo sui media – ma il contenuto informativo di quel titolo è seminullo, e quelle argomentazioni che ci sorgono istintivamente sono

praticamente insensate. Potrebbe ben essere che nella regione X i tumori siano 10 volte meno probabili *per ogni singolo residente* che nella regione Y, ma la regione X ha più del decuplo di abitanti della regione Y, per cui *il numero assoluto di casi* della regione X sopravanza quello della regione Y, molto più inquinata e meno assistita medicalmente.

Nella situazione considerata, altrettanto legittimo sarebbe stato il titolo

La regione Y prima per tumori

e in un caso numerico estremo potrebbe essere vero addirittura

*Nella regione X la **minima** incidenza di tumori*

(Per una disamina del termine *incidenza*: ->[Link](#)).

Anche qua la scelta di un diverso parametro cambia completamente l'impostazione della questione.

Numero assoluto dei morti: non dice nulla su inquinamento, malasanità...

Tasso di mortalità: serio indizio, su varie possibilità.

Il numero assoluto, spesso espresso da cifre ad alto impatto emotivo ("milioni di tumori"), da un punto di vista medico e farmaceutico ci dice meno della frequenza relativa, che in generale si potrà misurare in casi per 1 000 residenti, o per 100 000, o per milione. (Per evitare scritte come 0.000077, scarsamente leggibili: molto più chiaro, come nell'esempio del link sopra riportato, 7.7 casi per 100 000).

Si noti che la frequenza relativa corrisponde alla probabilità, per 1 singolo residente preso a caso. Per esempio 50 casi ogni 1000 abitanti dà una probabilità di $\frac{1}{20}$. Si tratta della concezione frequentista della probabilità.

La frequenza assoluta ha una diversa importanza: corrisponde alla *dimensione economica* complessiva della questione: è ovvio che se in un certo luogo ci sono più casi di una malattia usualmente trattata con un certo farmaco, là c'è un maggior *mercato* per quel farmaco.

In ambito medico e farmaceutico. Ragionando con la logica comune – sbagliata – trovare una riduzione nelle aritmie che causano morte è un buon risultato per un farmaco, e si verifica abbastanza facilmente; verificare la sopravvivenza a lungo termine è molto più oneroso (e in passato ha dato risultati opposti).

La sopravvivenza a 5 anni dalla diagnosi di cancro è un indice di altissimo valore ma richiede molto tempo e denaro; la riduzione della massa tumorale in un breve tempo si verifica prima.

Cosa misuriamo è un serio problema generale della Medicina e della Farmacologia.

Vogliamo massa tumorale ridotta? A 2 mesi o a 6 mesi o a 1 anno dall'inizio del trattamento, o quando? O vogliamo diminuita mortalità, ma, di nuovo, quando misurata? O vogliamo aumentata sopravvivenza complessiva? Oppure aumentata *progression-free survival*? O migliore *response rate*, tasso di risposta del tumore alla terapia valutato misurando qualche *biomarker*? O diminuita spesa? (Che anche quello conta). O migliorata qualità della vita? (Sotto-problema: misurata come?) O un po' tutto, con una media pesata con coefficienti da decidere? In ognuno di quei casi si otterranno risultati in generale diversi.

Leggiamo sul Journal of Clinical Oncology⁽¹²¹⁾

produce results that are clinically meaningful to patients
(ie, significantly improved survival, quality of life [QOL],
or both)

(La riduzione della massa tumorale, così tanto usata, appare per certi versi la cosa che *meno* possa interessare il paziente).

Lo stesso discorso vale praticamente per ogni patologia.

Icasticamente:

L'operazione è riuscita ma il paziente è morto

Leggiamo in un articolo scientifico:⁽¹²²⁾

¹²¹American Society of Clinical Oncology Perspective: Raising the Bar for Clinical Trials by Defining Clinically Meaningful Outcomes. By: Lee M. Ellis et al. DOI: 10.1200/JCO.2013.53.8009 Journal of Clinical Oncology 32, no. 12 (April 20, 2014) 1277-1280.

¹²²Progression-Free Survival: Meaningful or Simply Measurable? By: Christopher M. Booth, Elizabeth A. Eisenhauer. DOI: 10.1200/JCO.2011.38.7571 Journal of Clinical On-

The last few years have seen an increase in the number of randomized controlled trials (RCTs) of new agents in metastatic solid tumors using progression-free survival (PFS) as the primary end point. Some trials showing improvement in PFS, without a corresponding increase in overall survival (OS), have led to approval of new drugs and/or changes in standard of care. This suggests a growing belief in the oncology community that delaying progression in metastatic disease is a worthy goal, even if OS is not improved. But is a new treatment that improves PFS really an advance for patients? Or is it only lowering the bar to declare some of our much-heralded new molecular targeted therapies?

21.7 Rispetto a quale standard fare le statistiche?

Una fonte di problematicità è quale standard scegliere. Se per esempio volessimo vedere se la carenza di una certa sostanza nel sangue è statisticamente correlata ad una malattia, avremmo – oltre al problema della diagnosi della malattia nei soggetti, problema medico che non considereremo – il problema di quali soggetti considerare carenti e quali no, in base alle analisi chimiche del sangue: infatti ci sono molti standard per i vari livelli delle sostanze nel sangue, e in generale per i parametri biomedici, come l'essere sovrappeso, e come la mortalità infantile e l'aspettativa di vita alla nascita, standard che variano nel tempo e nelle grandi entità scientifiche, sanitarie e/o geopolitiche ragionevolmente considerabili: Ministero della Salute italiano, Comunità Europea, OMS (Organizzazione Mondiale della Sanità), ONU, USA (NIH, CDC...), la rivista scientifica [The Lancet](#) e molte altre...

L'aspettativa di vita delle donne in Italia, e il ranking dell'Italia per essa, sarebbe⁽¹²³⁾

86.49 – 3^a secondo l'ONU (luglio 2015), dati 2010-2015

85.6 secondo l'OECD (2016), che considera solo stati OECD

85.1 – 12^a-13^a secondo [The World Factbook](#) (2017)

cology 30, no. 10 (April 01, 2012) 1030-1033. <https://ascopubs.org/doi/full/10.1200/JCO.2011.38.7571>

¹²³Si veda Wikipedia, l'enciclopedia libera, alla voce [List of countries by life expectancy](#).

