La Bella Addormentata
e altre illusioni probabilistiche

Aljoša Volčič

volcic@unical.it

Firenze, 25 novembre 2009

Che cosa è la probabilità?

"La probabilità di un evento A è la misura del grado di fiducia che un individuo coerente attribuisce, secondo le sue informazioni e opinioni, all'avverarsi di A".

Questa è la definizione che ha dato della probabilità un grande genio, il matematico italiano Bruno De Finetti (1906-1985).

Ci sono naturalmente diverse cose da spiegare.

a) Che cosa è un evento? Nel calcolo delle probabilità, gli eventi sono gli esiti di vari esperimenti casuali (aleatori): il risultato del lancio di una moneta, o di un dado, della vittoria dell'Inter in questo campionato, la carta che si sceglie da un mazzo, il numero della tombola estratto dalla nonna, il fatto che quel numero estratto sia presente sulla mia cartella, la mia sopravvivenza per altri 20 anni, l'ammontare del danno al carico trasportato da un mercantile, etc.

b) Che cosa vuol dire che un individuo è coerente? Vuol dire che è disposto a scommettere, con la stessa quotazione, su A ma anche sul suo contrario (o su combinazioni di eventi che coinvolgano A). Si presume che egli non segua un impulso irrazionale (del tipo: tifoso sfegatato della Fiorentina è disposto a scommettere alla pari che il campionato lo vincerà la sua squadra).

c) Come si "misura [il] grado di fiducia"? Con un numero, naturalmente, come tutte le cose misurabili di questo mondo. In matematica si preferisce misurare la probabilità con un numero compreso tra 0 ed 1 (anche se nel linguaggio comune si usano spesso le percentuali).

Ma quale numero?

Aiutiamoci con un esempio.

Supponiamo di avere un'urna contenente 20 palline, di cui 3 rosse e 17 bianche, indistinguibili al tatto. Dopo averle mescolate bene, facciamo estrarre una pallina ad un fanciullo innocente e bendato, affiancato da due carabinieri in alta uniforme.

i) Ci chiediamo punto qual è la probabilità dell'evento A "la pallina estratta è rossa"? Ebbene, questo è uno di quei casi semplici (detti "classici") in cui anche per una persona non esperta è facile dare la risposta giusta: la probabilità è data da

$$P(A) = \frac{3}{20}\,.$$

ii) Vediamo ora lo stesso problema da un altro punto di vista: supponiamo di ripetere più volte l'estrazione, diciamo 1000 volte, rimettendo ogni volta la pallina estratta nell'urna e prendendo nota della frequenza (assoluta e relativa) dell'evento A. Che cosa ci aspettiamo?

Anche per rispondere a questa domanda non occorre essere grandi esperti di probabilità: è intuitivo che la pallina rossa verrà estratta *circa* 150 volte. Quello che è meno intuitivo è la quantificazione di questo *circa*. Metodi un po' più avanzati permettono di dire che la frequenza osservata sarà, con probabilità appena superiore al 50%, tra 140 e 160. Dunque uno scostamento dal risultato più probabile (che è 150) superiore al 6 o 7% non ci deve affatto stupire.

Denotiamo con f la *frequenza relativa* di A. Quanto appena detto equivale ad affermare che sarà, con probabilità appena superiore al 50%,

$$\frac{2,8}{20} \leq f \leq \frac{3,2}{20}\,.$$

Se aumentiamo il numero delle estrazioni, la frequenza relativa sarà sempre più vicina a $P(A) = \frac{3}{20}$ e questo segue dalla famosa (e spesso mal interpretata) legge dei grandi numeri.

iii) Vediamo ora un terzo aspetto. Supponiamo di voler organizzare un gioco di scommesse sull'esito dell'estrazione dalla solita urna. Lo vogliamo fare senza guadagnarci (né ci vogliamo perdere).

Qual è la posta (scommessa) S "giusta" che il giocatore deve pagare se il monte premi è pari ad un euro?

Qui entra in campo un concetto fondamentale utilizzato da De Finetti, quello di **gioco equo**, cioè un gioco in cui *in media* nessuno vince e nessuno perde.

