



DATI AMMINISTRATIVI, METODI E STATISTICHE PER LE POLITICHE TERRITORIALI

A CURA DI
FLAVIO VERRECCHIA



FrancoAngeli

OPEN  ACCESS



Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più:

http://www.francoangeli.it/come_publicare/publicare_19.asp

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "Informatemi" per ricevere via e-mail le segnalazioni delle novità.

DATI AMMINISTRATIVI, METODI E STATISTICHE PER LE POLITICHE TERRITORIALI

A CURA DI
FLAVIO VERRECCHIA

FrancoAngeli
OPEN  ACCESS

Si ringrazia il Dipartimento di Economia, Management e Metodi Quantitativi dell'Università degli Studi di Milano per il supporto dato a questo progetto nell'ambito del programma quinquennale 2018-2022 "Dipartimenti di eccellenza" finanziato dal Ministero dell'Istruzione, dell'Università e della Ricerca al fine di incentivare l'attività dei Dipartimenti delle Università statali che si caratterizzano per l'eccellenza nella qualità della ricerca e nella progettualità scientifica, organizzativa e didattica (legge 11 dicembre 2016, n. 232 - Art. 1, commi 314-338).

Copyright © 2019 by FrancoAngeli s.r.l., Milano, Italy.

Publicato con licenza *Creative Commons Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale* (CC-BY-NC-ND 4.0)

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

INDICE

PREFAZIONE	pag.	7
<i>di G. Manzi</i>	»	9
INTRODUZIONE AL VOLUME		
<i>di G. Garofalo</i>		
PARTE PRIMA – <i>Dati e metadati</i>		
Introduzione	»	11
<i>di S. Figini e A. Venturi</i>		
Capitolo 1 – Microdati integrati economici e demosociali	»	13
<i>di S. Casacci e D. Ercolani</i>		
Capitolo 2 – Metadati per capire i dati	»	21
<i>di D. Ferrazza</i>		
Capitolo 3 – Qualità dei dati amministrativi	»	29
<i>di M. E. Comune e L. C. M. Viviano</i>		
PARTE SECONDA – <i>Metodi</i>		
Introduzione	»	37
<i>di P.M. Chiodini</i>		
Capitolo 4 – Gestire i dati amministrativi	»	39
<i>di S. Ballabio, D. Ferrazza e F. Verrecchia</i>		
Capitolo 5 – Costruire gli indici sintetici	»	47
<i>di M. Mazziotta</i>		
Capitolo 6 – Rappresentare le informazioni territoriali	»	53
<i>del GdL Istat - SPoT</i>		
PARTE TERZA – <i>Applicazioni su dati amministrativi negli Enti locali</i>		
Introduzione	»	63
<i>di A. Lentini</i>		
Capitolo 7 – L’esperienza del Comune di Brescia per le politiche locali	»	65
<i>di M. Trentini</i>		
Capitolo 8 – L’esperienza della Regione Lombardia per la previsione della domanda di lavoro e la definizione dell’offerta formativa	»	75
<i>di F. Rappelli e V. Ricciari</i>		

PARTE QUARTA – *Applicazioni sperimentali su dati amministrativi*

Introduzione	pag.	83
<i>di M. Filandri</i>		
Capitolo 9 – <i>Working poor</i> nel Nord-Ovest	»	85
<i>di S. Cariello e M. Carbonara</i>		
Capitolo 10 – Proposta di un indice di vulnerabilità socioeconomica	»	93
<i>di R. Succi, F. Vannoni e A. Vitalini</i>		
Capitolo 11 – Redditi delle famiglie e condizioni di fragilità in Lombardia	»	101
<i>di S. Casacci</i>		
Capitolo 12 – Disparità di genere nel mercato del lavoro in Lombardia	»	109
<i>di A. L. Carulli, L. Sicuro e D. Tucci</i>		
Capitolo 13 – Neet e Quasi-Neet in Lombardia	»	117
<i>del GdL Istat - SPoT Lombardia</i>		
Capitolo 14 – Non profit in Lombardia: una risposta resiliente?	»	123
<i>di S. Della Queva, M. Nicosia e S. Stoppiello</i>		
CONCLUSIONI E PROSPETTIVE FUTURE	»	131
<i>di F. Verrecchia</i>		
APPENDICI		
Appendice 1 – Redditi	»	133
Appendice 2 – Comuni Frontalieri	»	135
Appendice 3 – <i>Missing data</i>	»	137
RIFERIMENTI BIBLIOGRAFICI	»	139
AUTORI	»	147
RINGRAZIAMENTI	»	149
WORD CLOUDS DEI CAPITOLI	»	151

PREFAZIONE

di Giancarlo Manzi

“Data! data! data!” he cried impatiently. “I can’t make bricks without clay.” Questa esclamazione – attribuita a Sherlock Holmes nel racconto *“L’avventura dei faggi rossi”* di Conan Doyle, quando le supposizioni di Watson non portavano a nulla di dimostrabile per risolvere il caso del mistero dell’ala proibita della casa del signor Rucastle, in una delle più famose avventure del *detective* britannico – non è mai stata così attuale come in quest’era della cosiddetta *rivoluzione dei big data*.

Malgrado non esista una separazione rigorosa tra *big data* e altri dati, il termine *big data* indica genericamente una raccolta di dati così estesa da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore o conoscenza.

L’apporto informativo dei *big data* ha avuto un impatto enorme in ogni campo. Se si considera, ad esempio, l’unità di misura più comune per le dimensioni trattate in ambito di analisi dei dati, in pochi anni si è passati dall’utilizzo del gigabyte (10^9 byte) allo zettabyte (10^{21} byte) e più recentemente allo yottabyte (10^{24} byte). Le dimensioni dei *data center* di Google arrivano a comprendere estensioni ormai paragonabili a decine di campi di calcio, così come il loro consumo di energia richiede impianti di raffreddamento equiparabili a quelli delle centrali nucleari.

Ma non è solo relativamente alle dimensioni delle basi di dati da gestire che le tecnologie informatiche applicate ai dati hanno fatto passi da gigante. In generale, infatti, oltre al “volume” l’evoluzione ha riguardato anche la “velocità”, la “varietà”, il “valore” e la “veridicità” dei dati, vale a dire tutte le “5 V” che molti studiosi utilizzano per sintetizzare le “componenti” dei *big data*. A queste vanno aggiungendosi altre componenti dell’“ecosistema” dei *big data*, come il “*real-time analytics*”, il “*data linkage*” (Demchenko et al., 2014) e, nel tempo, altre ancora.

La crescente facilità di accesso, di gestione e l’economicità dei *big data* non è priva di controindicazioni. Ci sono infatti anche dei rischi legati alla loro diffusione. Ad esempio, se da un lato si tende sempre più a cercare in essi la soluzione informativa unica e risolutiva rispetto alle esigenze conoscitive di una società globale immersa in un processo evolutivo sempre più veloce e coinvolgente, dall’altro, spesso si dimentica di analizzare le forzature metodologiche sottostanti il loro utilizzo. In tal senso, se si considera il confronto tra una fonte di *big data* come i *social network* e l’indagine d’opinione tradizionale, le informazioni provenienti dai *social network* sono più tempestive (la V della velocità), più veritiere rispetto al *sentiment* reale (la V della veridicità) e più commerciabili (la V del valore). Tuttavia, al pari, non si può tralasciare di considerare che le analisi basate sui *big data* hanno problemi metodologici non trascurabili nell’ambito dell’intero processo informativo: dalla corretta interpretazione testuale nel caso dei *tweets*, alla non rappresentatività di intere sezioni della popolazione, all’incertezza nella valutazione della variabilità delle stime, ecc. Il *mainstream*, a favore dei *big data* e degli *analytics*, tende a non considerare queste problematiche che invece non possono essere trascurate, soprattutto dagli enti certificatori dell’informazione come gli Istituti nazionali di statistica.

Pertanto, nell'ambito della Statistica ufficiale, la sfida consiste nel considerare fonti di informazione quantitativa sempre nuove ma isolandone i difetti e valorizzandone i pregi. Oggi in Italia l'Istat si è dotato di una infrastruttura orientata allo studio e alle analisi basate sui *big data*, tanto che anche il censimento della popolazione – diventato permanente – ha visto l'adeguamento, coerentemente ai tempi che cambiano, delle procedure di rilevazione dei dati. Infatti, le informazioni su cittadini e famiglie sono fortemente basate sulle anagrafi dei comuni e su altre fonti amministrative che possono essere considerate vere proprie fonti di informazione di tipo *big data*.

Ecco dunque che anche strumenti di indagine statistica tradizionali come l'indagine campionaria, se adeguati e integrati al mondo dei *big data*, diventano anche più importanti rispetto al passato perché, possono certificare e validare l'informazione proveniente dai *big data*. Al giorno d'oggi, infatti, conviene integrare le fonti di informazione piuttosto che metterle in competizione se si vogliono ottenere i migliori risultati in termini di bontà dell'informazione statistica.

Tra le fonti di informazione che sono maggiormente sfruttate per una conoscenza più completa dei fenomeni collettivi e che devono essere trattate come *big data* ci sono appunto le fonti di dati amministrativi. L'Istat, per alcuni fenomeni, ha tradizionalmente utilizzato il dato amministrativo quale fonte statistica importantissima da integrare con tutte le altre. Tuttavia, le importanti potenzialità dello sfruttamento del dato amministrativo con finalità conoscitive sono state finora limitate per motivi sia tecnici sia di integrazione e di coordinamento fra i vari Enti titolari delle basi dati amministrative.

Questo volume raccoglie alcune sperimentazioni nate nell'ambito del sistema formato da amministrazioni pubbliche ed enti privati che produce tutte le statistiche ufficiali in Italia (Sistan), volte all'integrazione dei dati amministrativi – oltre che rispetto alle altre fonti di informazione – e al loro utilizzo per lo studio dei fenomeni finalizzato al governo dei territori. L'integrazione è importante soprattutto nel processo di armonizzazione delle fonti statistiche territoriali. La tempestività e la correttezza dell'informazione statistica a livello territoriale di granularità sub-provinciale assume importanza crescente soprattutto quando il dato elementare può essere aggregato e integrato così facilmente come oggi le tecniche informatiche consentono.

Migliorare le tecniche di integrazione delle fonti di dati, qualsiasi esse siano, aumentandone l'efficienza – anche riguardo alla riduzione dei costi – è diventato dunque estremamente importante. Allo stesso tempo, i dati amministrativi hanno tutto il potenziale per diventare una fonte di informazione rilevante per la statistica ufficiale. È necessario allora aumentare le sinergie tra di essi, captarne le potenzialità di *linkage* e di aggregazione e sfruttare le moderne tecnologie dell'*information technology*. È quello che il Sistan sta cercando di fare in questi ultimi anni e questo volume è uno dei risultati di questa prospettiva.