84.8 – 7^a secondo l’OMS (maggio 2016), dati 2015

83.90 – 7^a secondo (un articolo di) The Lancet (2012).

Abbastanza variabilità per fare affermazioni molto diverse, pure autorevoli, specialmente usando il dato del ranking, molto fuorviante: per esempio San Marino viene considerato solo nella lista del World Factbook.

Gli standard variano nel tempo, e a seconda delle agenzie che li fissano, pur autorevoli.

Esempio. Attualmente (giugno 2019) leggiamo su sito governativo statunitense a proposito dell’anilina, in <https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=79>

Cancer Classification: EPA: Probable human carcinogen. IARC: Not classifiable as to carcinogenicity to humans. NTP: Not evaluated

(Si veda per esempio il [Link ->](#)).

Altro esempio: leggiamo⁽¹²⁴⁾

several factors influence the likelihood of progression from prediabetes to Type 2 diabetes, including how we define prediabetes (there is no internationally agreed-upon standard)

Varia nel tempo e nei luoghi cosa sia un farmaco...

...o piuttosto una droga (come l’eroina, “venduta liberamente dalla multinazionale farmaceutica Bayer dal 1899”⁽¹²⁵⁾ per varie patologie), o un cancerogeno⁽¹²⁶⁾ (come la formaldeide: “Fino al momento

¹²⁴Online text in <https://www.levelshealth.com/blog/how-to-reverse-prediabetes> read December 15, 2022.

¹²⁵<https://it.wikipedia.org/wiki/Eroina>:

L’intento era quello di ottenere una molecola più efficace della codeina nel sedare la tosse, la tubercolosi e le patologie respiratorie. (...) In breve tempo l’impiego terapeutico si ampliò alle più disparate patologie pneumologiche, ma anche neurologiche, ginecologiche, o a semplici dolori; si diffusero pertanto svariate preparazioni farmaceutiche acquistabili liberamente, e questo fece sì che l’eroina divenisse velocemente uno dei farmaci più venduti in assoluto.

¹²⁶<https://it.wikipedia.org/wiki/Formaldeide>:

L’Associazione Internazionale per la Ricerca sul Cancro (AIRC) sin dal 2004 ha inserito la formaldeide nell’elenco delle sostanze considerate con certezza

in cui il prodotto è stato bandito, era uno degli ingredienti delle pastiglie Formitrol per il mal di gola"⁽¹²⁷⁾).



**Ah... se avesse preso
in tempo il Formitrol!**

Avrebbe evitato quel potente mal di gola che non lo lascia quasi respirare.

Date in tempo il Formitrol ai vostri bambini, se volete difenderli dai malanni invernali.

L'uso anche prolungato del Formitrol non dà luogo ad alcun disturbo.

Il Formitrol piace ai bambini per il suo gradevole sapore aromatico.

Formitrol

chiude la porta ai microbi



DR. A. WANDER S.A. VIA MEUCCI 39 MILANO

Le conoscenze scientifiche variano nel tempo, e con esse i vari standard, seppure in parte contraddittori, e in uno stesso tempo e luogo esistono opinioni mediche opposte.

cancerogene per la specie umana. (...) Dal 1° gennaio 2016, la formaldeide è passata dalla classificazione di "sospettato di provocare il cancro" a "può provocare il cancro" (Regolamento (UE) N. 605/2014).

¹²⁷ <https://it.wikipedia.org/wiki/Formaldeide>



Un orbitoclasto è uno strumento chirurgico utilizzato per eseguire lobotomie transorbitali. Esso fu inventato dal dottor Walter Freeman nel 1948 come sostituto all'unica forma di leucotomo allora utilizzato per gli interventi di tal genere. Lo strumento in questione è a forma di punteruolo rompighiaccio, con alcuni segni incisi sul manico ed una struttura a cuneo sull'estremità esterna. L'operazione consiste nell'inserire lo strumento dietro la palpebra del paziente, perforare il sottile strato osseo che è posto inferiormente ad essa, e attraverso colpi di martello far penetrare lo strumento fino al lobo frontale. L'orbitoclasto viene poi mosso in varie direzioni al fine di distaccare i lobi frontali dal talamo. Nel 1948, Freeman perfezionò la procedura introducendo un profondo taglio frontale ed un movimento ulteriore della punta nel lobo, al fine di evitare - come succedeva occasionalmente in precedenza - che l'orbitoclasto si rompesse nella testa del paziente, necessitando così una rimozione chirurgica. In più, Freeman riuscì a commissionare la produzione di orbitoclasti di materiale più resistente ed affidabile.

Poiché durante i primi interventi venivano usati veri e propri punteruoli da ghiaccio, la procedura fu soprannominata "lobotomia con punteruolo da ghiaccio" ("ice pick lobotomy"). Nel 1949 fu assegnato il Premio Nobel per la Medicina a Egas Moniz,

"per la sua scoperta del valore terapeutico della lobotomia in alcune psicosi". (Da Wikipedia, l'enciclopedia libera.



21.8 Col ranking facciamo ben figurare il Paraguay

Supponiamo che abbiamo questi dati, relativi a una cosa qualunque che i lettori di un giornale percepiscono in generale come *buona*:

USA 3723

Canada 3518

Paraguay 18

India 16

eccetera

Oggi vogliamo dire bene del Paraguay, a beneficio di qualche potentato, ovvero interesse geopolitico – il lettore acuto può trovarne uno – a cui il nostro giornale, per il quale scriviamo, risponde. Che fare, con quel misero valore 18? Potremmo dire che

il Paraguay è terzo dopo USA e Canada, davanti all'India.

21.9 Illusioni percettive nella presentazione dei dati

In questo paragrafo usiamo sia lo standard del punto decimale che della virgola decime.

La dose giornaliera raccomandata di vitamina C è stata portata da 60 milligrammi a 90 milligrammi:

possiamo dire che adesso è del 50% maggiore

possiamo che prima era del 33% minore.

Ovviamente col primo modo diamo l'impressione di una variazione maggiore, almeno al lettore poco esperto di numeri.

