

## Introduzione

I numeri non hanno modo di parlare da soli, siamo noi a parlare per loro, li carichiamo di significato.

NATE SILVER, *Il segnale e il rumore*<sup>1</sup>.

### *A che cosa serve la statistica.*

Harold Shipman è stato condannato per il maggior numero di omicidi della Gran Bretagna, benché non corrispondesse allo stereotipo del serial killer. Tra il 1975 e il 1998 questo mite medico di famiglia della periferia di Manchester praticò ad almeno 215 pazienti anziani un'iniezione con una forte overdose di oppiacei. Fece infine l'errore di falsificare il testamento di una delle vittime, inserendo un lascito di denaro a proprio favore: la figlia della paziente era avvocata, insorsero sospetti, poi un'ispezione del suo computer rilevò cartelle cliniche modificate a posteriori, per aggravare lo stato di salute delle vittime. Shipman, noto fan delle nuove tecnologie fin dai primordi, non si era però reso conto che a ogni modifica era associato un orario (fra l'altro, questo è un buon esempio di dati che permettono rivelazioni inaspettate).

Furono riesumati 15 pazienti che non avevano subito la cremazione: l'autopsia constatò livelli letali di diamorfina, la forma dell'eroina per uso medico. Nel 1999 Shipman fu processato con l'accusa di 15 omicidi, ma scelse di non difendersi e in tribunale non disse mai nulla. Fu condannato all'ergastolo; si avviò poi un'inchiesta per appurare se avesse commesso altri crimini, e se lo si sarebbe potuto smasche-

<sup>1</sup> N. SILVER, *The Signal and the Noise*, Penguin, New York 2012, p. 9; trad. it. di M. Giffone, *Il segnale e il rumore*, Fandango Libri, Roma 2013, p. 116. Questo libro è un'ottima introduzione all'uso della statistica per realizzare previsioni nello sport e in altri ambiti.

rare prima. Fui uno degli statistici chiamati a testimoniare; l'inchiesta identificò 215 vittime certe di Shipman, e altre 45 vittime possibili<sup>2</sup>.

Questo libro si concentrerà sull'uso della **statistica**<sup>3</sup> per rispondere al tipo di interrogativi che insorgono quando vogliamo capire meglio il mondo; alcune domande saranno evidenziate in un riquadro. Per analizzare il comportamento di Shipman, è naturale chiedersi per prima cosa:

Che tipo di persone ha ucciso Harold Shipman, e quando sono morte?

Per ciascuna vittima, l'inchiesta ha fornito dettagli sull'età, il genere e la data di morte. La figura 0.1 è una visualizzazione piuttosto sofisticata di questi dati: il grafico di dispersione mostra l'età della vittima in funzione della data di morte, e il colore dei punti indica se si trattava di un uomo o di una donna. Sugli assi sono riportati gli anni solari e grafici a barre che mostrano l'andamento delle età (suddivise in intervalli di cinque anni) e degli anni.

Basta guardare un attimo la figura per trarre alcune conclusioni. Poiché i punti neri sono più numerosi di quelli grigi, le vittime di Shipman erano soprattutto donne. Il grafico a barre a destra della figura mostra che le vittime erano per lo più settantenni e ottantenni, ma la distribuzione dei punti rivela che, benché all'inizio fossero tutte anziane, con il passare degli anni si manifestarono casi più giovani. Il grafico a barre in cima mostra chiaramente un intervallo senza omicidi attorno al 1992. Si scoprì che prima di quell'anno Shipman lavorava in uno studio collettivo con altri medici ma poi, forse perché si sentiva sospettato, lo lasciò per aprirne uno in

<sup>2</sup> I dati di Shipman sono analizzati in dettaglio in D. SPIEGELHALTER e N. BEST, *Shipman's Statistical Legacy*, in «Significance», I (2004), n. 1, pp. 10-12. Tutti i documenti dell'inchiesta pubblica sono disponibili all'indirizzo <http://webarchive.nationalarchives.gov.uk/20090808155110/http://www.the-shipman-inquiry.org.uk/reports.asp>.

