

La statistica: uno strumento essenziale per capire la realtà



UNIVERSITÀ
DEGLI STUDI
FIRENZE
DiSIA
DIPARTIMENTO DI STATISTICA,
INFORMATICA, APPLICAZIONI
"GIUSEPPE PARENTI"

Leonardo Grilli

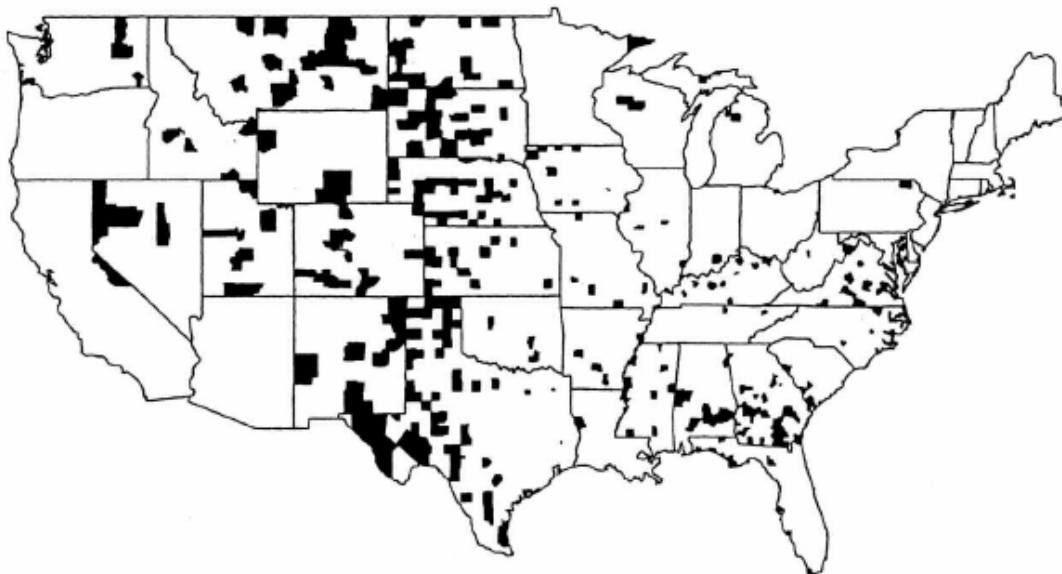
grilli@disia.unifi.it

local.disia.unifi.it/grilli

La statistica come strumento cognitivo

- Lo psicologo sperimentale Daniel Kahneman, premio Nobel per l'Economia nel 2002, racconta la sua attività di ricerca nello splendido libro *Pensieri lenti e veloci*
- Una parte essenziale dell'attività di ricerca di Kahneman (spesso in coppia con Tversky) riguarda le **distorsioni cognitive** (= errori nella comprensione della realtà)
- Alcune distorsioni cognitive sono dovute all'*ignoranza di principi basilari di probabilità e statistica*, ad esempio la relazione tra dimensione campionaria e variabilità (e quindi precisione delle stime)