L'epidemia di influenza del 2016/17 ha fatto circa 25 000 morti, circa il 28% in meno della prima ondata del covid-19 nel 2020 che ha fatto circa 35 000 morti, il 40% in più...

In linea generale, i titoli faranno più conseguenze dei testi sottostanti.

Se il partito dei Vispi Volpini un anno riceve lo 0,6% e l'anno dopo lo 0,9% (è un partito piccolo...) la televisione dirà che è cresciuto dello 0,3%, ma il Corriere dei Vispi Volpini dirà che il loro elettorato è aumentato del 50%, il che è vero.

(Magari anche un telegiornale lo dirà, chissà chi è il direttore...)

N.d.A. Ricordo, decenni fa, un esempio di presentazione fuorviante dei risultati elettorali del tipo "il partito A scende dal ...% al ...%, il partito B prende il ...% contro il ...% della volta precedente" dove nel profluvio di cifre e decimali l'ascoltatore medio ricorda solo che A *scende* e B *prende*, ma in effetti B scendeva più di A...

Fare raffronti con dati collaterali può permettere di titolare "aumento" un articolo che poi parla di un fenomeno in diminuzione, ma moltissimi si limiteranno a leggere il titolo.

In ambito medico, affermare che con diagnosi precoci si allunga la speranza di vita dalla diagnosi può essere molto fuorviante: “Una volta il paziente viveva mediamente due anni dalla diagnosi, adesso dodici anni”... ma facendo la diagnosi mediamente quanti anni prima?

Attenzione alle fallacie della memoria: [vi pare di ricordare](#) che secondo l’Organizzazione Mondiale della Sanità ogni anno 12.6 milioni di persone muoiono a causa dell’inquinamento, e [adesso leggete](#) che ogni anno 4.2 milioni di persone lasciano questo mondo a causa dell’inquinamento dell’aria? Basta rileggere con attenzione.

Attenzione soprattutto alle rappresentazioni grafiche.

Iniziare un diagramma a colonne o un grafico da un certo tempo invece che da un tempo precedente può permettere di fare affermazioni completamente diverse, magari entrambe formalmente vere, ma che genereranno emozioni e poi reazioni completamente diverse. Grafici che sull’asse delle ordinate non iniziano da 0 possono dare l’impressione di variazioni enormi, anche raddoppi o dimezzamenti, a variazioni minuscole.

Diagrammi a torte tridimensionali in rappresentazione prospettica possono far apparire più grandi certi dati rispetto agli altri.

21.10 Di cosa parliamo?

Diamo solo un cenno di un argomento ciclopico. Ogni persona vive immersa in un flusso di informazioni che gli arrivano, in base al quale costruisce la sua percezione della realtà, che lui crede vera.

Anche se tutte quelle informazioni fossero vere (per ipotesi assurda, ovvio) l’immagine complessiva che l’individuo si forma è in generale (molto!) distorta: una *diversa scelta* di informazioni *altrettanto vere*, da far arrivare coi media – dal livello più basso fino agli articoli scientifici – creerebbe in quella stessa persona un’immagine del

mondo completamente diversa, per molti versi quasi opposta.

Supponiamo qua in via ipotetica che sul pianeta Kepler-22 b, simile alla Terra, il danno, in qualche modo misurato, delle sigarette si disponga a campana gaussiana, con un'ampia zona mediana di un danno medio per un grande generalità di fumatori. E che il danno delle sigarette elettroniche si disponga analogamente, ma più a sinistra sull'asse delle ascisse, cioè con un danno mediamente minore. Per quanto il danno sia mediamente minore, in moltissime persone è grave, per molte addirittura gravissimo, per alcune perfino maggiore di quello medio delle sigarette. Ecco, dai, invece di fare alcuni milioni di articoli su quelli mediamente meno danneggiati (impossibile a farsi), e invece di dire che in media il danno è minore (chi mai ci obbliga?) facciamo un bell'articolo titolando "ragazzo fuma sigarette elettroniche e ha i polmoni devastati".

La gente dirà: ecco, vedi, non serve, soldi buttati, meglio continuare con le sigarette normali!

È impressionante quante persone ragionino per casi singoli, appresi dai media, senza garanzia che siano rappresentativi di situazioni medie invece che eccezionali.

21.11 Cherry picking – bias di selezione

Affine alla questione precedente. Scegliere le prove favorevoli a una tesi e ignorare quelle contrarie. Se un proprietario di media volesse sostenere la tesi che bisogna chiudere le scuole per evitare il propagarsi del covid, potrebbe cercare di citare affermazioni favorevoli alla sua tesi, al livello autoritativo maggiore possibile, trascurando le prove contrarie; e similmente se vuole far passare la tesi opposta. Troverà articoli scientifici per qualunque dei 2 desiderata. L'articolo di review sistematica in <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8375447/> considera decine articoli scientifici di alto livello (dopo averne scartati migliaia di altri) e verifica che metà trovano che la chiusura delle scuole ha ridotto la diffusione

del contagio, e metà hanno trovato che non l'ha ridotto.

21.12 Approfittare delle fluttuazioni statistiche



Mentre le funzioni elementari di base della Matematica hanno andamenti molto regolari, come x^2 o x^3 , e similmente le successioni propriamente dette come quella di Fibonacci e quella dei fattoriali, le successioni di valori numerici rappresentativi di fenomeni reali normalmente presentano nel tempo fluttuazioni, alternando crescite a diminuzioni.