La vincita media (che si chiama anche *speranza matematica*) del nostro gioco, in cui S è ancora da determinare, si calcola nel modo seguente:

$$P(A) \cdot (1 - S) - (1 - P(A)) \cdot S$$

Per quanto visto prima (al punto ii)), $P(A)$ dà una valutazione della frequenza relativa dell'evento A. Quando vince, il giocatore riceve 1, ma il banco naturalmente si tiene la puntata, quindi la vincita effettiva è $1 - S$.

Il giocatore perde invece con frequenza relativa stimata da $1 - P(A)$, e quando questo succede, perde la posta S.

Se vogliamo che il gioco sia equo, deve essere

$$P(A) \cdot (1 - S) - (1 - P(A)) \cdot S = 0\,,$$

ovvero

$$P(A) = S\,.$$

Questa equazione esprime in termini quantitativi la frase di De Finetti che conviene ora rileggere:

"La probabilità di un evento A è la misura del grado di fiducia che un individuo coerente attribuisce, secondo le sue informazioni e opinioni, all'avverarsi di A".

Consideriamo un generico gioco G, nel quale il giocatore guadagna a se si verifica l'evento A e guadagna b se A non si verifica (sul *segno* di a e b non ci pronunciamo!). Allora la speranza matematica di G si definisce come

$$E(G) = ap + bq\,,$$

dove $p = P(A)$ e $q = 1 - P(A)$.

Il gioco si dice equo se $E(G) = 0$, sfavorevole (al giocatore) se $E(G) < 0$ e favorevole se $E(G) > 0$.

Naturalmente chi organizza i giochi, fa in modo che questi non siano equi.

Se giochiamo alla roulette puntando sul numero, il banco paga 36 volte la posta (che non viene restituita). Siccome la roulette ha 37 numeri (da 0 a 36), la probabilità di vincere è $\frac{1}{37}$.

La speranza matematica in questo caso è negativa. In media il giocatore perde e concede al banco un margine di guadagno medio del $2,7\%$ circa.

Nella roulette americana oltre allo 0 c'è anche il doppio 0 e il margine a favore dei casinò è del $5,3\%$ circa. Nei casinò della Cuba precastrista le roulette avevano anche il triplo zero e il margine per il banco era quasi del 7,7%!

Il paradosso della Bella Addormentata

M. Piccione e A. Rubinstein hanno pubblicato nel 1997 un articolo sulla rivista Games and Economic Bahavior nel quale hanno introdotto il cosiddetto "paradosso della Bella Addormentata" (*sleeping beauty paradox*). L'articolo ha sollevato discussioni, soprattutto tra gli economisti e gli psicologi, e sono stati successivamente creati altri problemi, ben più complessi, per evidenziare meglio certi aspetti che riguardano le decisioni prese in caso di memoria incompleta.

La sceneggiatura è la seguente: dei ricercatori sottopongono la Bella Addormentata ad un esperimento, spiegandole come intendono procedere.

Dopo averla addormentata, verrà lanciata una moneta equilibrata.

A seconda del risultato del lancio, essa verrà risvegliata una o due volte: una volta se il lancio della moneta darà testa, due volte se darà croce.

Ad ogni risveglio verrà chiesto alla Bella Addormentata quale probabilità lei assegni al fatto che la moneta ha dato testa.

Dopo ogni risveglio la Bella Addormentata viene riaddormentata con un narcotico che le fa dimenticare le fasi precedenti dell'esperimento.

Lei non sa quindi, nella fase di risveglio, se è già stata svegliata in precedenza o se quello è il suo primo (e magari unico) risveglio.

La Bella Addormentata conosce le regole dell'esperimento e le ricorda per tutta la sua durata; non conosce però, ovviamente, quale risultato abbia prodotto il lancio della moneta.

Viene posta a questo punto la seguente domanda: quando la Bella Adormentata viene risvegliata, qual è la probabilità che essa assegna al fatto che il lancio della moneta ha dato testa?

Gli autori hanno proposto due risposte diverse e gli "amatori" che si sono dedicati a questo problema sono in effetti divisi in due campi.

Ci sono quelli che vengono chiamati i "halfters", noi diremmo i "mezzisti", i quali ritengono che la Bella Addormentata dovrebbe rispondere $\frac{1}{2}$.