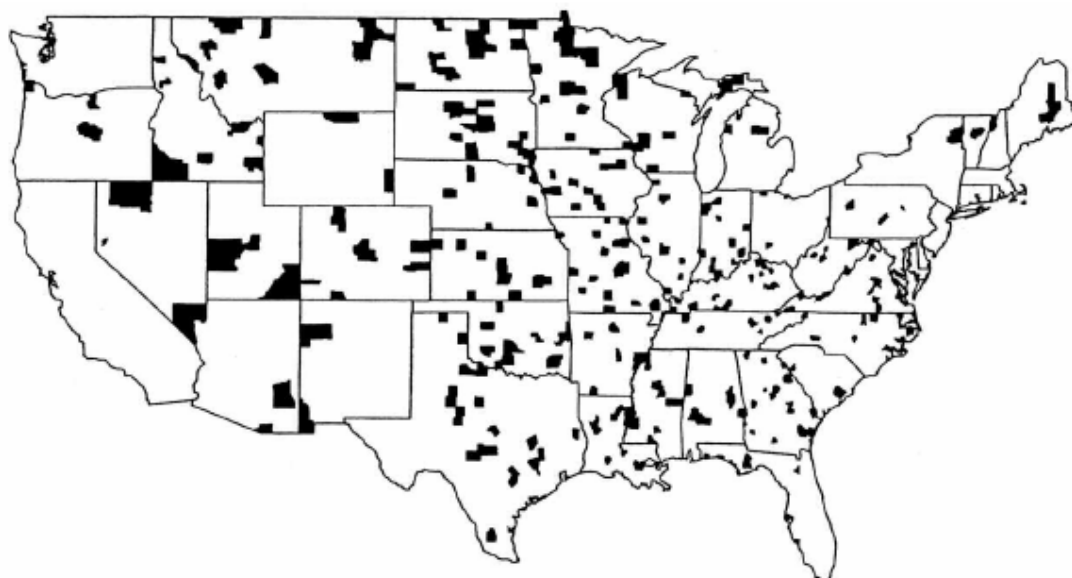




Le aree scure indicano le contee degli Stati Uniti che hanno registrato, nel periodo 1980-1989, i più **bassi tassi** di mortalità per cancro al rene (10% inferiore, tassi relativi ai maschi e standardizzati per età).

Queste contee salutari si trovano principalmente nelle **aree rurali** del Midwest, Sud e Ovest.

Perché in tali contee i tassi sono più bassi?



In questa figura, invece, le aree scure indicano le contee degli Stati Uniti che hanno registrato, nel periodo 1980-1989, i più **alti tassi** di mortalità per cancro al rene (10% superiore, tassi relativi ai maschi e standardizzati per età).

Anche le contee insalubri si trovano principalmente nelle **aree rurali** del Midwest, Sud e Ovest.


Com'è possibile???



Ah, la diversa variabilità!

- Le **aree rurali** includono sia le contee con il più **basso rischio** che quelle con il più **alto rischio** → lo stile di vita rurale non può essere il motivo dei rischi particolarmente bassi o alti
- Il motivo è semplicemente statistico: le contee delle aree rurali hanno **minore popolazione** e quindi sono soggette a **maggior variabilità**

Esempio: quanto si modifica il tasso con un caso in più?



regione grande: $\frac{1000 \text{ casi}}{1000000 \text{ abitanti}} = 1 \text{ per mille}$
un caso in più → $\frac{1001}{1000000} \cong 1 \text{ per mille}$ quasi invariato

regione piccola: $\frac{1 \text{ caso}}{1000 \text{ abitanti}} = 1 \text{ per mille}$
un caso in più → $\frac{2}{1000} = 2 \text{ per mille}$ doppio!

L. Grilli - Statistica 2014/2015

5

Le piccole scuole funzionano meglio?

- Negli anni 90 la Bill & Melinda Gates Foundation finanziò ricerche volte a identificare le caratteristiche delle scuole migliori
- I ricercatori selezionarono le scuole migliori, esaminandole attentamente ... e notarono che **tra le scuole migliori vi è una elevata frequenza di piccole scuole**
- Questi risultati indussero varie istituzioni a investire molte risorse in progetti per incentivare le piccole scuole (al 2001 la Gates Foundation aveva speso 1,7 miliardi di dollari)
- Peccato che i ricercatori abbiano trascurato le scuole peggiori: **anche tra le peggiori vi è una elevata frequenza di piccole scuole!**
- Semplicemente: **piccole scuole = grande variabilità**. A volte anche ricercatori esperti dimenticano questa semplice relazione ...



Wainer H. and Zwerling H.L. (2006) Evidence That Smaller Schools Do Not Improve Student Achievement, *Phi Delta Kappan*, Vol. 88 (4), pp. 300-303.

