

Statistica I

Introduzione al corso

Tommaso Rigon

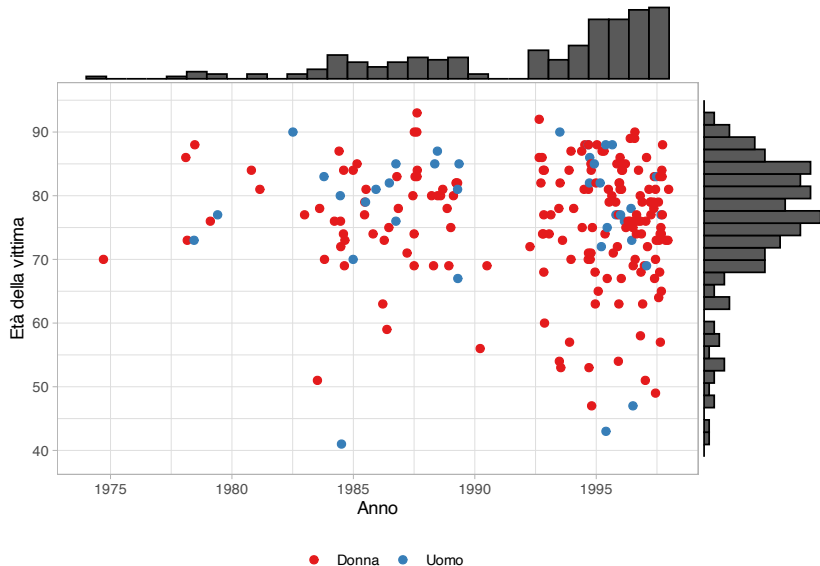
Università Milano-Bicocca



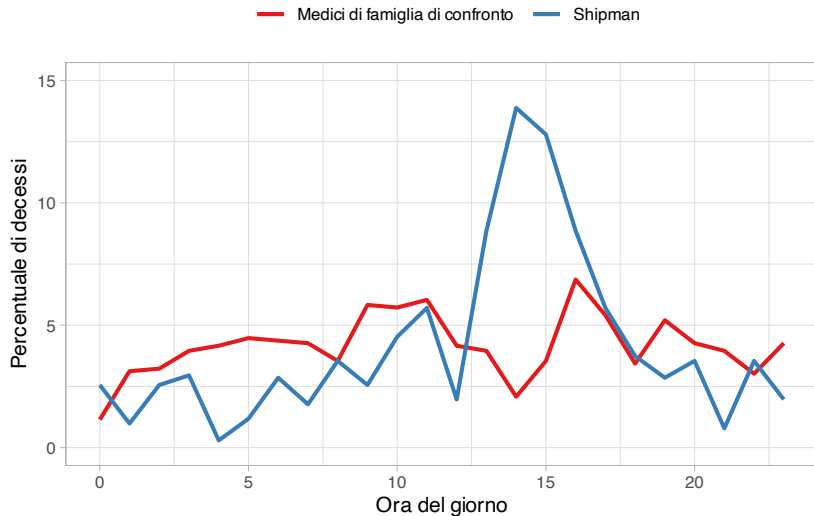
La storia di Harold Shipman

- Harold Shipman è stato un **medico di famiglia** della periferia di Manchester, nonché il **serial killer** più prolifico della storia del Regno Unito.
- Si stima che tra il 1975 ed il 1998 Shipman abbia ucciso almeno 215 pazienti anziani tramite un'iniezione con una forte **overdose di oppiacei** (diamorfina).
- Shipman fu **condannato all'ergastolo** per l'omicidio di 15 persone nel 1999. Non testimoniò al processo, non spiegò mai il movente. Morì suicida in carcere nel 2004.
- Un'indagine successiva (Baker, 2001), basata anche su metodi statistici, concluse che Shipman molto probabilmente uccise più di 215 pazienti.
- È naturale quindi chiedersi per prima cosa: **chi erano le vittime** di Harold Shipman? **Poteva essere fermato prima?**

Età ed anno di morte delle 215 vittime confermate



Orario del decesso



Statistica: cenni storici

- Sebbene oggi abbia un significato molto diverso, la parola **statistica** deriva dal vocabolo italiano **Stato**.
- **Ghislini** nel 1589 indica la statistica come *“descrizione delle qualità che caratterizzano e degli elementi che compongono uno Stato”*.
- La **collezione di dati** circa la popolazione e l'economia cominciò nelle città di Venezia e Firenze durante il Rinascimento. Venivano ad esempio registrate le **nascite**, i **decessi**, i **matrimoni**, soprattutto in occasioni delle pestilenze.
- Ancora oggi alcuni enti (ad esempio ISTAT ed EUROSTAT), intendono la statistica come “disciplina dello Stato”.
- Nel linguaggio moderno, chiameremmo questa disciplina “**statistica ufficiale**” oppure anche “**statistica economica**”.