I mezzisti sostengono infatti che inizialmente la stima della Bella Addormentata, che conosce le regole dell'esperimento e che sa che la moneta non è truccata, è di $\frac{1}{2}$. Nel momento in cui viene risvegliata, non ha a disposizione alcun nuovo elemento, poiché non sa se il suo è il primo (e magari unico) risveglio, oppure se si tratta del secondo risveglio, né ha alcun mezzo per scoprirlo.

Il fatto che essa venga risvegliata non le dà dunque alcuna nuova informazione, pertanto i mezzisti ritengono che la sua stima debba rimanere quella “a priori”, cioè $\frac{1}{2}$.

Contro i mezzisti sono schierati i “thirders” (noi diremmo i “terzisti”), i quali ritengono che la risposta corretta della Bella Addormentata sia $\frac{1}{3}$. Dicono infatti i terzisti che se l'esperimento viene ripetuto più volte, un terzo dei risvegli è, in media, susseguente ad un lancio nel quale la moneta ha mostrato testa, mentre due terzi dei risvegli seguono un lancio in cui si è ottenuto croce. Se quindi lei risponde “croce”, sbaglierà una volta su tre, e indovinerà due volte su tre. Il fatto che l'esperimento venga eseguito una sola volta non cambia la valutazione delle probabilità.

Io mi schiero nettamente con i terzisti.

Per avvicinarci al punto di vista di De Finetti, immaginiamo che alle risposte sia legata una simbolica scommessa. La Bella Addormentata vince 1 euro per ogni risposta giusta e ne perde uno per ogni risposta sbagliata.

Nella seguente tabella riportiamo tutte le possibili situazioni.

Esito del lancio	risvegli	BA dice C	BA dice T
testa (T)	1	-1	+1
croce (C)	2	+2	-2

Nella prima colonna è indicato l'esito del lancio della moneta, che può dare con uguale probabilità testa (T) o croce (C). Nella seconda colonna sono riportati i conseguenti risvegli, uno nel caso di T e due nel caso di C. Nella terza e nella quarta colonna sono riportate le vincite e le perdite conseguenti alle due possibili risposte che la Bella Addormentata (BA) dà ad ogni risveglio.

La conclusione è a questo punto ovvia: se la Bella Addormentata punta tutte le volte su "croce", allora vince 2 euro con probabilità $\frac{1}{2}$ e perde 1 euro con la stessa probabilità. Se invece punta tutte le volte su "testa", vince 1 euro con probabilità $\frac{1}{2}$ e con la stessa probabilità perde 2 euro.

Se punta su "croce", la speranza matematica è

$$2 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = \frac{1}{2} > 0\,,$$

quindi il gioco è a lei favorevole.

Se i mezzisti avessero ragione, il gioco dovrebbe essere equo.

L'importante, secondo me, è capire che conta il numero dei risvegli, non la regolarità della moneta.

Non c'è dubbio infatti che, ritornata a casa dopo l'esperimento, se il Principe Azzurro le chiedesse quale probabilità lei attribuisca al risultato "testa", lei dovrebbe rispondere $\frac{1}{2}$. Ma questa risposta avviene al di fuori dello schema previsto dall'esperimento.

La seguente situazione mi sembra presenti delle analogie con il problema della Bella Addormentata.

Un giornalista economico tiene sul suo giornale una rubrica fissa, tre volte alla settimana: sabato, domenica e lunedì. Egli vuole favorire un'azienda quotata in borsa (di cui ha acquistato delle azioni, o che è proprietaria del giornale) e quindi addotta la seguente linea comportamentale: se le azioni dell'azienda aumentano nel corso della settimana, ne scrive sabato, per poi riprendere la notizia anche la domenica e il lunedì. Se invece le azioni subiscono un calo, ne scrive solo il sabato. Supponiamo che le azioni abbiano avuto nel corso dell'ultimo anno un andamento mediocre: qualche aumento e qualche perdita; un valore, alla fine dell'anno, non lontano da quello iniziale.

Un lettore distratto della rubrica giornalistica avrà l'impressione di un ottimo andamento generale delle azioni (= risvegli dopo "croce"), perché avrà letto nel corso dell'anno più commenti positivi che negativi.

