

Tanto quanto?

Per i livelli essenziali di statistica nel territorio.

LE PAROLE PER CAPIRE I NUMERI I NUMERI PER CAPIRE IL MONDO

Giovanni A. Barbieri, Usci

Taranto, 4 luglio 2019

Sommario

- L'esperienza dell'Istat
- I numeri per capire il mondo: quantificare e misurare
- Le parole per capire i numeri: meta-informazione e linguaggio della statistica
- Un esempio: disoccupati, occupati, inattivi
- Dalla parte degli utenti: tradurre tra diversi meta-linguaggi
- Rifuggire dall'eccesso di semplificazione

L'esperienza dell'Istat

- Rendere i servizi della statistica pubblica **utili, rilevanti e affidabili** per tutti
 - [D.Lgs. 322/1989](#) «I dati elaborati nell'ambito delle rilevazioni statistiche comprese nel programma statistico nazionale sono patrimonio della collettività» (art. 10)
 - [Fundamental Principles of Official Statistics](#) (1992-94, 2013-14)
 - [Statistics: a Matter of Trust](#) (24 febbraio 1998)
 - L'esperienza del *Rapporto annuale* (1994-2018)
- **Rilevanza** e **fiducia** sono in primo luogo nelle mani degli utenti
- **Ascoltare** le esigenze e i bisogni degli utenti

I numeri per capire il mondo

- Dell'esistenza di quello che non si può **misurare**, o quanto meno **quantificare**, conviene dubitare
 - «Dico spesso che quando si può misurare ciò di cui si sta parlando, ed esprimerlo in numeri, si sa qualcosa al riguardo; ma quando non si può misurare, quando non si può esprimere in numeri, la conoscenza è magra e insoddisfacente; può essere l'inizio della conoscenza, ma si è raramente avanzati allo stadio della scienza, quale che sia la materia» [Lord Kelvin, 1883]
 - Quantificare: introdurre precisi elementi di valutazione quantitativa; tradurre in termini di quantità, tradurre in cifre, in valori numerici
 - Misurare: determinare la misura di una grandezza (lunghezza, distanza, superficie, peso...), confrontandolo con uno standard predefinito

L'importanza della meta-informazione, ovvero la risposta alla domanda fondamentale sulla vita, l'universo e tutto quanto

- «Quarantadue! – urlò Loonquawl – Questo è tutto ciò che sai dire dopo un lavoro di sette milioni e mezzo di anni?»
- «Ho controllato molto approfonditamente – disse il computer – e questa è sicuramente la risposta. A essere sinceri, penso che il problema sia che voi non abbiate mai saputo veramente qual è la domanda.»



Le parole per capire i numeri

- Un numero non dà alcun contributo alla conoscenza, se non sai che cosa questo numero significa: l'insieme delle sue ramificazioni semantiche
- Le parole dunque – anche nella statistica – sono importanti
 - «Chi parla male, pensa male e vive male. Bisogna trovare le parole giuste: le parole sono importanti!» [Michele Apicella in *Palombella rossa* di Nanni Moretti]
- Le parole della scienza
- Le parole della statistica

Il linguaggio della statistica

- Il linguaggio della statistica usa una strategia di classificazione
 - Classi esaustive, mutuamente esclusive e organizzate gerarchicamente
- È una strategia di digitalizzazione in senso lato
 - Migliora il rapporto segnale/rumore
 - Minimizza gli errori
- L'operazione di codifica ha comunque un costo, il costo dell'astrazione
 - Si astrae dalla complessità implicita nel grande numero di variabili e attori del processo modellizzato, ma questo significa che c'è un livello di dettaglio in cui le ipotesi assunte sono false
 - «All models are wrong, but some are useful» [George Box]

Il paese con le percezioni più sbagliate

- Duffy, Bobby. *The Perils of Perception: Why We're Wrong About Nearly Everything*. London: Atlantic Books. 2018 [*I rischi della percezione: perché ci sbagliamo su quasi tutto*. Torino: Einaudi. 2019]
- «Su 100 persone in età lavorative quante, secondo lei, sono disoccupate e cercano lavoro?»
 - 49 (invece all'epoca della rilevazione erano 12)
 - Siamo, per questa domanda e per l'insieme delle domande della rilevazione Ipsos-MORI (30 quesiti, 13 Paesi, 50.000 interviste) il Paese con le percezioni più sbagliate,
- Ma sbagliate rispetto a che?
 - Educare al linguaggio tecnico
 - Mettersi dalla parte degli utenti

