

La statistica è una scienza esatta. Se appare a volte poco

Bugie, statistiche



Viaggiare sicuri... ma come?

Treni e aerei si contendono il primato di "mezzo più sicuro", sfoderando statistiche opposte. L'aereo ha solo 3 morti ogni 10 miliardi di passeggeri/chilometro (il treno 9). Ma in treno perdono la vita

solo 7 persone ogni 100 milioni di passeggeri/ora (in aereo 24). Chi ha ragione? Nessuno. Perché in questo caso la definizione di sicurezza non è chiara (statistica non applicabile).



Cicogne e bambini

A Copenaghen, negli anni '50, si registrò un aumento di cicogne e un contemporaneo aumento di nascite. Se qualcuno avesse messo in relazione i due fatti

avrebbe commesso un errore (molto frequente) chiamato "errata correlazione". Si presenta quando 2 fenomeni indipendenti (che per qualche

motivo hanno andamenti simili) vengono presentati come dipendenti l'uno dall'altro. L'errore è unire 2 statistiche indipendenti.

Inarrestabile inflazione

Nel gennaio 2001, un giornale titolò: "Negli ultimi 17 mesi l'inflazione in Europa è cresciuta del 164%", forse per alimentare nei propri lettori l'ostilità all'Unione monetaria europea. Il dato era corretto, ma si riferiva a una percentuale di una percentuale (l'inflazione era infatti passata dall'1,1 al 2,9%). Ma i lettori meno attenti avrebbero potuto capire che il tasso d'inflazione era giunto al 164%.



credibile è solo perché viene usata male (o in malafede)

e mezze verità

In quanti modi una statistica può mentire? Ne abbiamo contati almeno 10. Dovuti a errori, fretta, ignoranza, inganno, vantaggio personale, superficialità...

L'inflazione aumenta! «No, diminuisce». I prezzi sono cresciuti del 7%. «Ma nell'ultimo semestre sono calati del 2%». Statistiche, percentuali, ricerche di mercato a volte sembrano proprio dare i numeri. In realtà la statistica è una scienza esatta... o, meglio, lo sarebbe se i dati fossero raccolti

bene, i calcoli corretti e i risultati presentati in modo completo. Purtroppo, non sempre è così: ecco come nascono gli errori e (e, a volte, gli inganni) della statistica.

● Grande è meglio

Si parte con la raccolta dei dati. Facile, in apparenza; basta distribuire questionari o utilizzare rile-

vazioni demoscopiche. Ma se il campione non è ben scelto, può accadere ciò che si verificò con il rapporto Kinsey (un'indagine fatta negli anni '50 sul comportamento sessuale negli Usa): domande mal poste e un campione troppo piccolo portarono alla conclusione che il 10% degli americani era o era stato omosessuale (il

dato corretto è dal 2% al 3%). Anche indagini mediche fatte su campioni piccoli (v. riquadro nella prossima pagina) possono dare risultati poco affidabili.

● "Aggiustare" i numeri

Ci sono poi anche statistiche viziate all'origine. Condotte soltanto nelle grandi città, per esempio, ►

Crescita inarrestabile

Una statistica può prevedere il futuro... ma lo farà con esattezza solo se il fenomeno ha un andamento costante. Altrimenti, andrà incontro a errori dovuti a errata estrapolazione. Come è accaduto agli studiosi del Club di Roma che negli anni '60 prevedevano l'esaurimento di molte risorse naturali entro il 2000... se i parametri

considerati fossero rimasti costanti (ma non lo sono rimasti). Gli errori di estrapolazione appaiono assurdi quando consideriamo fenomeni ben noti (come l'aumento di statura di un uomo, nel disegno) ma sono sempre in agguato quando si parla di futuro. Per questo, si fanno 3 stime diverse di quanti uomini vivranno sulla Terra nel 2050.

42 anni

Altezza presumibile per estrapolazione: 300 cm

La statura di un essere umano non cresce in modo così regolare come mostrano queste foto, ma fino ai 18 anni aumenta. Poi, però, si ferma. Altrimenti, a 40 anni...