Uno più attento invece capirà che c'è un trucco, anche indipendentemente dal valore finale delle azioni, perché si sarà reso conto che il giornalista tendeva a nascondere sotto il tappeto le prestazioni scadenti delle azioni, sottolineandone invece le prestazioni positive. Questo lettore cercherà quindi di liberarsi di quelle azioni e vedrà di trovare un giornale con una pagina economica più affidabile.

In internet si sono creati dei forum di discussione che raccolgono gli interventi dei vari "amatori" che desiderano esprimere la propria opinione. La discussione su questi blog non ha ancora raggiunto un consenso generale e i gruppi di discussione che se ne occupano continuano a pubblicare argomentazioni sia di mezzisti che di terzisti.

Gli interventi sono in stragrande maggioranza a favore della fazione terzista, ma naturalmente la prevalenza numerica non sarebbe di per sé significativa.

Viene in mente la barzelletta che racconta di un matematico che espone nel corso di una conferenza un suo nuovo risultato. Uno dei presenti si alza e dice che il risultato deve essere senz'altro falso, perché nella letteratura è ben noto un controesempio. Ma a questa obiezione il conferenziere risponde sorridente e sicuro di sé: "Ma io ho trovato ben tre dimostrazioni di questo teorema!".

Il paradosso delle due buste

In un gioco a premi simile a quelli che si vedono spesso in televisione ("Affari tuoi", ad esempio), ad un concorrente vengono presentate due buste chiuse. Nelle buste c'è l'indicazione del premio che il concorrente vincerà. Il concorrente sa che il premio più alto è esattamente il doppio del premio più basso, ma naturalmente non sa quale busta contenga il premio più alto, né conosce l'entità del premio.

Al concorrente viene data la possibilità di aprire una busta, senza impegno, e di vedere l'ammontare in euro (un numero intero positivo) che c'è scritto. A questo punto il conduttore dà la possibilità al concorrente di incassare il premio trovato, oppure di optare per il premio che è indicato nell'altra busta. Che cosa conviene al concorrente?

Noi sappiamo che cosa bisogna fare per fare una scelta razionale: calcolare la speranza matematica!

Sia n il premio indicato nella busta aperta. Se si tratta del premio minore, cambiando si guadagna la differenza, cioè $2n - n = n$. Se invece n è il premio maggiore, cambiando busta il concorrente perderà $n - \frac{n}{2} = \frac{n}{2}$.

La probabilità di aver scelto la busta con il premio maggiore è uguale a quella di aver scelto quella con il premio minore, perciò è naturale associare ad entrambi gli eventi la stessa probabilità.

La speranza matematica del gioco è dunque

$$E(G) = -\frac{n}{2} \cdot \frac{1}{2} + n\frac{1}{2} = \frac{n}{4} > 0 \,.$$

Pertanto sembrerebbe che al concorrente convenga sempre cambiare, seguendo il motto "l'altra busta è sempre più verde".

Il fatto paradossale è che questa convenienza c'è, qualunque sia la busta prescelta per prima e indipendentemente dalla cifra che vi è indicata!

La trappola in cui cadiamo leggendo il problema sta infatti nel considerare "naturale" che i due importi abbiano la stessa probabilità, anche dopo che che è stata letta la cifra scritta nella busta prescelta. Invece questa cifra può contenere informazioni interessanti, alle volte decisive.

Infatti, chi organizza il gioco ha dei vincoli di bilancio che dipendono dalla collocazione della trasmissione, dallo share, dalla politica adottata dalle reti concorrenti, dalle trasmissioni che si svolgono in contemporanea e alla fin fine dal bilancio generale della emittente.

Se il cartoncino nella busta che il giocatore ha aperto indica, ad esempio, 500.000 euro (che, per dire, è il premio massimo previsto dalla trasmissione “Affari tuoi”), il buon senso gli dovrebbe suggerire che l'altra busta non offrirà un milione di euro, ma molto più probabilmente 250.000 euro, e dunque deciderà di tenersi il mezzo milione.

In altre parole, assegnerà a $2n$ una probabilità molto piccola, prossima a zero.