Esempio, morti in Italia:

2013 600 000

2014 599 000 $-1\ 000$ rispetto all'anno precedente

2015 653 000 $+53\ 000$

2016 616 000 $-37\ 000$

2017 650 000 $+34\ 000$

2018 632 000 $-18\ 000$

2019 645 000 $+13\ 000$

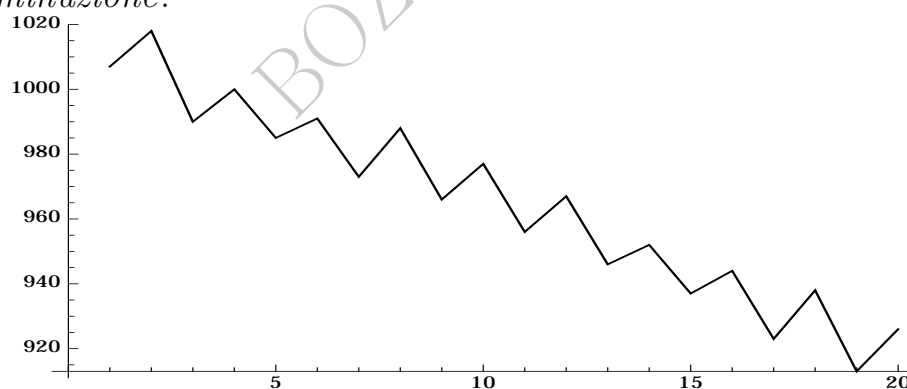
Consideriamo come ulteriore esempio il grafico *Sperimentazioni autorizzate dall'Autorità competente* a pagina 10 del documento [La](#)

Sperimentazione Clinica dei Medicinali in Italia – 17° Rapporto Nazionale – Anno 2018 dell’AIFA, Agenzia Italiana del Farmaco. L’ultimo dato è del 2017 e relativamente ad esso potremmo dire *veridicamente*

- aumentate rispetto al 2000
- aumentate rispetto a 15 anni prima
- diminuite rispetto all’anno precedente
- al livello minimo da 10 anni

ottenendo nel lettore emozioni diverse, che potranno supportare interventi decisionali diversi.

Si noti che anche in situazioni di palese diminuzione sostanziale sul lungo periodo, tali fluttuazioni possono permettere in vari anni di parlare di aumento. E negli anni in cui proprio non si può dirlo... si può parlare della situazione dell’anno precedente, che magari vedeva un aumento rispetto all’anno precedente – mica il lettore sa qual è l’ultimo dato rilasciato, in generale. Agendo così, si potrà *ogni anno* parlare di *aumento* in questo esempio di evidente *sostanziale diminuzione*:



21.13 Capziosità verbali

Un’intera enciclopedia si potrebbe scrivere, di infiniti volumi, sulle capziosità verbali, e, tranne che per la Medicina e la Farmacia, in cui c’è da piangere, sarebbe molto divertente.

Leggiamo per esempio in un articolo⁽¹²⁸⁾ online:

(Titolo)

Perché le trivelle non c'entrano niente con il terremoto nelle Marche

(Sottotitolo)

Le insinuazioni dei contestatori dell'estrazione di gas sul territorio italiano e la spiegazione dell'esperto

(Corpo dell'articolo)

(...) ha chiarito che "la trivellazione di per sé non comporta alcuno scuotimento sismico, sono semmai le attività di estrazione e stoccaggio che possono farlo. (...)

Insomma i contestatori delle trivelle sono fuori strada: che si trivelli pure, come *divertissement*, casomai si contestino le successive "attività di estrazione e stoccaggio".

21.14 Illustrazione di un triplice esempio reale

Aumento dell'aspettativa di vita negli USA: [link<-](#)



¹²⁸<https://www.today.it/cronaca/trivelle-terremoto-oggi-9-novembre-2022-marche.html> letto il 24 novembre 2022

Aumento del [tasso di mortalità negli USA](#):

2009	7.9
2010	8.0
2011	8.1
2012	8.1
2013	8.2
2014	8.2
2015	8.4
2016	8.5
2017	8.6

Capire questi diagrammi richiede:

- 1) conoscere la Statistica Descrittiva, notando in particolare che il diagramma considera solo gli ultimi 3 anni; il diagramma ha una linea base posta a circa 78.5 anni.
 - 1 bis) conoscere la Demografia, che possiamo considerare semplicemente una branca specialistica della Statistica Descrittiva. Questo ci farebbe capire che non è per nulla strano che si alzino da tempo i tassi di mortalità, anche prima della riduzione dell'aspettativa di vita. Il tasso di mortalità può ben alzarsi mentre l'aspettativa di vita si alza, proprio perché la popolazione mediamente invecchia, in compresenza di un fattore demografico completamente diverso: la scarsità di nuovi nati.
 - 2) Conoscere la situazione reale il che è completamente diverso dal guardare questi numeri o grafici.
- Per una parziale spiegazione della situazione reale dietro alla diminuzione dell'aspettativa di vita negli USA si segua questo [link](#)<-.

21.15 Conclusioni sulla fallace presentazione di dati veri

Pensare di aver capito una situazione perché si è vista una qualche statistica o lista di numeri è in generale *ingenuo*, e i grafici sono, per le persone non istruite, perfino più fuorvianti. Le parole, poi, possono essere ancora più ingannevoli anche se sembrano esprimere dati oggettivi, numerici: aumento, diminuzione...

È necessario acquisire una buona competenza statistica per non essere tratti in inganno. Sia da errori fatti in buona fede, che da manipolazioni volontarie, fatte da soggetti interessati.

Superato il problema della comprensione formale di statistiche e grafici, il problema delle statistiche e dei grafici, e perfino dei dati grezzi/completi, è che ce ne possono sempre essere altri, che non stiamo vedendo ovvero non ci stanno mostrando, che cambierebbero completamente la nostra prospettiva sui fenomeni considerati.

In linea generale, le statistiche economiche dei media non valgono la carta su cui sono scritte. Per le statistiche scientifiche, non è vietato sperare una maggiore oggettività.

Molti testi divulgativi affermano che *la maggior parte* dei contagi da toxoplasmosi non avvengono dai gatti ma dalla sporcizia. Anche se questa statistica – praticamente impossibile da verificare – fosse vera, quegli Autori magari trascurano di dire, nei loro articoli divulgativi, che sebbene il toxoplasma si trovi in tanti animaletti, si riproduce solo nei felini, per cui se tutti i gatti venissero inceneriti (orrore!) la toxoplasmosi semplicemente sparirebbe dall'Italia, dove praticamente l'unico felino che frequenta l'uomo è il gatto.

I gattofili non costituiscono una lobby molto pericolosa e allora possiamo prendercela scherzosamente con loro senza rischi, per spiegare cosa vuol dire comunicare una statistica fuorviante perché omette fatti che cambierebbero radicalmente la prospettiva.