18 anni

Altezza di circa 180 cm

12 anni

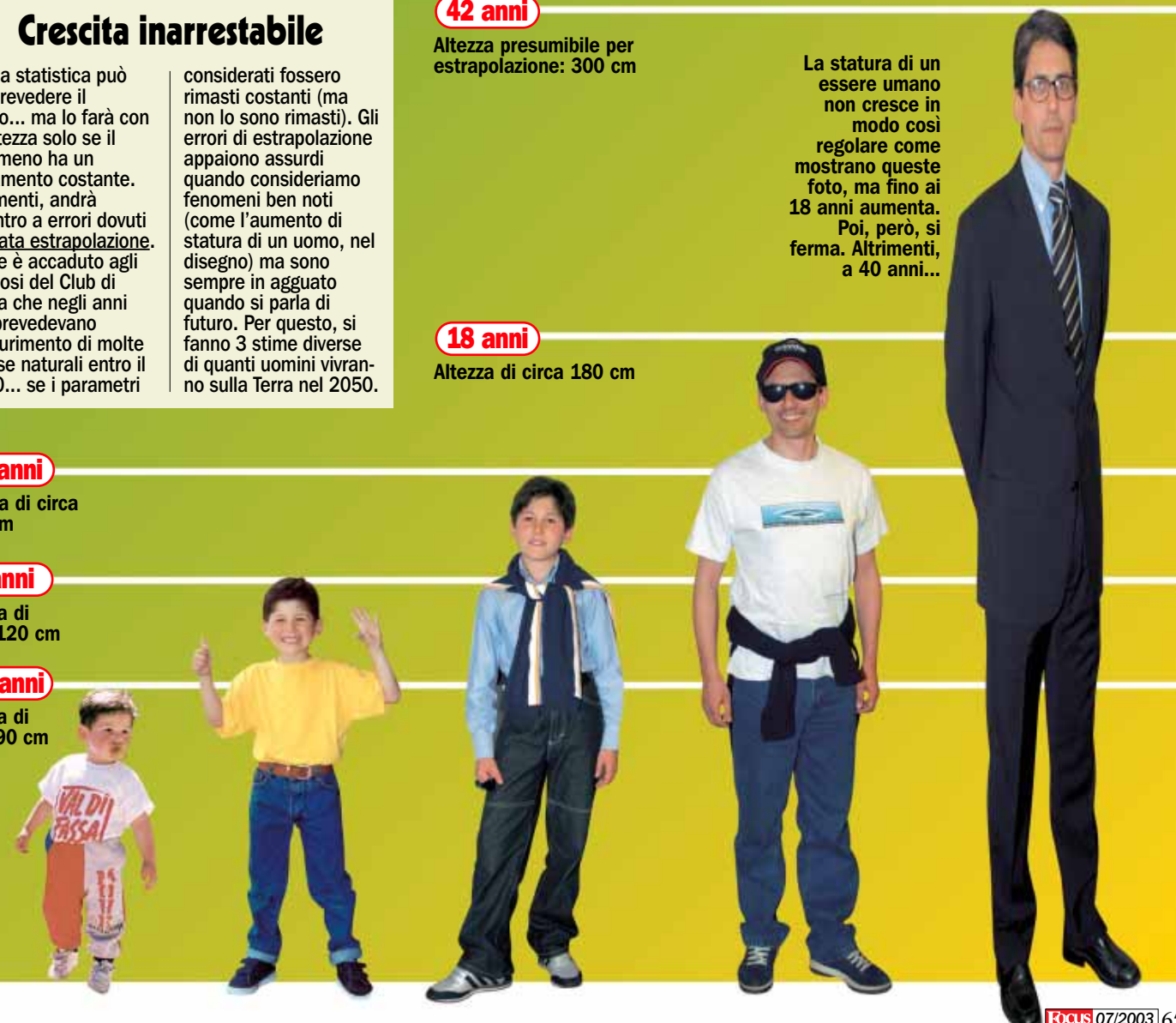
Altezza di circa 150 cm

6 anni

Altezza di circa 120 cm

1,5 anni

Altezza di circa 90 cm



Attenti alle variabili nascoste: possono ribaltare il significato dei dati statistici

► benché il 51 per cento degli italiani viva in paesi con meno di 10 mila abitanti. In questo caso gli statistici parlano di "campione non significativo".

