

Statistico vs data scientist? Le nuove sfide della professione

Domenica Fioredistella Iezzi

Dipartimento di Scienze storiche, filosofico-sociali, dei Beni culturali e del Territorio
Università di Roma "Tor Vergata"

e mail: stella.iezzi@uniroma2.it

Sommario

Nella nomenclatura e classificazione delle Unità professionali dell'ISTAT (CP2011), la figura dello statistico è presente in tre diversi macro gruppi con differenti denominazioni: statistico, tecnico statistico e addetto ai servizi statistici. La classificazione di queste unità professionali fa riferimento a tre livelli di istruzione, rispettivamente, alla laurea specialistica o di vecchio ordinamento, alla laurea triennale e al diploma di scuola media secondaria. Le competenze previste per questi lavoratori non includono, in maniera esplicita, la raccolta, l'analisi e l'interpretazione di Big Data. L'obiettivo di questo lavoro è individuare le nuove competenze dello statistico professionale e illustrare le nuove sfide che tale professione richiede.

1 Lo statistico nella classificazione ISTAT 2011

Secondo la nomenclatura e classificazione delle Unità Professionali dell'ISTAT (CP2011), la figura dello statistico è individuabile in tre diversi grandi gruppi di professioni (GGP): 1) intellettuali, scientifiche e di elevata specializzazione (GGP n.2); 2) le tecniche (GGP n.3); 3) le esecutive del lavoro d'ufficio (GGP n.4). Ogni GGP identifica tre diverse unità professionali, che presentano caratteristiche molto differenti sia per la formazione, che il lavoratore ha conseguito, sia per le competenze possedute dallo stesso (Iezzi, 2008). Lo Statistico (GGP n.2) per essere tale deve aver acquisito una laurea di secondo livello o una di vecchio ordinamento nelle discipline statistiche o equipollenti e deve condurre ricerche su concetti e teorie fondamentali della scienza attuariale e della statistica, incrementare la conoscenza scientifica in materia, applicare le relative teorie e tecniche per raccogliere, analizzare e sintetizzare informazioni, per definire modelli di interpretazione di dati, per individuare soluzioni statistiche da adottare nei vari settori della produzione di beni e servizi e della stessa ricerca scientifica. Esempi di questa

professione sono l'attuario, l'esperto in ricerca operativa, lo statistico, lo statistico applicato, lo statistico economico, lo statistico esperto in controlli di qualità, lo statistico metodologico e il demografo. I tecnici Statistici (GGP n. 3) hanno conseguito, invece, una laurea di primo livello nelle discipline statistiche o equipollenti e lavorano come specialisti nella ricerca sperimentale, nelle indagini e in altre ricerche demografiche, epidemiologiche ed economiche, nel controllo e nell'applicazione delle procedure di ricerca e di acquisizione dei dati, nella gestione sul campo delle rilevazioni, nel controllo della qualità dei dati rilevati e nella elaborazione statistica degli stessi. Esempi di tecnico statistico sono l'attuario junior, il tecnico di ricerca operativa, il tecnico di statistica demografica, il tecnico di statistica economica, il tecnico di statistica sanitaria, il tecnico statistico e il tecnico statistico di controllo di qualità. Infine, gli addetti ai servizi statistici (GGP n. 4) non sono necessariamente laureati, ma più verosimilmente hanno conseguito un diploma di scuola media secondaria. Nel loro lavoro, rilevano dati, fanno elaborazioni di tipo elementare, anche con l'uso di tecnologie informatiche. Esempi di professioni sono l'addetto agli uffici economici e statistici, l'addetto al calcolo attuariale, l'addetto al

controllo statistico, l'addetto alla raccolta di documentazione statistica, addetto alle rilevazioni statistiche, l'addetto rilevazioni sistematiche, l'assistente di ricerche di mercato, il codificatore, il rilevatore di mercato, il rilevatore di prezzi e l'attuario junior (Gallo & Loré, 2012; Istat, 2013). Tutti questi profili di lavoratori, oltre alle statistiche tradizionali, come l'uso di dati amministrativi, statistiche sociali e economiche e censimenti, s'imbattono in nuove sfide legate soprattutto all'uso di tecnologie digitali, che consentono l'acquisizione di dati da molte fonti non ufficiali, reperibili, in molti casi, dal web. La capacità di generare conoscenza attraverso l'integrazione di dati diversi e a differenti livelli territoriali, la possibilità di accedere a *open government* e *open data*, la necessità di fare *scaping the web* richiede allo statistico, per esercitare la professione, competenze acquisibili attraverso una formazione continua, la cui qualità possa essere riconosciuta anche da terzi, che si affideranno alla loro professionalità per risolvere nuovi e complessi problemi di analisi di dati nei campi più svariati.