Per contro, se la busta aperta offre un premio di 20.000 euro (o ancora più basso), basandosi anche sui premi elargiti nelle trasmissioni precedenti, il concorrente potrà ragionevolmente sperare che l'altra busta offra qualcosa di più, e quindi sarà più tentato di cambiare.

Detto in termini probabilistici, assegnerà a $2n$ una probabilità più prossima ad 1 che ad $\frac{1}{2}$.

Giovanni Prodi ha scritto una decina di anni fa un bellissimo articolo dedicato a questo paradosso, nel quale riporta le risposte dategli da alcuni dei più famosi probabilisti italiani e a queste aggiunge anche la sua. Le risposte sono tra loro molto diverse, anche se la radice del problema viene individuata da tutti nel fatto che non si può assegnare la stessa probabilità a tutti i numeri interi.

Il paradosso di Simpson

Questo "paradosso" è uno dei fenomeni più sorprendenti nel calcolo delle probabilità e nella statistica matematica. Esso si può manifestare quando si aggregano o disaggregano dati.

Nella letteratura il fenomeno prende il nome di "paradosso di Simpson", statistico inglese, il quale lo descrisse in un articolo del 1951. Ma quasi cinquant'anni prima di lui, nel 1903, ne scrisse George Udny Yule. Ne scrisse più tardi anche un altro grande statistico, Karl Pearson, ed altri ancora. Ecco un esempio di scoperta e riscoperta di un fenomeno, anche abbastanza importante, e della ingiusta attribuzione della priorità ad uno dei tanti che se ne occupò, non a chi per primo scoprì il fenomeno.

Incominciamo con un primo esempio.

Due ragazzi, uno di Roccacannuccia di Sopra e l'altro da Roccacannuccia di Sotto, devono decidere a quale campeggio estivo iscriversi.

I due hanno a disposizione due scelte: in entrambe le frazioni c'è il campeggio organizzato dalla parrocchia e quello laico, ma l'unico criterio che i ragazzi intendono seguire è quello di un rapporto più "favorevole" tra ragazze e ragazzi.

A Roccacannuccia di Sopra per il campeggio della parrocchia si sono iscritti 70 ragazze e 70 ragazzi, mentre per quello laico si sono iscritti 90 ragazze e 100 ragazzi. Il ragazzo opta dunque per il campeggio parrocchiale dove il rapporto è $1:1$, mentre nell'altro campeggio le ragazze sono in minoranza.

Nella frazione di Roccacannuccia di Sotto la situazione è simile: per il campeggio parrocchiale si sono iscritti 60 ragazze e 200 ragazzi, mentre per l'altro campeggio gli iscritti sono, rispettivamente, 20 ragazze e 80 ragazzi. Ovviamente anche l'altro amico opta per il campeggio parrocchiale, dove la percentuale delle ragazze è circa del 23%, mentre nell'altro la percentuale è del 20%.

Motivi organizzativi costringono ad un certo punto tanto le due parrocchie quanto le due associazioni laiche ad unire i gruppi. Intuitivamente sembrerebbe naturale ritenere che entrambi i ragazzi debbano confermare la propria scelta, ma ad uno dei due (che non ha debiti formativi in matematica) viene lo scrupolo e rifà i conti:

Nel campeggio parrocchiale "unificato" ci sono ora $70+60=130$ ragazze, contro i $70+200=270$ ragazzi,

In quello laico ci sono invece $90+20=110$ ragazze, contro i $100+80=180$ ragazzi.

Nel primo campeggio il rapporto tra ragazze e ragazzi è $\frac{130}{270} \simeq 0,48$, mentre in quello laico è $\frac{110}{180} \simeq 0,61$, e quindi il secondo è nettamente preferibile, stando al criterio di scelta che hanno i due amici.

Da non crederci! I due fanno e rifanno i conti, ma i numeri sono quelli. Non resta dunque che, squotendo la testa, optare per il campeggio laico!

L'esempio seguente è ispirato ad un famoso caso accademico (e giuridico).

Lo statuto dell'Università della Calabria (che è un'università residenziale) prevede che ogni corso di laurea stabilisca, anno per anno, il numero massimo degli iscritti al primo anno, tenendo conto del numero dei docenti, delle aule, dei laboratori, e così via.

La selezione dei candidati avviene solitamente in base ad una graduatoria per la quale conta solo il voto di maturità, ma supponiamo che la Facoltà di Scienze decida di far fare ai candidati un esame attitudinale e di formare la graduatoria esclusivamente in base ad esso. Per ogni corso di laurea viene dunque creata una commissione che procede all'esame degli aspiranti.

Presso la Facoltà di Scienze la graduatoria è effettivamente selettiva solamente per i due corsi di laurea più richiesti, quelli di Informatica e di Biologia, perché per tutti gli altri corsi di laurea le domande sono solitamente meno numerose dei posti disponibili.

Supponiamo che per Informatica i posti disponibili siano 100, mentre per Biologia siano 140. Per questi due corsi di laurea vengono presentate in tutto 400 domande, così suddivise, per corso di laurea e genere: a Informatica presentano domanda 150 ragazzi (M) e 50 ragazze (F), a Biologia 50 M e 150 F. In totale 200 M e 200 F, 200 a I e 200 a B.

Questi dati sono realistici, perché le ragazze preferiscono biologia, mentre i ragazzi preferiscono informatica. Inoltre biologia offre più posti di quelli di informatica.

Le due commissioni predispongono dei test che non privilegiano né i ragazzi, né le ragazze. A informatica viene accettata la metà delle domande e a seguito del test viene ammessa la metà dei ragazzi e la metà delle ragazze che hanno fatto domanda. Questo vuol dire che ad Informatica vengono ammessi 75 ragazzi e 25 ragazze.

A biologia viene accolto il 70% delle domande. Anche qui il test seleziona con equità e passa la selezione il 70% dei ragazzi ed il 70% delle ragazze. Vengono quindi ammessi 35 ragazzi e 105 ragazze.

Rivediamo i dati nella seguente tabella.

	I	B	
domande M	150	50	
domande F	50	150	
ammessi M	75$*$		
ammessi F	25$\sharp$		
ammessi M		35$*$	
ammessi F		105$\sharp$	
totale M			110$*$
totale F			130$\sharp$
totale			240

I dati vengono riportati un giornale locale, che constata che dei 240 ammessi ai due corsi di laurea 110 sono ragazzi e 130 sono ragazze, ed osserva che tra i candidati c'erano 200 ragazzi ed altrettante ragazze. Il titolo dell'articolo è "L'Università della Calabria privilegia le ragazze?".

Naturalmente noi abbiamo seguito passo passo come si è arrivati a quel dato globale e sappiamo che il titolo del giornale avanza un'ipotesi che non ha alcun fondamento. In questo, appunto, sta il paradosso.

Si noti che anche se i risultati dei test fossero stati leggermente più favorevoli ai ragazzi che alle ragazze, questo fatto sarebbe rimasto "mascherato" nell'aggregazione dei dati: se ad esempio ci fossero stati 78 ragazzi ammessi ad Informatica e 41 a biologia, le ragazze complessivamente ammesse sarebbero state comunque più numerose.

La spiegazione dell'inversione di tendenza nell'aggregazione dei dati sta nel fatto che i ragazzi hanno fatto prevalentemente domanda ad informatica, che ha messo meno posti a disposizione e dove era quindi più difficile entrare. Al contrario le ragazze hanno preferito nella loro scelta il corso di laurea in biologia, dove c'erano più posti a disposizione ed era quindi più facile entrare.

L'esempio presentato è un modello "giocattolo" di un famoso caso accademico che ha coinvolto una delle più famose università americane.

Nel 1973 vennero ammessi ai programmi "graduate" della prestigiosa università di Berkeley il 44% dei ragazzi che avevano fatto domanda ed il 35% delle aspiranti ragazze.

Le associazioni femministe non solo protestarono vivacemente, ma fecero causa all'università, accusandola di discriminazione sessista. L'università si difese (con successo) dimostrando che il risultato cumulativo dipendeva dal paradosso di Simpson. Se si guardava infatti alle situazioni dei singoli dipartimenti, si vedeva che le ammissioni erano sostanzialmente equilibrate, spesso anzi con una prevalenza a favore delle ragazze.

Il sostegno scientifico alla linea difensiva dell'amministrazione dell'università di Berkeley è apparso nell'articolo scientifico di Bickel, Hammel e O'Connell, pubblicato sulla prestigiosa rivista *Science* nel 1975.

Il paradosso di Simpson, a differenza degli altri due "paradossi" presentati prima, è un fenomeno importante, del quale spesso si deve tener conto in situazioni reali.

Consiglio, a chi è interessato a vedere altri esempi, di cercare su Google, ad esempio su

www.math.grinnell.edu/∼mooret/reports/
SimpsonExamples.pdf

dove, oltre al caso Berkeley, vengono presentati altri esempi reali, riguardanti la puntualità di due compagnie aeree americane, dati clinici, la relazioni (negli Stati Uniti) tra le esecuzioni capitali e la razza della vittima e la sopravvivenza dei neonati in dipendenza della assistenza prenatale avuta dalle madri.

Gli esempi sin qui visti potrebbero far pensare che in generale i dati analitici siano più affidabili di quelli cumulativi, perché danno un'informazione più dettagliata. Viene in mente la famosa poesia di Trilussa

Sai ched'è la statistica? È 'na cosa
che serve pe' fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.

Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, lì, la media è sempre eguale
puro co' la persona bisognosa.

Me spiego, da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che se ne magna due.

Il dato cumulativo può dunque nascondere la verità.

Ma succede anche che sia il dato globale quello che dà maggiori informazioni.

Vediamo come questo possa accadere, analizzando un ultimo esempio.

Nel comune di Roccacannuccia vengono assunti 47 nuovi impiegati comunali. Di questi, 32 risultano residenti nella frazione di Roccacannuccia di Sotto (frazione in cui il sindaco in carica ha la sua base elettorale), mentre solo 15 dei nuovi assunti sono residenti nella frazione di Roccacannuccia di Sopra. L'opposizione, guidata da un politico molto votato a Roccacannuccia di Sopra, protesta vivacemente e presenta un'interrogazione. Le due frazioni hanno più o meno la stessa popolazione e gli aspiranti erano altrettanto numerosi: 100 domande erano pervenute da una frazione e 100 dall'altra. Gli esclusi minacciano di ricorrere al TAR per bloccare le assunzioni.

Il sindaco ha la coscienza sporca, perché effettivamente qualche favoritismo c'è stato (ma, si consola il sindaco, in Italia si è visto di peggio). Inoltre la sua coalizione traballa perché, malgrado gli imbrogli, non è riuscito ad accontentare tutti i partiti che lo sostengono. Per questo motivo si rivolge al suo amico e vecchio compagno di liceo, il professor Bravo Affardiconto.

Il professore si fa dare tutti i bandi, i dati anagrafici dei partecipanti e degli assunti ed i provvedimenti di assunzione. Dopo una settimana passata a disaggregare i dati in tutti i modi possibili (per età, sesso, titolo di studio, data del bando, data dell'assunzione, tipologia dell'impiego, ecc), finalmente trova ciò che cercava e mostra trionfalmente la sua relazione al sindaco.

Suddividendo i bandi tra quelli pubblicati prima della estate e quelli dopo l'estate, si ottengono i seguenti risultati.

Per il primo gruppo di bandi hanno presentato domanda 90 *candidati della frazione del sindaco, dei quali* 31 *sono stati assunti (*$\sim 34,4\%$*) e* 10 *dell'altra frazione, dei quali* 4 *sono stati assunti (*40%*). Per il secondo gruppo di bandi hanno presentato domanda* 10 *candidati della frazione del sindaco, dei quali* 1 *è stato assunto (*10%*), e* 90 *candidati dell'altra frazione, dei quali* 11 *sono stati assunti (*$\sim 11,1\%$*).*

Il sindaco può dunque replicare efficacemente alle accuse dell'opposizione, minacciare di denunciarne il capo per calunnia, accusandolo di fare una vergognosa propaganda e di fomentare dannosi scontri tra le due frazioni del comune. Infatti la verità scientificamente dimostrata è che i candidati della frazione di Roccacannuccia di Sopra sono stati favoriti in entrambi i periodi concorsuali!