Verso la statistica moderna

- Il ruolo della statistica cambiò notevolmente a partire dalla fine del 1800, anche grazie allo sviluppo del **calcolo delle probabilità** e del **gioco d'azzardo**.
- La **Royal Statistical Society** (RSS) venne fondata nel 1834, mentre l'**American Statistical Association** (ASA) nel 1839. Il primo dipartimento di statistica al mondo venne fondato nel 1911 all'University College di Londra (UCL).
- Nella prima metà del '900 avvenne la **prima rivoluzione** della statistica. Furono gettate le fondamenta della teoria della statistica moderna.
- A partire dagli anni '70 e '80 del secolo scorso è cominciata una **seconda rivoluzione**, grazie all'immensa potenza di calcolo e disponibilità di dati derivanti dallo **sviluppo informatico**.
- Questa fase rivoluzionaria è ancora in atto. Nel contesto moderno, la statistica viene anche chiamata **data science**.

La percezione della statistica

- *“Ci sono tre tipologie di menzogne: le bugie, le maledette bugie e le statistiche”.* Pronunciata da **Mark Twain**.
- *“Come i sogni, le statistiche sono una forma di appagamento”.* **Jean Baudrillard**.
- A volte purtroppo la statistica è accolta con sospetto.
- La statistica e i dati, di per sé, non mentono mai. Chi mente sono due categorie di statistici: quelli **disonesti** e quelli **incompetenti**.
- Sfortunatamente, anche quando un bravo statistico svolge correttamente il suo lavoro, questo potrebbe venire **percepito** come “falso” se mancano le competenze quantitative di base.

La statistica, Trilussa (Carlo Alberto Salustri)

*Sai ched'è la statistica? È na' cosa
che serve pe fà un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che spósa.*

*Ma pè me la statistica curiosa
è dove c'entra la percentuale,
pè via che, lì, la media è sempre eguale
puro co' la persona bisognosa.*

*Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:*

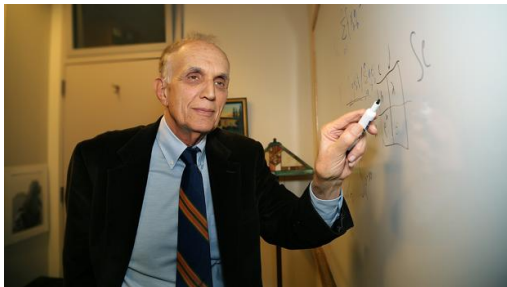
*e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perch'è c'è un antro che ne magna due.*

La statistica nella cultura popolare



- La figura qui sopra illustra un esempio di **statistico incompetente**.
- La poesia di Trilussa evidenzia invece lo scetticismo di chi è sprovvisto delle nozioni statistiche di base.

In risposta alle precedenti provocazioni



- *"E' semplice mentire con la statistica, ma è molto più semplice mentire senza di essa".* **Frederick Mosteller.**
- *"La statistica è la grammatica della scienza".* **Karl Pearson.**
- *"Coloro che non conoscono la statistica sono condannati a reinventarla".* **Bradley Efron.**

La definizione di statistica

- Nel tempo, la statistica ha cambiato sembianze e nome: oggi viene a volte chiamata **Data Science**.
- La statistica non è più la “scienza dello Stato”. Non esiste una definizione univocamente accettata, ma le varie proposte hanno spesso un ampio respiro.

Alcune definizioni

- “La statistica riguarda **ciò che si può imparare dai dati**”. **Anthony Davison**.
- “La statistica è essenzialmente lo **studio dell'incertezza**. Il ruolo degli statistici è di **assistere** i lavoratori di altri campi, o i propri clienti, i quali incontrano l'incertezza nel loro lavoro.” **Dennis Lindley**.
- “La statistica è la tecnologia necessaria per **estrarre significato dai dati** e per **gestire l'incertezza**”. **David Hand**.
- “La statistica è la **scienza delle decisioni** in condizioni di incertezza”. **Domenico Piccolo**.

Il lavoro dello statistico

- *“L'aspetto migliore dell'essere uno statistico è che puoi giocare nel giardino di chiunque”.* John Tukey.
- La statistica viene spesso chiamata una **scienza ancillare**, perché è presente trasversalmente in quasi tutte le discipline scientifiche.
- Di conseguenza, lo statistico troverà **lavoro** e potrà giocare in tanti **diversi giardini**:
 - **Imprese ed industria**
 - Ricerche di mercato
 - Controllo di qualità
 - Marketing quantitativo
 - **Finanza ed assicurazioni**
 - Risk management
 - Analisi dei mercati finanziari
 - **Pubblica Amministrazione**
 - Studi sulla popolazione, demografia
 - Statistiche ufficiali
 - **...e tanti altri settori**
 - Medicina & epidemiologia
 - Psicologia
 - Ecologia

Il corso di laurea - Percorso comune

- