

La statistica pubblica per conoscere e decidere

Giovanni A. Barbieri
III giornata per la diffusione della statistica
Firenze, 29 maggio 2006



Informazione e decisioni

- Nella vita quotidiana, non c'è situazione in cui non abbiamo bisogno di informazione per prendere una decisione
- Il più delle volte, non ce ne accorgiamo neppure (reazioni istintive)
- Il processo: informazione → decisione → azione
- Situazioni di scelta in cui le nostre conoscenze e le informazioni che riceviamo sono più scarse e in cui l'esito della decisione è incerto

Qui entra in gioco la statistica



Statistica e vita quotidiana

- Le situazioni in cui abbiamo bisogno di informazione – e in particolare informazione statistica – per prendere una decisione sono molte:
 - un acquisto importante
 - la scelta del corso di studi o della professione
 - una decisione sulla salute
- Nell'impossibilità di conoscere con certezza l'esito della nostra decisione, è comunque meglio poter valutare la probabilità degli esiti possibili, sulla base di un gran numero di situazioni simili

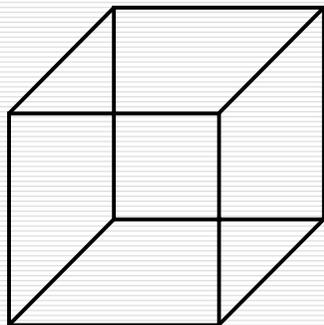


Ma allora perché è così difficile "ragionare statisticamente"? L'illusione della certezza

- "In the world nothing can be said to be certain except death and taxes" (B. Franklin, 13 novembre 1789)
- L'illusione della certezza è profondamente radicata nel *modus operandi* della mente umana (*hard-wired*)
- Secondo alcune teorie, i meccanismi percettivi sono fondati sull'idea che la percezione operi risolvendo le ambiguità
- Illusioni ottiche e illusioni cognitive



Il cubo di Necker



Ragazza o befana?



Lo vedete?



Come funziona?

- Nel primo caso:
 - La profondità è ambigua per la mancanza voluta degli indizi consueti (linee nascoste, ombre...)
 - Ma non percepiamo un'ambiguità, ma piuttosto un'oscillazione
 - Gestalt switch
 - Mai contemporaneamente!
- Nel terzo:
 - Una volta "vista" la figura nascosta, non puoi smettere di vederla
 - Il tuo sistema percettivo è cambiato permanentemente

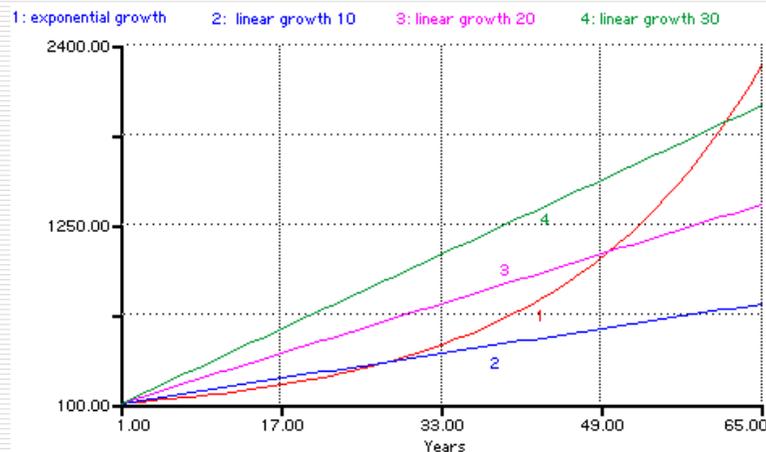
Illusioni cognitive

- Fenomeni esponenziali
- Processi di sconto
- Decisioni razionali
- Il ruolo del buon senso

Fenomeni esponenziali

- Manchiamo di un modulo mentale per capire la crescita esponenziale, e l'approssimiamo con la crescita lineare
 - La leggenda dell'invenzione degli scacchi
 - Thomas R. Malthus (1798), *An Essay on the Principle of Population*
 - *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind* (1972)
- Un modo di ragionare "intuitivamente lineare" (Ray Kurzweil, 2005): diamo per scontato che il tasso di sviluppo corrente continuerà nel futuro
- Per la crescita lineare, il passato è una buona guida per il futuro; per quella esponenziale, è una guida ingannevole (Herman E. Daly)
- Comprendere la crescita esponenziale è la chiave della comprensione delle tendenze future
- Nel lungo termine, la crescita esponenziale determina cambiamenti di scala
- I fenomeni di crescita esponenziale sono pervasivi