L'obiettivo di questo lavoro è individuare le nuove competenze richieste allo statistico professionale e illustrare le nuove sfide professionali.

2 Le nuove frontiere della statistica: il data scientist

La statistica tradizionale si è da sempre confrontata con la scarsità di dati, per questo motivo ha dovuto predisporre tecniche per inferire regolarità, probabilità, previsioni. Nell'era dei Big Data, le informazioni sono cresciute in maniera esponenziale, quindi, dalla mancanza di dati si è passati a una sovrabbondanza di informazioni, spesso non strutturate, da codificare e ripulire, analizzare e interpretare. Questo nuovo scenario ha spinto lo statistico a confrontarsi non solo con le competenze tipiche della sua disciplina, ma ad acquisire nuovi *skill*, soprattutto nel campo dell'informatica e delle Scienze Sociali. Fin dagli anni '90, Brunet (1995) ha riconosciuto al web un ruolo preminente per la raccolta di informazioni, lo stesso, inoltre, ricordava che "la statistica ha sempre amato i grandi spazi, le grandi masse di dati e la legge dei grandi numeri. Internet apre le sue ricchezze alla

statistica, senza controllo, senza ritardo, senza spesa e senza limiti". Nella sua sperimentazione la parola "Rome" compariva 7.111 volte, posizionandosi al 24esimo posto tra le città più citate su web. Oggi, 2013, la stessa espressione compare trecentosessantotto milioni di volte (Bolasco, 2013).

La disponibilità di materiali in Internet è una nuova e importante fonte per la raccolta di dati, ma richiede che lo statistico abbia delle competenze informatiche per raccogliere informazioni e organizzarle in modo tale da rendere utilizzabili per un'analisi statistica; conoscenze nel campo della Linguistica per disambiguare le informazioni non strutturate; pratica nelle Scienze Sociali per compenetrare il settore in cui si opera.

Negli Stati Uniti, l'esperto di big data è stato denominato "Data Scientist" ossia lo "scienziato del dato". Alcuni atenei statunitensi hanno già progettato corsi di laurea dedicati a questa nuova figura professionale, che ha come competenze fondamentali la conoscenza della statistica e del calcolo delle probabilità. Secondo un recente studio della McKinsey, negli Usa, sono necessari tra i 140 e 190 mila laureati che abbiano gli *skill* utili per maneggiare i big data¹, che siano in grado, ad esempio, di monitorare la mappa dei movimenti di milioni di persone, di aggregare i dati mediante la localizzazione dei loro cellulari. In Italia, si comincia a parlare della necessità di professionisti in grado di lavorare con i big data. Nel 2013, il sole 24 ore, ha pubblicato 106 articoli su questo tema, sono state usate 49.118 occorrenze, 8.261 forme grafiche differenti e 3.888 hapax (47,06% delle forme grafiche e 7,80 delle occorrenze), un piccolo vocabolario per descrivere le nuove sfide lanciate da Internet. Le parole lemmatizzate più ricorrenti sono "dato" che appare 275 volte, "informazione" 145, "azienda" 138, "dollaro" 110, mentre le parole "statistica/statistico" solo 10 volte (Figura 1).

¹Cfr il sito web:

<http://www.teradatamagazine.com/v12n03/Features/The-Rise-of-the-Data-Scientist/>.

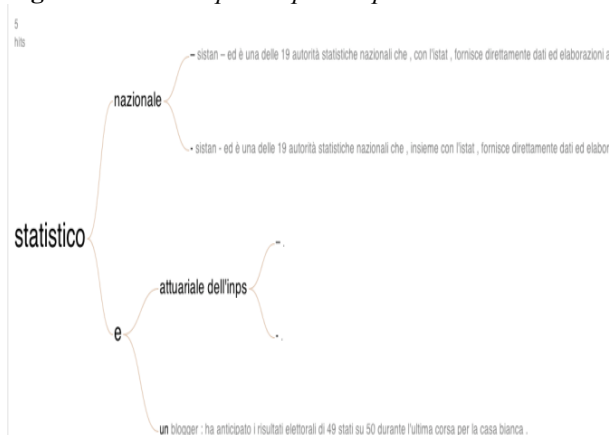
Figura 1 Nuvola di parole proveniente dal corpus degli articoli de il sole 24 ore che contengono l'espressione "big data" – anno 2013