Educare al linguaggio tecnico

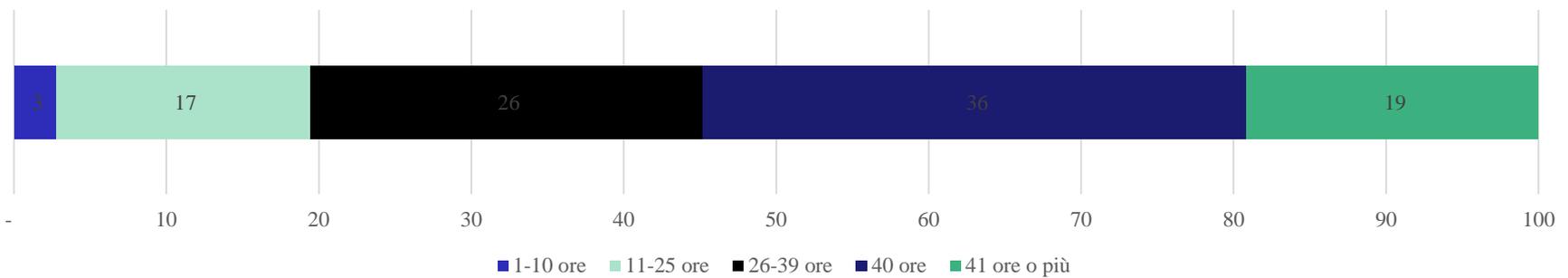
Come usa le parole la statistica: disoccupati

- In Italia i disoccupati sono circa 2,5 milioni
- Ma chi sono i disoccupati? Sono disoccupati (o in cerca di occupazione) le persone non occupate tra i 15 e i 74 anni che:
 - hanno effettuato almeno un'azione attiva di ricerca di lavoro nelle quattro settimane che precedono la settimana di riferimento
 - **e** sono disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive
 - **oppure**, inizieranno un lavoro entro tre mesi dalla settimana di riferimento
 - **e** sarebbero disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive, qualora fosse possibile anticipare l'inizio del lavoro
- Come verifico la disponibilità a lavorare nelle due settimane successive all'intervista?
- E poi, per individuare i disoccupati in questa accezione tecnica, ho necessità di distinguerli dagli **occupati**, da una parte, e dagli **inattivi**, cioè da coloro che non lavorano e non cercano lavoro, dall'altra

Educare al linguaggio tecnico

Come usa le parole la statistica: occupati

- In prima battuta: gli occupati sono quelli che lavorano; gli inattivi sono quelli che non lavorano e non intendono lavorare perché studenti, o pensionati, o impegnati in attività domestiche ('casalinghe')
- Quante ore alla settimana occorre lavorare per essere statisticamente occupati? «**almeno un'ora di lavoro** in una qualsiasi attività che preveda un **corrispettivo** monetario o in natura»
- Scandalo? Non proprio, perché oltre alla durata media della settimana lavorativa (su oltre 23 milioni di occupati, la maggioranza relativa, il 36%, lavora 40 ore la settimana), ci sono anche delle 'code' di occupati che lavorano molto di più o molto di meno: il 19% dalle 41 ore in su, ma soltanto il 3% dieci ore o meno