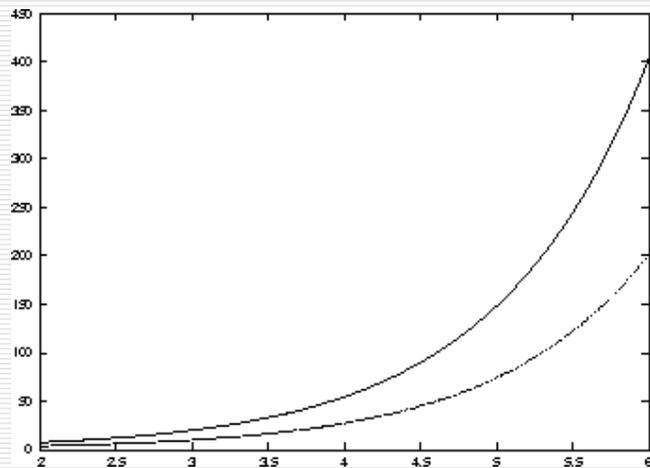
Un esempio grafico



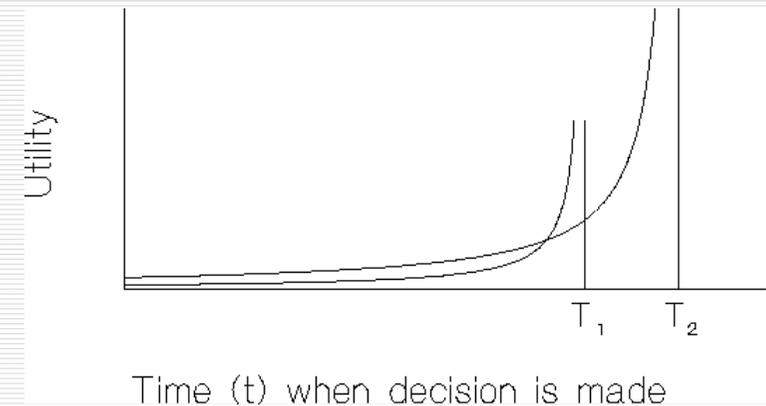
Processi di sconto iperbolico

- George Ainslie (*Breakdown of Will*, 2001) parte dall'osservazione, contenuta già nel *Protagora* di Platone, che le persone tendono a preferire il beneficio immediato e a scontare quello futuro:
 - Se offro ai soggetti d'un esperimento 50 euro subito, o 200 tra cinque anni, la maggioranza sceglie i 50 subito
 - Ma se offro 50 euro tra 10 anni, o 200 tra 15, la maggioranza sceglie i 200 tra 15 anni
- Apparentemente illogico, perché le preferenze cambiano anche se restano costanti sia i compensi assoluti (50 o 200 euro) sia l'intervallo temporale (5 anni)
- Ainslie spiega la situazione con processi di sconto iperbolico, anziché esponenziale, in cui il valore cresce più rapidamente all'avvicinarsi della scadenza
- Nell'esempio grafico, c'è una regione in cui le preferenze sono invertite
- Ad esempio, la domenica pomeriggio preferisco essere riposato e in forma il lunedì, piuttosto che guardare un film a tarda notte. La sera, all'avvicinarsi della proiezione, la prospettiva di guardare il film si fa più allettante ... Il lunedì mattina, essere riposato e in forma è di nuovo più importante ...

Sconto esponenziale



Sconto iperbolico



Decisioni razionali

- Secondo la teoria delle decisioni, le decisioni razionali sono assunte sulla base del calcolo dell'utilità attesa
- In economia, ad esempio, l'idea è che per prendere una decisione importante – sposarsi, o comprare delle azioni – devi seguire questi passi:
 - Elencare tutte le conseguenze possibili di ogni decisione
 - Assegnare a ciascuna una probabilità
 - Associare un valore
 - Tirare le somme
 - Scegliere il valore atteso più alto