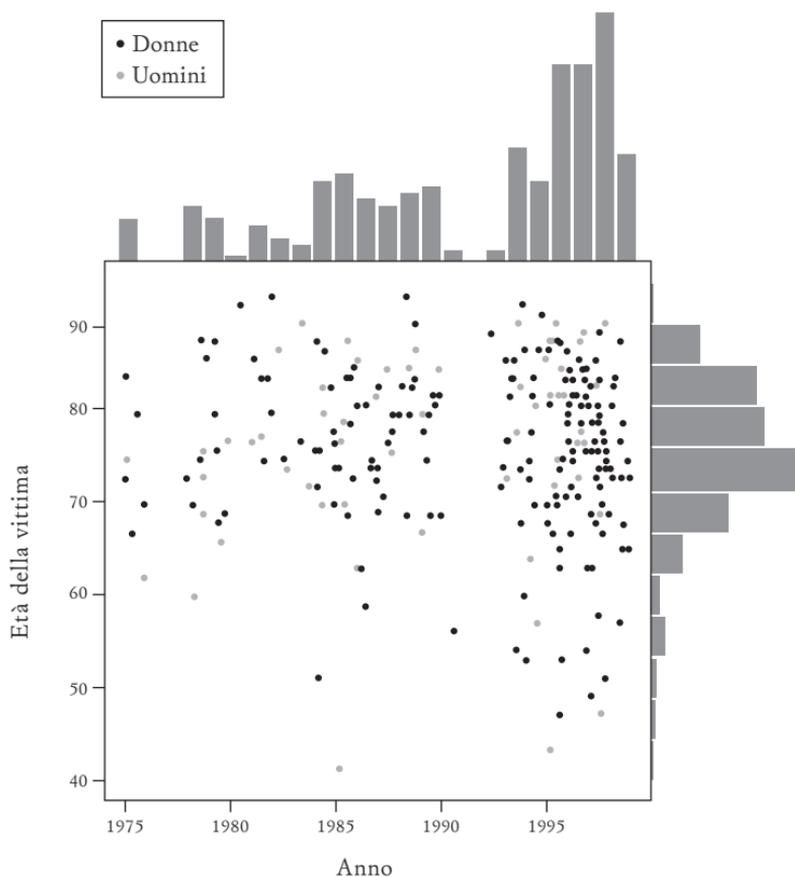
<sup>3</sup> I termini in **grassetto** appaiono nel *Glossario* alla fine del libro, che fornisce definizioni sia elementari sia tecniche.

proprio. In seguito le sue attività accelerarono, come mostra il grafico a barre in cima.

Questa analisi delle vittime identificate dall'inchiesta solleva altre domande sulla meccanica degli omicidi. Alcuni indizi statistici sono forniti dall'orario in cui morirono le sue presunte vittime, reperibile sul certificato di morte. La figura 0.2 è un grafico a linee che confronta gli orari di decesso dei pazienti di Shipman e quelli di un campione di pazienti

Figura 0.1.

Grafico di dispersione che mostra l'età e l'anno di morte delle 215 vittime confermate di Harold Shipman. Agli assi sono aggiunti grafici a barre che mostrano l'andamento delle età e degli anni in cui ha commesso gli omicidi.