Fonte: Elaborazioni su il sole 24ore

Sorprende che l'associazione tra "dato" e "statistico/e" sia così bassa, mentre è evidente la relazione con il mercato del lavoro (aziende appare 136 volte e impresa 58 volte) e l'informatica (software appare 95 volte). L'analisi delle concordanze mostra che l'espressione "statistico" è associata alle parole "sistema statistico nazionale", "attuario" e "previsioni in campo elettorale" (Figura 2. I verbi più impiegati sono: "usare" (60 volte), "elaborare e analizzare" (50 volte) e "costruire" (47 volte), inoltre, la sequenza di parole di senso pieno più utilizzata è "analisi/elaborazione dati", ma le professioni più citate sono informatico (42 volte), *data scientist* (19 volte) e ingegnere (10 volte).

Figura 2 Albero di parole per l'espressione "statistico"



Fonte: Elaborazioni su il sole 24ore

Il *data scientist* non è stato ancora codificato nel nostro sistema delle professioni, ma di fatto è un lavoratore le cui competenze hanno già un effetto dirompente in tutti i settori della conoscenza, ad esempio, nell'epidemiologia un gruppo di ricercatori è riuscito a prevedere ad individuare per l'HIV il ruolo fondamentale degli antigeni eritrocitari, utilizzando le ricerche pubblicate sul motore di ricerca PubMed dal 1984 al 2013 (MOTSWALEDI *et al.*, 2013). Una ricerca sistematica su cinque banche dati (Medline, Global Health, Embase, Social Policy e Web of Science) ha permesso di stimare la prevalenza dei femminicidi nel mondo (Stöckl *et al.*, 2013). I due noti social media Facebook e LinkedIn utilizzano i Big data per analizzare le reti di conoscenze di milioni di persone iscritte, suggerendo altri nostri amici che sono già in rete, con un tasso di successo elevatissimo. Questi sono soltanto alcuni degli innumerevoli esempi che possono essere fatti nei campi più svariati del sapere. In tutte le scienze, infatti, si sta affermando il principio che la conoscenza ha un effetto *butterfly*, ossia che se una farfalla batte le ali a Pechino in marzo, allora in agosto il tracciato degli uragani dell'Atlantico sarà completamente diverso. Ciò ha portato a riflettere sui confini dei diversi campi di ricerca e le nuove competenze richieste ai professionisti di oggi e di domani.

L'avvento del XXI secolo si è caratterizzato per la disponibilità di enormi quantità di informazioni digitali, secondo il rapporto IDC – Global Go to Market Services 2011, in Internet transitano ogni mese un flusso di dati pari a 150 exabytes². Se lo statistico applicato saprà diventare anche *data scientist*, acquisendo nuove competenze e valorizzando quelle giù possedute, allora sicuramente riuscirà a districarsi in questo nuovo mondo virtuale, ottenendo risultati in tutte le scienze, aprendo un nuovo ciclo di ricchezza e benessere, generando un valore aggiunto all'analisi di dati complessi.

3 Riferimenti bibliografici

Bolasco S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*, Carocci, Roma.
 Brunet E. (1995). *Web et statistique. L'exemple*

² Cfr Digital Universe, sito web: italy.emc.com

du mot Rome, in *Cahiers de Lexicologie*: 71-94
Didier érudition, Paris.

Gallo F., Lorè B. (2012). Training on the new occupational classification: the Italian experience. *Working Papers*, **12**, ISTAT, Roma.

Iezzi D.F. (2008). *I lavoratori come informatori delle qualità delle professioni*. In: Fabbris L. Definire figure professionali tramite testimoni privilegiati. **1**: 135-151, Padova: CLEUP.

ISTAT (2013). *La Nomenclatura delle Unità Professionali*, ISTAT, Roma.

Motswaledi MS, Kasvosve I, Oguntibeju OO. (2013) The role of red blood cells in enhancing or preventing HIV infection and other diseases. *Biomed Res Int*. Epub 2013 Oct 10. Review.

Stöckl H, Devries K, Rotstein A, Abrahams N, Campbell J, Watts C, Moreno (2013), The global prevalence of intimate partner homicide: a systematic review. *Lancet*. **382** (9895):859-65.