Il ruolo del buon senso

- Le persone "non sono massimizzatori dell'utilità probabilisticamente razionali" (Kahneman and Tversky, *Choices, Values and Frames*, 1984)
- Gli atteggiamenti di fronte alle probabilità di guadagno possono essere molto diversi da quelli di fronte al rischio di perdita
 - Per esempio, di fronte alla scelta di ricevere certamente 1.000 euro o di avere una probabilità del 50% di guadagnarne 2.500, i soggetti dell'esperimento tendono a preferire i 1.000 euro certi (anche se la speranza matematica dell'opzione incerta è pari a 1.250 euro)
 - *Avversione al rischio*
 - But Kahneman and Tversky hanno anche trovato che gli stessi soggetti, di fronte a una perdita certa di 1.000 euro oppure alla probabilità del 50% di perderne 2.500, tendono a scegliere l'alternativa più rischiosa
 - *Ricerca di rischio*
- Le scelte umane sono asimmetriche!

Ma allora perché è così difficile “ragionare statisticamente”?

- Esempi tratti da Amos Tversky e Daniel Kahnemann:
 - Il gioco d'azzardo
 - La paura degli aerei
 - I numeri in ritardo (ma anche la borsa)
 - Lisa, 31 anni, single, brillante, attiva sui temi della discriminazione e della pace, laureata in filosofia. Quante probabilità ci sono che sia un'impiegata di banca? E che sia un'impiegata di banca attiva nel movimento femminista?

Un esempio medico

- Una donna di 40 anni fa una mammografia di routine e risulta positiva
- “Che cosa significa?” – chiede al radiologo – “Ho il cancro? Con che probabilità? Il 99%, il 95, il 90? Soltanto il 50%? Che cosa sappiamo?”
- Un terzo dei radiologi della Harvard Medical School ha risposto il 90%
- Ma:
 - La probabilità che una donna di quell'età abbia il cancro al seno è intorno all'1%
 - Se ha il cancro, la probabilità di risultare positiva alla mammografia è del 90%
 - Se non ha il cancro, la probabilità di risultare comunque positiva è del 9%

Un modo diverso di comunicare l'informazione

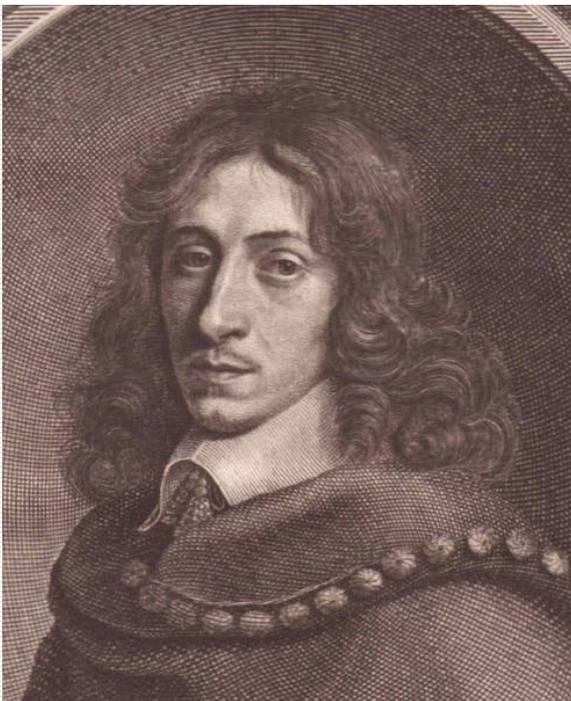
- Pensate a 100 donne
 - Tutte fanno la mammografia
 - Una ha il cancro (questo è l'1%)
 - Fa la mammografia e risulta positiva
 - Delle altre 99 che fanno la mammografia, 9 risultano positive ma non hanno il cancro
 - Quindi, 10 risultano positive al test
 - Di queste, quante hanno realmente il cancro?
- Non il 90%, né il 50, ma il 10%!
- Vi immaginate come è stata quella signora?

Vuol dire che siamo stupidi?