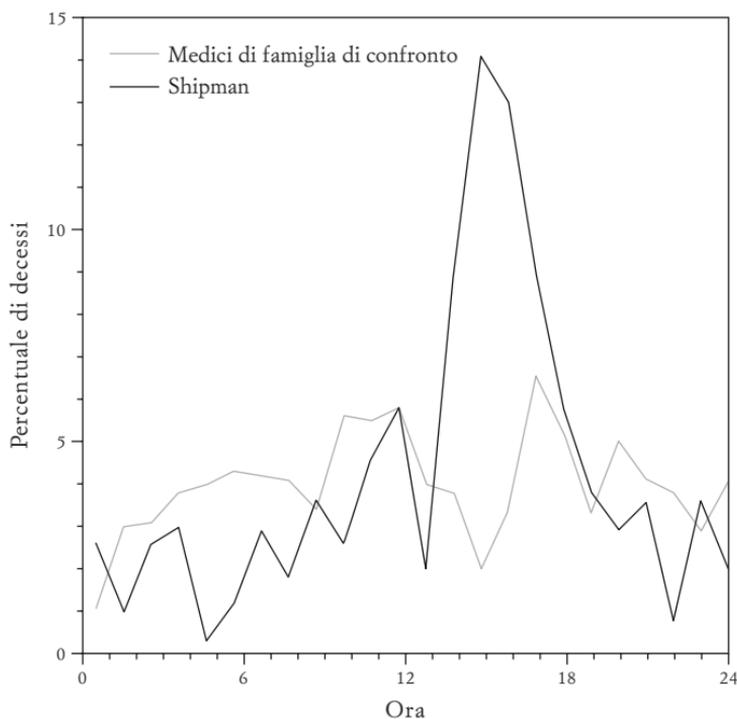


di altri medici di famiglia della zona. L'andamento non richiede un'analisi molto approfondita: questo genere di conclusione è a volte detto «interoculare», perché salta agli occhi. I pazienti di Shipman tendevano quasi tutti a morire nel primo pomeriggio.

I dati non possono dirci *perché* tendessero a morire a quell'ora, ma altre indagini rivelarono che Shipman svolgeva le visite domiciliari dopo pranzo, quando in genere era solo con i pazienti anziani. Proponeva loro un'iniezione che, a suo dire, li avrebbe fatti sentire meglio, ma che in realtà era una dose letale di diamorfina: dopo che la vittima gli era morta serenamente davanti, ne modificava la cartella clinica perché il de-

Figura 0.2.

L'ora del decesso dei pazienti di Harold Shipman, confrontata con quelli di altri medici di famiglia della zona. L'andamento non richiede un'analisi statistica sofisticata.



cesso sembrasse naturale e prevedibile. Come ha detto tempo dopo Dame Janet Smith, comandante dell'Ordine dell'Impero Britannico che ha presieduto l'inchiesta: «Mi sembra ancora spaventoso oltre ogni dire, proprio inesprimibile e impensabile e inimmaginabile che giorno dopo giorno, fingendosi un medico di grande empatia, portava nella borsa la sua arma letale [...] che tirava fuori senza pensarci due volte».

Shipman correva qualche rischio, perché una sola autopsia sarebbe bastata a smascherarlo, ma non ne fu svolta nessuna, vista l'età dei pazienti e le cause di morte in apparenza naturali. Il movente di questi omicidi non fu mai spiegato: Shipman non testimoniò al processo, non parlò mai dei suoi crimini con nessuno, neanche con la famiglia, e si suicidò in carcere, appena in tempo perché la moglie percepisse la pensione.

Studi iterativi e investigativi del genere possono essere considerati statistica «forense»; in questo caso è stato proprio così. Non si parla di teoria né di matematica, ci si limita a cercare le regolarità che potrebbero sollevare domande più interessanti. I dettagli dei crimini di Shipman furono determinati usando indizi specifici per ciascun caso, ma l'analisi dei dati permise di capire in generale come attuava i suoi intenti omicidi.

Più avanti nel libro, nel capitolo x, vedremo se un'analisi statistica formale avrebbe permesso di incolpare Shipman in anticipo<sup>4</sup>. Nel frattempo, la storia di Shipman dimostra ampiamente la grande potenzialità dei dati per aiutarci a capire il mondo e formulare valutazioni migliori. Questa è l'essenza della statistica.