Ma la triste verità per uno statistico è che le notizie di statistica medica che arrivano al grande pubblico dalla televisione e dai media *lasciano a desiderare*, diciamo così.

Ulteriori problematiche si presentano nella Statistica Inferenziale.

If you torture the data long enough, it will confess to anything. (In

questo [link](#) è in articolo su sito governativo statunitense).

Esempio finale. Da La Peste, aprile 2022, su Facebook:

Il primo gennaio 2015 sono decedute 126 persone di 95-99 anni, mentre il primo gennaio 2021 ne sono decedute 198; 72 in più? Per fare il confronto corretto dobbiamo considerare come è variata la numerosità della popolazione di persone di 95-99 anni. Queste erano 88.799 al primo gennaio 2015, ma sono diventate 158.581 al primo gennaio 2021. Con una variazione così imponente non si può fare un confronto tra le cifre assolute. Il problema si può risolvere considerando la mortalità. Il primo gennaio 2015 sono morte lo 0,142% delle persone di 90-95 anni mentre il primo gennaio 2021 ne sono morte lo 0,125%. La mortalità è un po' inferiore.

A seconda del risultato che si vuole ottenere, si potrà dire che ne sono morti di più, o di meno, dicendo sostanzialmente il vero.

PARTE II – DATI QUASI VERI (QUASI FALSI)

Già abbiamo visto in [18.9](#) le problematiche del trasformare dati ordinali in dati numerici. Nella pratica, se abbiamo già il dataset ordinale ma siamo ancora liberi nel fissare le corrispondenze numeriche, possiamo largamente manipolare il risultato.

21.16 Il valore anomalo: outlier

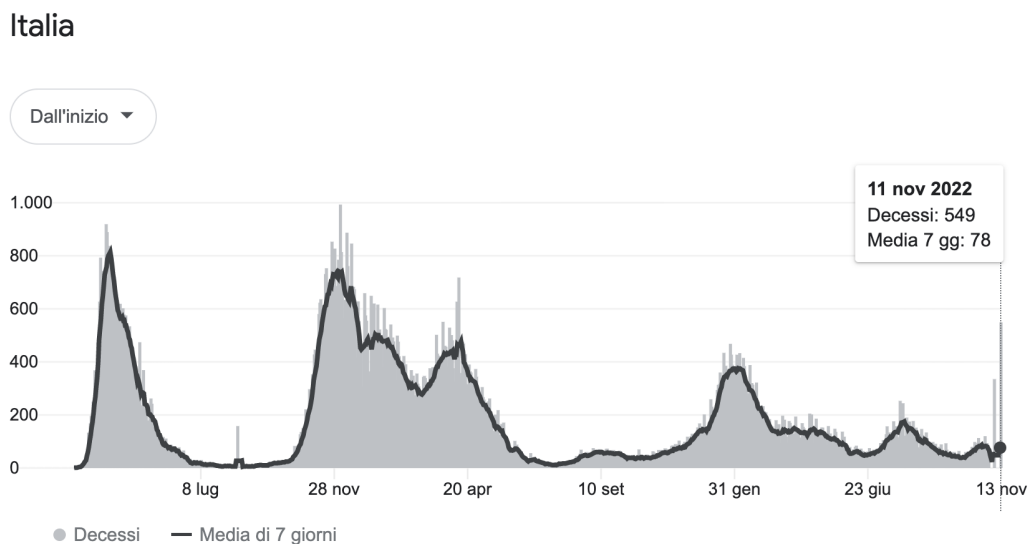


Figure 29: In questa peraltro autorevole statistica – screenshot da Google – l’esorbitante numero di 549 morti dell’11 novembre 2022 è dovuto al fatto che il Ministero ha cambiato modo di comunicare i dati: quei morti non sono di un solo giorno bensì 7.

Per un’infinità di motivi, in un dataset ci può essere qualche valore che il ricercatore ritiene *assurdo*, e tenderà ad eliminarlo prima di procedere con qualunque analisi dei dati stessi, per non comprometterla.

Questi *outlier*, valori anomali ovvero aberranti, provengono da risposte scherzose a questionari, truffe (“reddito zero”...), malfunzionamento di apparati, confusione fra punto decimale e virgola decimale, confusione fra milligrammi e microgrammi, eccetera. Ma talvolta sono dati *veri* benché *eccezionali*.

La decisione di quali valori siano da considerare outlier ed eventualmente da eliminare, può essere

soggettiva

oppure dettata da formule, che comunque non considereremo,

e in ogni caso è problematica, anche perché ce ne sono sono varie. In una ricerca seria *et* onesta, si *dovrebbe* avvertire se sono stati eliminati outlier.

Leggiamo su Wikipedia, l'enciclopedia libera, alla voce "Outlier", e non è chi non veda le pericolose conseguenze di ciò:

There is no rigid mathematical definition of what constitutes an outlier; determining whether or not an observation is an outlier is ultimately a subjective exercise. There are various methods of outlier detection. (...) Deletion of outlier data is a controversial practice

21.17 Omissione di dati ritenuti poco significativi

Dati ritenuti poco significativi vengono talvolta omessi, per esempio San Marino da molte statistiche; facendoli rientrare si possono ottenere affermazioni molto diverse, ma per certi versi effettivamente meno significative.

21.18 "Fatta la legge, trovato l'inganno"

Fissato uno standard per valutare statisticamente un qualche aspetto della realtà, qualcuno trova conveniente, diciamo così, fare azioni che migliorano di molto la valutazione statistica con scarso o nullo miglioramento della realtà. Se per esempio una legge attribuisce ai "comuni montani" particolari finanziamenti, fissando ad una certa altitudine del municipio la denominazione di "comune montano", state sicuri che qualche comune sposterà la sede del municipio, più a monte... *et voilà*, il comune – esattamente lo stesso di prima – diventa "montano" (e ottiene i finanziamenti).

In ambito scientifico e biomedico in particolare, a rischio, diciamo così, sono le statistiche che valutano la *credibilità percepita* ("impact factor") delle riviste scientifiche; si veda per esempio *A user's guide*

to inflated and manipulated impact factors, di John P. A. Ioannidis e Brett D. Thombs, European Journal of Clinical Investigation (17 giugno 2019) [Link->](#)

21.19 Traslazioni e confusioni linguistiche

Spesso le statistiche vengono presentate con quelle che potremmo considerare “traslazioni di linguaggio”.