Oppure i dati sono perfetti... ma non ci vengono comunicati tutti (v. riquadro sotto). Si parla allora di variabili nascoste, che possono far cambiare completamente il significato di uno studio. Oppure si possono omettere alcuni dati, senza falsare gli altri. «Può accadere per i risultati particolarmente sgraditi all'azienda committente» spiega Enrico Finzi, di Astra-Demoskopea. «Noi in questo caso preferiremmo non pubblicare i risultati».

● Crescita limitata

Se i dati sono giusti e completi, bisogna poi interpretarli. Con vari rischi di errore. Il primo è quello dell'errata estrapolazione. Un caso concreto? Se un determinato investimento, un mix di azioni o di obbligazioni è cresciuto dell'8% in un anno, non è detto che continui a farlo per tutto il periodo di investimento. Così come, se un bambino a due anni è alto 90 centimetri, a 6 anni 120 cm, a 12 anni 150 cm, non è detto che a 18 anni sia alto 180 cm... né, soprattutto, che a 40 anni l'altezza possa arrivare a 3 metri. Sembra ovvio, eppure le er-

rate estrapolazioni sono frequenti. Una delle più clamorose la fece il Club di Roma nel suo rapporto *I limiti dello sviluppo*, che negli anni '70 prospettava l'esaurimento di parecchie risorse naturali entro la fine del secolo (esaurimento che non c'è mai stato). Le ipotesi su cui si basavano gli studiosi erano molto pessimistiche, e questo toglieva ai loro calcoli ogni valore (a parte quello di ammonirci che le risorse non sono inesauribili, un fatto di cui non dovremmo mai dimenticarci).

● Fiducia

Ma l'errore più grave (e più frequente nei giornali) è quello che si può definire errata correlazione. Se due fatti accadono in concomitanza, non è detto che il primo sia causa del secondo (v. riquadro *sulle cicogne nelle pagine precedenti*). Un altro esempio è dato da una recentissima ricerca della World Values Survey sul livello di fiducia nei confronti degli stranieri, nei va-



Allarme ingiustificato?

Esiste una relazione tra elettrosmog e leucemia infantile? Forse sì, forse no: le ricerche finora fatte non sono conclusive. Lo sostengono gli autori di uno studio fatto a Roma, nella zona intorno alle antenne di Radio Vaticana: pur avendo rilevato nel periodo 1987-1998 un aumento di leucemie

rispetto alla media, hanno scritto che «il numero di eventi osservati è piccolo, quindi le statistiche sono affette da grande instabilità». Un esempio: nel "primo cerchio" intorno alle antenne (2 km di raggio) c'è 1 caso di leucemia infantile, quando la probabilità statistica è circa 0,16. "Sei volte più

della media!" hanno titolato i giornali, senza badare al fatto che non era certo possibile che si ammalassero frazioni di persone. In realtà, i dati indicano solo che sono necessarie altre ricerche.



Analfabeta sarà lei!

Per l'Unesco, l'analfabetismo in Germania è cresciuto in modo consistente dall'inizio del secolo. Questa affermazione paradossale non è

dovuta alla statistica, ma, ancora una volta, a chi la applica male. Oggi, infatti si definisce analfabeta chi non sa "leggere, scrivere e far di conto in modo da

poter migliorare se stesso e la società"... ieri lo era chi non sapeva scrivere il proprio nome. Insomma, i dati non sono comparabili.

Maledette variabili

Ecco un esempio di variabili nascoste: in un lago ci sono ottimi lucci e spinosissimi pesci-gatto. Il pescatore compra un'esca che i lucci mangiano al 94% e i pesci-gatto solo al 6%... eppure pesca solo questi ultimi.

Pochi pesci. Come mai? La sua pesca è stata rovinata da una delle tante possibili variabili nascoste: per l'esattezza, la presenza quasi nulla di lucci nella zona del lago in cui ha fermato la sua barca.