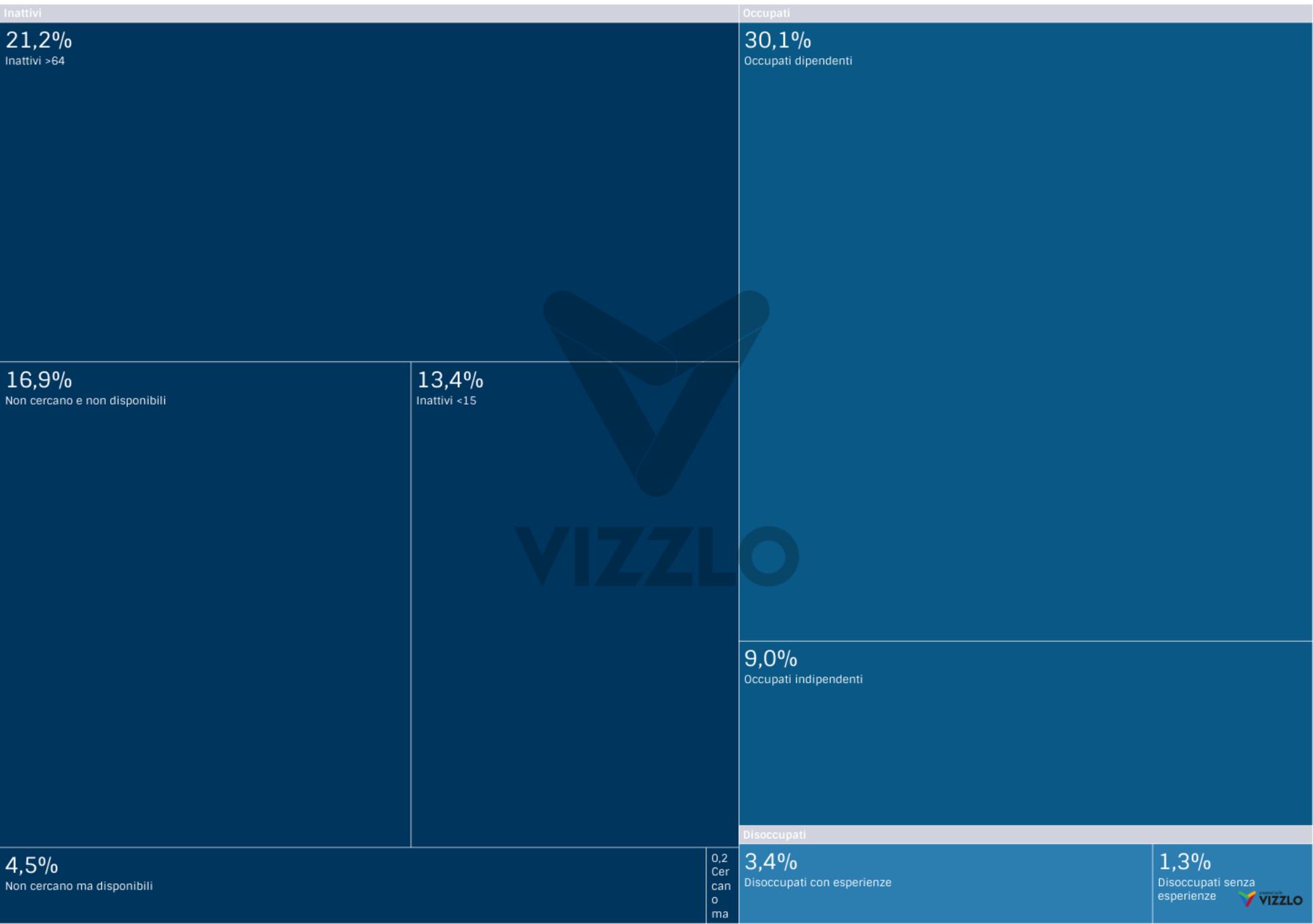


Educare al linguaggio tecnico

Come usa le parole la statistica: inattivi

- Almeno questo non dovrebbe creare problemi, perché sono **definiti in modo residuale**, come la parte della popolazione che non lavora (altrimenti sarebbero occupati) e non cerca un'occupazione (altrimenti sarebbe disoccupati)
- Coerenza con le altre definizioni, complicata ma fruttuosa:
 - non hanno cercato lavoro nelle ultime 4 settimane **e** non sono disponibili a lavorare entro 2 settimane dall'intervista (per lo più di studenti, ritirati dal lavoro e 'casalinghe')
 - pur **non** avendo **cercato** un lavoro nelle ultime 4 settimane si sono dichiarati **disponibili** a iniziare un lavoro entro 2 settimane
 - hanno **cercato** un lavoro nelle ultime 4 settimane, ma **non** sono **disponibili** a iniziare un lavoro entro 2 settimane
- **Forze di lavoro potenziali**
 - Confine fluido
 - La ricerca di lavoro ha un costo

Partecipazione al mercato del lavoro della popolazione residente



Mettersi dalla parte degli utenti: tradurre tra diversi meta-linguaggi

- Il problema è che gli utenti pongono domande alla statistica utilizzando il linguaggio naturale
 - Ma il linguaggio naturale non funziona come il linguaggio tecnico
 - Wittgenstein (*Ricerche filosofiche*) affronta il problema della meta-informazione parlando di «somiglianze di famiglia» (e di fibre e fili)
- Un problema di traduzione
 - Da una parte abbiamo il linguaggio della statistica (classi esaustive, mutuamente esclusive e organizzate gerarchicamente)
 - Dall'altro il linguaggio naturale (somiglianze di famiglie e fibre intrecciate)
- Chi lo deve fare?
 - Lo scarto tra linguaggio naturale e linguaggio tecnico è una **componente dello *statistical burden***
 - L'onere della traduzione deve ricadere sull'Istat

Verso un motore di ricerca semantico: le prospettive sono buone

- *Topic modelling*: scoprire strutture semantiche nascoste in un corpus di testo
 - Un esempio: “divari retributivi di genere” vs. “diseguaglianze salariali uomo/donna” → associazioni, declinazioni e sinonimie (divario/diseguaglianza/differenza, genere/maschio-femmina/uomo-donna, retribuzione/salario/stipendio)
- Obiettivi:
 - classificare i risultati della ricerca per rilevanza e dare suggerimenti sulle informazioni correlate
 - agevolare la traduzione dei termini della lingua corrente in termini specifici dei metadati statistici (lavorando sui due *corpus*)
- Tecniche:
 - LDA (Latent Dirichlet Allocation)
 - Word2Vec

Rifuggire dall'eccesso di semplificazione

- La semplificazione non è priva di rischi
 - Non parlo soltanto della semplificazione nei modelli (Einstein 1933: «Models should be as simple as possible, but not simpler»)
 - Parlo della semplificazione dei concetti che si spaccia per semplificazione del linguaggio: non ci sono spiegazioni semplici di fenomeni complessi
 - «nello studio della scienza tutto sta nel prendere su di sé la fatica del concetto (*die Anstrengung des Begriffs*)» [Hegel, *Fenomenologia dello spirito*, I, p. 48]
- **Chiedersi sempre:** Quale contributo all'informazione statistica, alla conoscenza e allo sviluppo della cultura e della coscienza civile dà il testo con cui commento i dati? In che modo contribuisce a rendere i servizi della statistica pubblica utili, rilevanti e affidabili per tutti?
- **Dedicare alla cura dei commenti e alla scelta delle parole la stessa maniacale attenzione che dedichiamo al controllo dei dati!**

Grazie!

Per domande e commenti:

gabarbieri@gmail.com