Per necessità di sintesi, soprattutto nei titoli – a non voler pensare male.

I cani non sono la terza causa di morte nel mondo, no. Sono, più precisamente, la terza causa di morte *fra quelle causate da animali*, dopo zanzare e serpenti.

Titolo (vero, al momento ancora reperibile in rete, ottobre 2020):
Piene al 150% le rianimazioni in Abruzzo,
prima Regione in Italia ad esaurire i posti
letto

ATTUALITÀ / REGIONE ABRUZZO Venerdì 16 Ottobre 2020

Più discretamente la prestigiosa agenzia di stampa ANSA, nel testo di un articolo, non nel titolo, e con qualche parola in più:

Secondo il report settimanale dell’Alta Scuola di Economia e Management dei Sistemi Sanitari dell’Università Cattolica, campus di Roma (Altems) l’Abruzzo ha saturato il 150% dei posti letto aggiuntivi implementati

Dunque, vediamo che hanno saturato dei posti letto *aggiuntivi*, non quelli ordinari che avevano già prima della crisi.

Un’attenta analisi delle fonti che nessun normale lettore farà rivela che si tratta di 10 posti in terapia intensiva – non sappiamo quanti eventualmente per covid, di cui parlano gli articoli considerati, o incidenti stradali o altro – su 7 di certi posti definiti “aggiuntivi”,

aggiunti per la crisi covid, mentre *molti più di 100 restavano non occupati*.

È stato tremendo, ho fatto 3 ore di straordinario
(Nelle 8 ore lavorative ordinarie riposavo)

Naturalmente questi titoli e queste notizie, oltre all'incalcolabile danno di ansia e paura che possono procurare alle persone fragili – in un determinato tempo e luogo nell'epidemia covid del 2020 si è constatato un *triplicamento* degli infarti, forse almeno in parte dovuti alla paura di morire – manipolano il consenso popolare verso una o l'altra azione politica, che decreterà questo o quell'altro *spostamento di risorse* o, addirittura, determinate limitazioni alla vita delle persone.

21.20 Il problema della discrezionalità

Nella rilevazione dei dati può essere presente un ampio margine di discrezionalità: classificare al microscopio cellule in "regolari/ir-regolari", "simmetriche/asimmetriche"...

Esempio lampante di discrezionalità:⁽¹²⁹⁾

"Moribund larvae were considered dead and included in the analyses."

21.21 Esempio importante: le cause di morte

Il referto sulle cause di morte presenta una certa discrezionalità.

Le persone in generale giungono alla morte con diverse patologie concomitanti (si veda *comorbidità*).

E alcune muoiono pure dissanguate in incidenti automobilistici, in realtà causati dalle loro patologie (non è il dissanguamento la "vera" causa, a monte). Si pensi per esempio all'incidente automobilistico

¹²⁹Nartey, Rita et al. "Use of *Bacillus thuringiensis* var *israelensis* as a viable option in an Integrated Malaria Vector Control Programme in the Kumasi Metropolis, Ghana." *Parasites & vectors* vol. 6 116. 22 Apr. 2013, doi:10.1186/1756-3305-6-116

da crisi ipoglicemica.

Freddie Mercury è morto di polmonite (causa formale) o di AIDS (causa sostanziale)?

Anche alcol e droga forse non gli hanno fatto molto bene. Ma la causa di morte non viene ripartita in percentuali.

Il problema diventa critico per concomitanti patologie entrambe registrabili come causa di morte. Se sul piano di verità entrambe hanno favorito la morte, nelle statistiche ufficiali una sola l'ha causata, e quelle statistiche hanno conseguenze sulle politiche sanitarie e farmaceutiche – magari con la mediazione dei media.

Si pensi all'enorme difficoltà che può esserci dietro questa scelta:

morto per leucemia ma con morbillo
morto con leucemia ma per morbillo.

La decisione di cosa scrivere spetta a persone che sono generalmente molto competenti, tuttavia la situazione talvolta può essere complessa.

A proposito del Coronavirus un virologo avrebbe (2020) affermato:

Tutte e 7 le vittime avevano anche altre patologie, quindi sarebbe più opportuno parlare di Coronavirus come concausa non come causa diretta.⁽¹³⁰⁾

Molto problematico da un punto di vista statistico questo passaggio di Repubblica:

Si tratta in gran parte inoltre - viene precisato - di pazienti contagiati dal virus nei mesi scorsi, nel frattempo negativizzatisi, ma che su indicazione del Ministero della Sanità vanno registrati comunque come soggetti con infezione da Covid.

Per capire meglio la complessità della questione, su un sito internet del Ministero della Salute leggiamo⁽¹³¹⁾

¹³⁰Matteo Bassetti, direttore della clinica di malattie infettive dell'ospedale policlinico San Martino di Genova, citato in <https://www.primocanale.it/notizie/vittime-coronavirus-1-esperto-con-1-influenza-probabilmente-sarebbero-morti-lo-stesso--216491.html>, letto il 28 febbraio 2020

¹³¹In <https://www.epicentro.iss.it/influenza/sorveglianza-mortalita-influenza>, accesso il 9 marzo 2020.

il virus influenzale non viene identificato o perché **non ricercato** o perché il decesso viene attribuito a polmoniti generiche.

(Enfasi aggiunta).

L'attribuzione di una causa di morte piuttosto che un'altra può avere grandi ripercussioni nelle decisioni di politica sanitaria, sostenuta dalle istanze dei cittadini, mediate o indirizzate dai media.

*

Non per mettere in dubbio nulla, però leggiamo, a proposito di un notissimo caso di cronaca di Trieste, di come possano cambiare i risultati da un'autopsia all'altra. Inizialmente, per la vittima, trovata in un bosco racchiusa in vari⁽¹³²⁾ sacchi, si era ipotizzato il suicidio.

Leggiamo⁽¹³³⁾

nuova autopsia, eseguita dopo la riesumazione del corpo, che indicano la presenza di lesioni compatibili con un'aggressione da parte di terzi:

- Frattura alla lamina della seconda vertebra toracica, risalente a poco prima della morte.
- Lesioni al volto e un ematoma sulla mano destra.