- Non necessariamente: a volte le scelte sono semplicemente preferenze personali
 - È il caso della paura del gioco d'azzardo, ma anche della paura di volare
- A volte, quelle che abbiamo chiamato “reazioni istintive” sono esatte nella maggior parte dei casi
 - La vista non ci inganna quasi mai ...
 - Anche la credenza che il futuro sia influenzato dal passato è quasi sempre corretta (nei rapporti sociali, e persino nelle previsioni del tempo ...)
 - Ma ci sono situazioni “create” apposta:
 - Le estrazioni del lotto sono congegnate in modo da essere indipendenti dall'estrazione precedente
 - Lo stesso accade per la roulette, che è una macchina di precisione progettata per ingannare il nostro “intuito statistico”
 - Anche la borsa opera sulla base di un meccanismo che tende ad annullare ogni deviazione da un percorso casuale

Statistica e democrazia (1)

- L'ignoranza pubblica del rischio ha una base storica
- A differenza della narrativa (storia, mitologia, pettegolezzo) che è elemento essenziale del nostro modo di pensare dai primordi dell'umanità, la statistica pubblica è un risultato culturale recente
 - 1654: uno scambio di lettere tra Pascal e Fermat
 - Il fatto che la nozione di probabilità matematica si sia sviluppata così tardi – più tardi della maggior parte dei concetti filosofici – è stato definito "lo scandalo della filosofia" (Ian Hacking 1990, *The Taming of Chance*)
 - Girolamo Cardano, *De Ludo Aleae* (1525-1565)
 - Probabilità (basata sulla propensione)
 - Fortuna



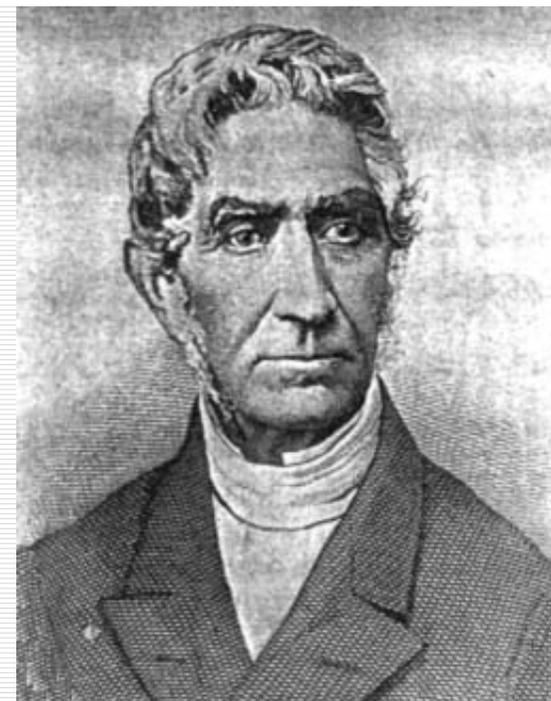
John Evelyn (1620-1706)

- Progressista, tra i fondatori della Royal Society, orticoltore, viaggiatore, diarista
- Ci dà il modo di vedere il mondo con gli occhi di uno scienziato che ignora la statistica
- Avido di numeri:
 - La campana di Notre Dame (alta 13 piedi, 32 di circonferenza, 40.000 libbre di peso)
 - L'obelisco di Piazza San Pietro (costato 37.975 corone ed eretto con il lavoro di 907 uomini e 75 cavalli)
- Ma persone e prodotti non sono ancora entrati nel campo del numerabile:
 - Le città sono "popolose"
 - Misura la ricchezza dell'Olanda contando i canali
- Le assicurazioni esistono, ma sono a interesse invariante o basate meticolosamente su casi individuali
- Orticoltore, ma le annotazioni meteorologiche sono puramente aneddotiche
- I vestiti sono fatti su misura
- La vita umana è basata sull'ideale biblico di 70 anni (anche se Edmund Halley prepara la prima tavola empirica)
- La Royal Society riconosce la variabilità dei fenomeni, ma esita ad assegnarla al caso e a cogliere le regolarità sottostanti
- L'unità fondamentale della realtà è l'individuo singolo



Nessuno scandalo, dopo tutto!

- Nelle società antiche:
 - Chance (chaos)
 - La volontà divina
 - Fortuna (fortuna, fas/nefas)
- Previsione e profitto:
 - Un buon motivo per esplorare il futuro in maggior dettaglio!
 - Il cuore della probabilità: guardare avanti alla ricerca di regolarità nel lungo periodo, per fare pronostici ragionati su ciò che potrebbe accadere nel futuro, sulla base di scommesse, aspettative e verosimiglianza
 - Spostamento dal breve al lungo termine, e dal caso individuale alla regola generale (la legge dei grandi numeri di Bernoulli)
 - Il rischio come espressione del calcolo dell'incertezza



Un passo avanti

- Pensatori sociali come Auguste Comte, Adolphe Quetelet ed Emile Durkheim scoprono che la legge dei grandi numeri può essere applicata alle popolazioni, oltre che ai fenomeni naturali
- Anche se il comportamento individuale è caotico e imprevedibile nel breve periodo, nel lungo il comportamento di gruppi sociali più ampi presenta caratteristiche di stabilità e regolarità
- Le statistiche sulla popolazione rivelano regolarità nella criminalità, nell'epidemiologia, nei tassi di nuzialità e di suicidio; diventa possibile calcolarne e prevederne il rischio e, dunque, controllare e governare la società
- L'aggregato e il lungo termine rivelano ordine: "il caos scompare nei grandi numeri" (Gigerenzer)

Che cos'è la statistica

- Scienza che ha per oggetto lo studio dei fenomeni collettivi suscettibili di misurazione e di descrizione quantitativa
 - Specialmente quando il numero degli individui interessato è talmente elevato da escludere la possibilità o la convenienza di seguire le vicende di ogni singolo individuo
 - Si perviene alla formulazione di leggi di media che governano tali fenomeni, dette leggi statistiche
 - basandosi sulla raccolta di un grande numero di dati inerenti ai fenomeni in esame, e partendo da ipotesi più o meno direttamente suggerite dall'esperienza o da analogie con altri fenomeni già noti
 - mediante l'applicazione di metodi matematici fondati sul calcolo delle probabilità
 - Spesso la raccolta dei dati viene limitata a un campione più ristretto, opportunamente predeterminato in modo da rappresentare fedelmente le caratteristiche generali
- Concepita inizialmente come attività descrittiva di certi fatti sociali e in particolare come attività amministrativa dello Stato, ha via via ampliato i suoi confini, fino a diventare una vera e propria «scienza del collettivo», disciplina con finalità non solo descrittive dei fenomeni sociali e naturali, ma orientata anche a finalità di ricerca

Alcuni punti fermi

- Rapporto tra “statistica” e “fenomeni collettivi”: il tipo di incertezza che la statistica cerca di tenere sotto controllo è insita nel numero degli individui o dei fenomeni
- Semplificare la complessità attraverso un modello – basato sul calcolo delle probabilità – capace di conservare gran parte dell’informazione riducendo drasticamente la molteplicità dei fenomeni
- Misurabilità
- Rilevanza

Misurabilità

- Misura: Il valore numerico attribuito a una grandezza, ottenuto ed espresso come rapporto tra la grandezza data e un'altra della stessa specie assunta come unità (*unità di misura*), e determinato con opportuni metodi o strumenti di misurazione
- Misurazione: L'operazione del misurare, consistente nel confrontare una determinata grandezza fisica con la sua unità di misura, allo scopo di determinare il valore (o misura) della grandezza stessa [...]
 - diretta (o fondamentale, o relativa), quella che consente di determinare direttamente la misura di una grandezza (senza quindi far ricorso alla misurazione di altre grandezze), confrontandola con un campione (per es., la misurazione della massa di un corpo mediante una bilancia, sull'altro piatto della quale vengono poste masse campione)
 - indiretta (o derivata), quella in cui la grandezza da misurare dipende, secondo una relazione funzionale nota, da altre grandezze misurate direttamente (per es., la misurazione della velocità media di un corpo come rapporto tra lo spazio percorso e il tempo impiegato a percorrerlo)

Rilevanza

- Il fatto, la caratteristica di essere rilevante, cioè di notevole importanza o anche gravità, soprattutto riguardo a determinati fini
- Relevance is term used to describe how pertinent, connected, or applicable some information is to a given matter
 - Concetto importante ma elusivo:
 - La soluzione di un problema richiede l'individuazione *ex ante* degli elementi rilevanti che possono concorrere alla sua soluzione
 - Ma la logica convenzionale non lo cattura (una proposizione falsa implica tutte le altre proposizioni, ma non tutte sono rilevanti)
 - La rilevanza è funzione di un obiettivo (*goal dependent*): un elemento (oggetto o proposizione) è rilevante per un obiettivo se e solo se è essenziale all'interno di un piano per conseguirlo (Gorayska e Lindsay)
 - Soddisfa sia le esigenze della logica sia quelle delle attività di problem-solving
 - È definita empiricamente, non sulla base dello stato delle conoscenze o delle credenze

Modello

- Un modello astratto (o concettuale) è una costruzione teorica che rappresenta processi fisici, biologici o sociali, con un insieme di variabili e un insieme di relazioni logiche e quantitative tra loro
- In questa accezioni, il modello consente di ragionare all'interno di uno schema logico astratto e semplificato:
 - Astratto (idealizzato) perché il modello può formulare ipotesi esplicite di cui è noto che – a un certo livello di dettaglio – sono false
 - Semplificato perché ciò consente di pervenire a soluzioni ragionevolmente accurate, trascurando la complessità implicita nel grande numero di variabili e attori del processo modellizzato

Statistica e democrazia (2)

- La parola *statistica* viene dal latino *statisticum collegium* (consiglio di stato) ed è apparentata con *statista*
- "Statistica" è il femminile sostantivato dell'aggettivo "statistico", il cui significato originario era "che riguarda lo Stato, la vita e i problemi dello Stato"
- Il tedesco *Statistik* fu introdotto da Gottfried Achenwall (1749) e faceva riferimento all'analisi di dati sullo Stato
- Assume il significato di raccolta e classificazione dei dati all'inizio dell'Ottocento
- In questo modo, il principale scopo originario della statistica riguarda dati a uso del governo e dell'amministrazione (spesso centralizzata)
- Per tutto il Settecento e gran parte dell'Ottocento, l'informazione statistica era un segreto di Stato, nota solo a un'élite e tenuta nascosta al pubblico
 - Il potere insito nell'informazione statistica è noto da secoli ai governi
 - L'appetito di Napoleone per i dati del suo *bureau de statistique* è leggendario
- La volontà di rendere pubblici i dati economici e demografici è piuttosto recente (circa 1830)

Statistica e democrazia (3)

- La crescente disseminazione di informazione statistica ai cittadini – durante l'Ottocento e il Novecento – è stata associata alla diffusione della democrazia nell'Occidente (Porter 1986)
 - Una delle caratteristiche delle democrazie è di rendere disponibile una grande massa di informazioni statistiche ai cittadini
 - Ma i cittadini sono molto selettivi nei loro interessi



L'impero della statistica

- All'inizio del XX secolo, H. G. Wells, padre della fantascienza, scrive: "Se vogliamo avere cittadini istruiti nella moderna società tecnologica, dobbiamo insegnare tre cose: a leggere, a scrivere e a pensare in modo statistico"
- Nel 1954, R. A. Fischer scrive: "la statistica è l'aspetto particolare del progresso umano che ha dato la sua impronta al XX secolo"



Statistica e probabilità sono pervasive

- La statistica descrittiva su ogni numero di ogni quotidiano
- La statistica matematica indispensabile:
 - In epidemiologia e nella ricerca medica
 - Nel marketing e nel controllo di qualità delle aziende
 - Nella contabilità
 - Nelle previsioni, meteorologiche ed economiche
 - Nei sondaggi politici
 - Nello sport
 - Nella ricerca e sviluppo
 - Nelle assicurazioni ...
- In molte scienze, procedure standardizzate di inferenza statistica definiscono il processo di formazione delle ipotesi, condurre gli esperimenti e analizzarne i risultati: in una parola, il metodo scientifico

I nuovi esperti

- L'autorità degli esperti non si basa tanto sull'esperienza concreta e sul giudizio personale, quanto sull'informazione e sull'applicazione di tecniche formali
- La statistica è utilizzata in una pluralità di accezioni:
 - Descrizione
 - Inferenza
 - Misurazione
 - Modelli
- Strumenti statistici, sostenuti da ipotesi sulle sottostanti distribuzioni di probabilità, misurano (e contribuiscono a creare) entità come il QI, il PIL, il clima di fiducia ...
- La statistica ha (ri)definito concetti come "intelligenza", "opinione pubblica" e "ricchezza"
- Ha sostituito i giudizi con le regole

La statistica pubblica

- In tutti i Paesi democratici, la statistica ufficiale è finanziata con i soldi dei contribuenti e messa a disposizione di tutti i cittadini
- Ma perché gode di una posizione simile a quella dei servizi pubblici essenziali?
 - A un primo sguardo, la statistica appare meno necessaria dell'assistenza sanitaria o della pubblica sicurezza

L'informazione statistica come bene pubblico

- Si deve a Stiglitz (1999) la connotazione della conoscenza come bene pubblico globale
- Due caratteristiche essenziali:
 - Non-rivalità: il consumo di un individuo non diminuisce quello di un altro
 - Non-escludibilità: escludere un individuo dalla fruizione del bene è impossibile o quanto meno difficile
- La conoscenza di un teorema matematico, per esempio, soddisfa entrambi i criteri