Test e rischi

Se vi dicessero che siete positivi a un test per il tumore all'intestino ne sareste preoccupati. Però questi test danno un risultato falsamente positivo (cioè si è sani) nel 3,2% dei casi. E la percentuale dei malati senza sintomi (oltre i 75 anni) è dello 0,2%. Quindi, su 100 mila persone esaminate, solo 200 (lo 0,2%, *in rosso nel disegno*) sono davvero malate, ma ben 3.200 (il 3,2%, *in rosa nel disegno*) sono sane e positive al test: tutta colpa delle "probabilità condizionate".



Ogni anno muoiono negli Usa miliardi di bimbi... Possibile?

► ri Paesi del mondo. La ricerca ha messo in luce che le nazioni che meno si fidano degli altri sono quelle più inquinate, mentre un'intensa attività sessuale comporterebbe un alto livello di fiducia. Possibile? Ovviamente no: forse il risultato dipende dal fatto che Paesi molto liberali (anche in fatto di costumi sessuali) come quelli scandinavi sono anche attenti alle leggi contro l'inquinamento. Certamente però non può esistere un rapporto di causa-effetto tra ambiente/sexo e fiducia o, se ci fosse, solo altri tipi di ricerca potrebbero metterlo in luce. Allo stesso modo, se tra le auto incidentate prevalesse il colore bianco, non sarebbe legittimo dire che il bianco aumenta il rischio di incidente. Anche se, purtroppo, c'è chi lo dice.

● Analfabeti e no

Le cose si complicano ancora di più quando non abbiamo di fronte una statistica, ma due. Il confronto fra due gruppi di dati deve infatti partire da basi simili. Ecco un caso concreto: all'inizio del secolo scorso c'erano in Germania 10 mila persone che non sapevano né leggere né scrivere. Oggi, secondo l'Unesco, gli analfabeti in Germania sono 4 milioni. Un'assurdità che nasce dal confronto tra dati non omogenei. Un secolo fa, infatti, era analfabeta chi non sapeva scrivere il proprio nome; oggi è analfabeta "una persona che non può prendere parte a tutte le atti-

vità con obiettivi definiti del suo gruppo e della sua comunità, per le quali è necessario saper leggere, scrivere e far di conto, e allo sfruttamento di questi strumenti culturali per il proprio sviluppo e per quello della società".

● Miliardi di bambini uccisi

Un'ultima serie di errori, molto più difficile da scoprire, può derivare da semplici incomprensioni o letture sbagliate dei risultati. Come nel caso di alcuni test medici (v. riquadro qui sopra) per malattie poco diffuse: si può risultare malati ma, per assurdo, avere una grande probabilità di essere sani, perché il

significato delle percentuali di errore non è facile da capire.

Infine si può proprio sbagliare a dare un risultato, come racconta l'americano Joel Best, nel libro *Lies, damned lies and statistics* (Bugie, bugie sfacciate e statistiche). Qualche tempo fa un giornale scrisse: «Dal 1950 il numero di bambini americani uccisi con armi da fuoco è raddoppiato ogni anno». Partendo da 1 bambino nel 1950, si arriverebbe nel 2003 a 4.503.599.627.370.496 bambini uccisi... La frase andava letta invece così: «Dal 1950 il numero di bambini americani uccisi ogni anno con armi da fuoco è raddoppiato». Piccolo spostamento nella frase, enorme errore nel risultato. ■

Marco Ferrari

L'ingannevole media

Pochi mesi fa hanno suscitato incredulità i dati sugli stipendi medi degli insegnanti: troppo alti! Nei dati ministeriali

erano però compresi gli stipendi dei docenti universitari, che alzavano la media. Nella percezione comune, però, si identifica lo stipendio medio con il più frequente, confondendo la

media con la moda. Se 500 persone hanno stipendio 1.000 e 30 stipendio 10 mila, la media sarà 1.509: molto più della moda (che è il dato più frequente, cioè 1.000).



Focus

© Gruner und Jahr - Mondadori SpA
Tutti i diritti di proprietà letteraria e artistica riservati.



Gruner und Jahr-Mondadori SpA

Gruner und Jahr-Mondadori SpA
Corso Monforte, 54 - 20122 Milano

Elaborazione **ELEVER SRL**