Questi elementi contrastano con le conclusioni della prima perizia medico-legale, che aveva escluso l'intervento di terze persone.

Leggiamo⁽¹³⁴⁾ altrove:

L'avvocato (2024) Gentile: "Quadro incompatibile con un suicidio".⁽¹³⁵⁾

21.22 Cose vere in articolo titolato oppostamente

In un articolo divulgativo, per esempio su un quotidiano, si possono riportare correttamente i risultati di uno studio scientifico, così nessuno può obiettare, mettendo però un titolo che dice l'opposto, così nessuno può obiettare comunque, perché il titolo l'ha messo il titolista, mica l'autore che firma l'articolo – e senz'altro gliela si

¹³²“si rafforza la tesi del suicidio (...) Dna di Liliana dappertutto. Sui sacchi neri in cui era infilato il corpo, sui sacchetti di nylon che le avvolgevano la testa” in <https://necrologie.ilpiccolo.gelocal.it/news/137558>

¹³³<https://www.triestecafe.it/it/news/cronaca/caso-resinovich-a-porta-a-porta-l-analisi-di-vladimir-luxuria-sul-mistero-della-sua-morte.html>

¹³⁴<https://www.fanpage.it/attualita/liliana-resinovich-cosa-e-emerso-dalla-seconda-autopsia-frattura-a-una-vertebra-e-altre-lesioni/>

¹³⁵Precedenti (2022) osservazioni critiche del medesimo avvocato in <https://www.triestepima.it/cronaca/liliana-resinovich-associazione-penelope.html>

perdona. Ovviamente il titolo farà immensamente più effetto del corpo dell'articolo.

Covid, effetto boomerang; a ogni reinfezione conseguenze più gravi

di Fabio Di Todaro



Il lavoro di un gruppo di ricercatori dell'Università di Washington pubblicato s possibili conseguenze di infezioni plurime da coronavirus: ad ogni nuovo cont incontro a complicanze

11 NOVEMBRE 2022 AGGIORNATO ALLE 09:44

f In assoluto determinano quadri meno gravi rispetto a quelli relativi al primo contagio,
t ma le reinfezioni da Sars-CoV-2 non è detto che evolvano sempre senza particolari
... conseguenze. Ragion per cui continuare ad adottare le misure di protezione non

Figure 30: Fotografia di la Repubblica online, 11 novembre 2022. Si legga la prima riga dell'articolo.

PARTE III – DATI FALSI

21.23 La falsificazione dei dati

Finora abbiamo considerato solo dati *veri*, magari presentati con omissioni o in modo fuorviante ma comunque veri: nella realtà pratica, esiste anche la falsificazione dei dati.

Falsificazioni negli articoli divulgativi

Per dare solo una vaga idea del livello stratosferico di menzogna continua – costantemente monitorata dallo scrivente – della generalità dei media italiani (l'Italia è internazionalmente classificata nel 2022 al [58esimo posto nel mondo per libertà di stampa](#)) durante la pandemia del covid-19, si consideri il caso della Svezia, che

ha tenuto una politica anticovid diversissima dall'Italia, e in particolare non ha mai (2020 - OTT 2022) chiuso le scuole: la Svezia veniva presentata come un luogo in cui il covid faceva strage, molto più che in Italia; il lettore interessato troverà in rete che la politica anticovid della Svezia veniva addirittura assimilata a un sacrificio umano. La verità fattuale è opposta:

– la Svezia ha costantemente avuto una mortalità covid molto più bassa dell'Italia;

– questo fatto non può essere liquidato dicendo che là non hanno contato bene i morti per covid, perchè anzi, la Svezia ha avuto un eccesso di mortalità *per tutte le cause* rispetto agli anni precedenti minore dell'Italia.

Cumulative confirmed COVID-19 deaths per million people

Due to varying protocols and challenges in the attribution of the cause of death, the number of confirmed deaths may not accurately represent the true number of deaths caused by COVID-19.