Non-rivalità

- La non-rivalità della conoscenza – cioè che il costo marginale di un individuo aggiuntivo ammesso a goderne i benefici sia nullo – ha implicazioni importanti
- Anche se fosse possibile escludere l'individuo dal godimento dell'informazione, ciò non sarebbe desiderabile, proprio perché il costo marginale della condivisione è nullo
- Se l'informazione deve essere utilizzata in modo efficiente, non può dunque essere fornita privatamente
- Il prezzo, pari al costo marginale, dovrebbe essere nullo. Ma a questo prezzo sarebbe prodotta soltanto l'informazione il cui costo di produzione è nullo
- Beninteso, l'acquisizione e l'uso della conoscenza possono avere un costo per il fruitore (è quello che accade, ad esempio, per il prelievo d'acqua da una fonte pubblica, o per scaricare informazione statistica dal web)
- Inoltre, la trasmissione della conoscenza può avere costi associati rilevanti e un privato può fornire i relativi servizi a un prezzo che riflette i costi marginali di trasmissione
- Ma questo non ha alcun riflesso sulla natura pubblica del bene "informazione"

Non-escludibilità

- La non-escludibilità, dal canto suo, comporta che nessuno *possa* essere escluso dal godimento e implica l'impossibilità – sotto il profilo della razionalità economica – della produzione privata di conoscenza
 - Si supponga che sia stato scoperto un teorema con rilevanti implicazioni pratiche e che esso non possa essere tenuto segreto. In regime di concorrenza il suo prezzo sarebbe pari a zero
- Alcune forme di conoscenza, però, sono (o possono essere rese) escludibili, come nel caso dei segreti industriali
- L'impresa, però, corre rischi elevati: un concorrente potrebbe impossessarsi del segreto e trarne vantaggio, tanto più se il segreto riguarda un'informazione "pura"
 - È più difficile (anche se non impossibile) impossessarsi della formula di un nuovo prodotto che non impossessarsi di informazioni di mercato...
- Un altro criterio di protezione consiste nell'assegnare diritti esclusivi di sfruttamento (tramite brevetto o copyright) per un periodo limitato di tempo, a condizione che i dettagli dell'invenzione siano resi pubblici: in questo modo, tuttavia, una parte consistente dell'informazione originaria viene resa pubblica
 - Ad esempio, l'invenzione del rayon nel 1910 ha diffuso la conoscenza della "fattibilità" delle fibre artificiali e sintetiche, in sé un'informazione di enorme valore commerciale
- La natura cumulativa e incrementale della conoscenza fa sì che, di norma e in buona sostanza, il criterio della non-escludibilità sia soddisfatto.

Tre raccomandazioni dell'Onu

- Il rapporto dell'Undp in cui è pubblicato l'intervento di Stiglitz citato sottolinea la necessità di porre attenzione a tre aspetti del carattere pubblico della conoscenza:
 - La conoscenza deve essere pubblica per quanto attiene al suo *consumo*, perché ogni cittadino deve avere libero accesso all'informazione
 - Deve essere pubblica quanto alla *partecipazione*, perché il processo politico che porta a decidere quale e quanta informazione debba essere prodotta deve prendere in considerazione le opinioni e la volontà di tutti gli attori interessati (*stakeholders*), in un processo aperto
 - Deve essere pubblica con riferimento alla *diffusione dei benefici*, perché ogni cittadino ha titolo a goderne
- Queste connotazioni del carattere pubblico dell'informazione lasciano aperta la questione se essa possa o debba essere prodotta soltanto da un soggetto pubblico

Statistica pubblica e correzione delle asimmetrie informative

- Il costo della raccolta delle informazioni elementari è molto elevato
- I soggetti più deboli (individui, imprese, istituzioni, gruppi sociali) in posizione di svantaggio, non potendo acquisire informazione statistica sul mercato (asimmetrie informative)
- In una società democratica, i cittadini devono essere posti in una posizione di eguaglianza iniziale: l'eguaglianza delle opportunità e delle possibilità
- Lo Stato provvede a produrre statistiche ufficiali e pubbliche – cioè certificate e messe a disposizione di tutti – come strumento di garanzia dell'eguaglianza dei diritti dei cittadini