L. Grilli - Statistica 2014/2015

6

Dimensione campionaria e variabilità

Ecco una tipica domanda che Kahneman e Tversky ponevano alle persone durante i loro esperimenti:



In una data città ci sono due ospedali:

nell'ospedale **grande** nascono ogni giorno circa **45 bambini**

nell'ospedale **piccolo** nascono ogni giorno circa **15 bambini**

In media circa metà dei neonati sono maschi e metà sono femmine; però la percentuale esatta di maschi e di femmine varia di giorno in giorno. Per un anno intero ogni ospedale tenne regolarmente i conteggi del **numero di giorni in cui i nati maschi erano più del 60% del totale**.

In quale ospedale pensate che questo sia avvenuto più spesso?

- Nell'ospedale più grande *risposte 22%*
- Nell'ospedale più piccolo *risposte 22%*
- In entrambi in modo simile *risposte 56%*

Qui non si chiedeva di **quantificare** la relazione tra dimensione campionaria e variabilità, bastava capire il **verso** della relazione ... eppure ci riuscivano in pochi!

Leonardo Grilli - DISIA

7

La legge dei piccoli numeri

- Naturalmente i ricercatori che applicano metodi statistici hanno ben chiaro il verso della relazione tra dimensione campionaria e variabilità (e quindi precisione delle stime) ...
- ... ma questo non basta: per progettare un esperimento occorre quantificare la relazione al fine di decidere **la dimensione campionaria** (es. quante persone intervistare)
- Nel classico articolo del 1971 «**Belief in the law of small numbers**», Kahneman e Tversky mostrano che anche persone con conoscenze di statistica hanno la tendenza a *sottostimare la dimensione campionaria necessaria per ottenere stime sufficientemente accurate* → negli esperimenti di psicologia le dimensioni campionarie erano (sono?) spesso insufficienti perché scelte dai ricercatori in modo intuitivo anziché tramite le note formule

Riferimenti – analfabetismo statistico

- Huff D. (1954) *How to lie with statistics* (in italiano: *Come mentire con la statistica*).
- Krämer W. (2009) *Le bugie della statistica*.
- Paulos J. A. (1988) *Innumeracy: Mathematical Illiteracy and its Consequences*.
- Paulos J.A. (2009) *Un matematico legge i giornali. Difendersi con la logica dai trucchi dell'informazione*.
- Gigerenzer G. (2003) *Quando i numeri ingannano. Imparare a vivere con l'incertezza*.
- Mlodinow L. (2009) *La passeggiata dell'ubriaco. Le leggi scientifiche del caso*.

Grilli L. , La probabilità: capire la realtà e prendere decisioni migliori. Atti di *Pianeta Galileo 2013*.
<http://local.disia.unifi.it/grilli/conferences.htm>

9

Riferimenti – didattica della statistica

- **Guidelines for Assessment and Instruction in Statistics Education (GAISE)** by the American Statistical Association (2005)
<http://www.amstat.org/education/gaise/>
 - Sviluppare il *pensiero statistico*
 - Usare *dati reali*
 - Rinforzare la *comprensione concettuale* piuttosto che la mera conoscenza delle procedure
 - Incentivare l'*apprendimento attivo* in classe
 - Usare la *tecnologia* per sviluppare la comprensione concettuale e l'analisi dei dati
 - Usare *prove con valutazione* per migliorare e valutare l'apprendimento
- Wild C.J. et al. (2011) Towards more accessible conceptions of statistical inference. *J. R. Statist. Soc. A*, 174, pp. 247–295 [interessante per la vivace discussione]
- **Statistica e Probabilità per la scuola**, sito web a cura di Antonio Moro e Giovanni Marchetti <http://local.disia.unifi.it/gmm/scuola/>

Due report: uno per la scuola (Pre K-12) e uno per l'università (college – introductory statistics)

10