Our World
in Data

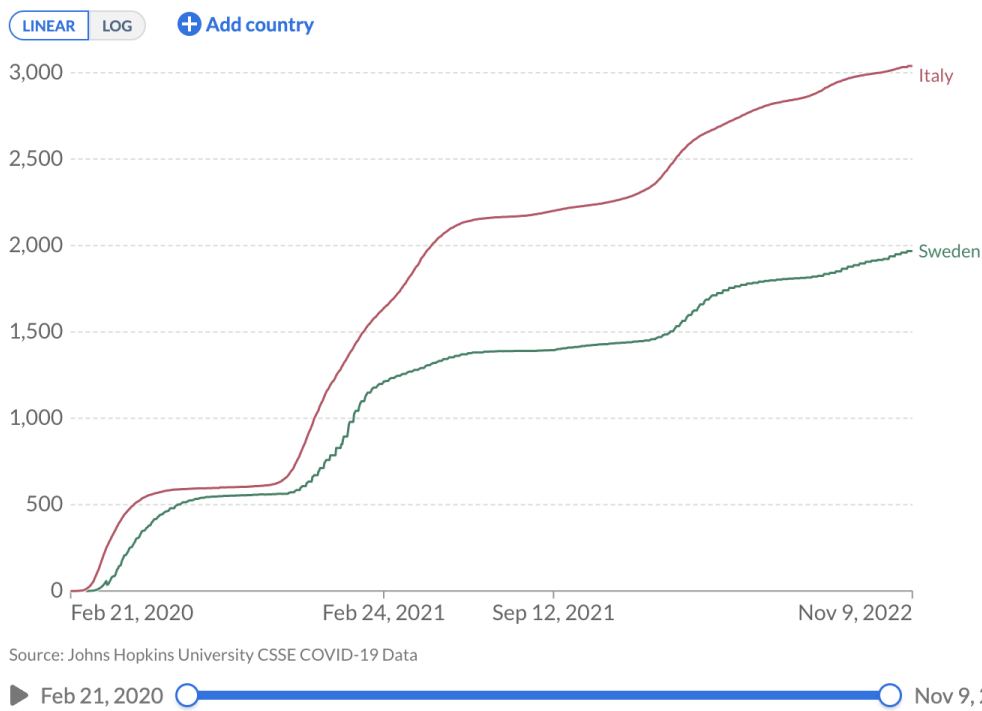


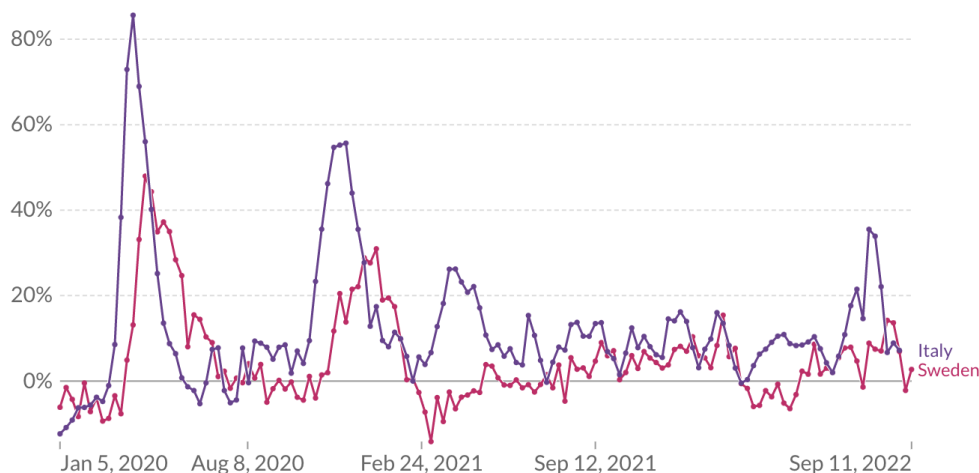
Figure 31: <https://ourworldindata.org/covid-deaths?country=ITA~SWE>, screenshot.

Excess mortality: Deaths from all causes compared to projection

The percentage difference between the reported number of weekly or monthly deaths in 2020–2022 and the projected number of deaths for the same period based on previous years. The reported number might not count all deaths that occurred due to incomplete coverage and delays in reporting.



+ Add country



Source: Human Mortality Database (2022), World Mortality Dataset (2022)

OurWorldInData.org/coronavirus • CC BY

Note: Comparisons across countries are affected by differences in the completeness of death reporting. Details can be found at our Excess Mortality page.

► Jan 5, 2020 ○ ○ Sep 11, 2022

Figure 32: <https://ourworldindata.org/excess-mortality-covid>, screenshot.

E negli articoli scientifici è legale mentire?

Leggiamo⁽¹³⁶⁾

Giugno 2019 [...] Scienziati impegnati nella ricerca sul cancro hanno manipolato le immagini dei loro studi, riuscendo così a ottenere successo, carriera, nuovi fondi per le loro ricerche. La Procura (...) ha appena concluso un'indagine lunga e complessa, che fornisce un quadro devastante

e volendo troviamo⁽¹³⁷⁾ molti dettagli fra cui

¹³⁶In <https://www.ilfattoquotidiano.it/in-edicola/articoli/2019/06/30/ricerca-sul-cancro-risultati-ritoccati-per-ottenere-milioni/5291204/>

¹³⁷<https://codacons.it/ricerca-sul-cancro-risultati-ritoccati-per-ottenere-milioni/>

la procura deve archiviare (...) non è reato

Dal punto di vista giuridico, le cose stanno nei termini seguenti.

Le *bugie* possono essere eventualmente immorali, ma non sono illegali, tranne in alcune fattispecie elencate dalla legge: il falso in atto pubblico, il falso in bilancio, il falso ideologico, eccetera, ma assolutamente non è compreso il falso in articolo scientifico, che quindi è assimilabile a una bugia fra privati cittadini – anche se l'articolo contiene descrizioni di esperimenti mai avvenuti, cioè non semplici *errori* ma falsità scritte con dolo.

Naturalmente, un cittadino che ritenesse che dalla pubblicazione di affermazioni false in un articolo scientifico gliene sia conseguito un danno economico, o un danno biologico, potrebbe ipotizzare una querela, o una denuncia – ma si capisce bene che in generale sarà impresa ardua, specialmente per articoli scientifici che non dicono bene di farmaci specifici, ma facciano *Scienza di Base*.

Tuttavia, l'autore truffaldino rischia pur sempre qualcosa.

Se egli è, come avviene quasi sempre, dipendente di un'istituzione scientifica (per esempio un'università), o anche solo affiliato scientificamente ad essa, è ben probabile che la pubblicazione di dati deliberatamente falsificati sia contraria al codice deontologico dell'istituzione. Allora al falsario l'istituzione può comminare una sanzione disciplinare in base al proprio regolamento, per violazione del codice deontologico (comunque solo in base a regolamenti e non a leggi, e quindi ad un livello molto più basso: le pene detentive sono escluse per la violazione di regolamenti).

Inoltre se la falsificazione diviene di dominio pubblico, l'istituzione può rivalersi contro il truffaldino per avere il risarcimento per danno di immagine - con una denuncia o querela (e quindi al livello di legge).

Si noti comunque che mentre la Giustizia tende a perseguire le violazioni delle leggi (cioè i reati), l'istituzione scientifica avrebbe in generale tutto l'interesse a occultare o almeno a minimizzare la violazione del regolamento (avvenuta con la falsificazione dei dati) per evitare un danno d'immagine, e ciò fa intuire come verosimilmente andrebbero le cose.

Esercizio μ Si afferma che un intervento del 1845 ha causato una rapida diminuzione del parametro X, e si presenta questo diagramma a colonne:

```

1845 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1846 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1847 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1848 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1849 XXXXXXXXXXXXXXXXXXXXXXXX
1850 XXXXXXXXXXXXXXXXXXXXX

```

È plausibile quell'affermazione? Ecco un dataset più completo:

```

1842 7.68
1843 6.14
1844 4.92
1845 3.93
1846 3.19
1847 2.58
1848 2.09
1849 1.69
1850 1.37

```

(Risposta: no, non è plausibile. Infatti fino al 1845 il parametro diminuiva circa del 20% annuo e poi circa del 19%).

Sezione A2 – Calcolo Infinitesimale

BOZZA - DRAFT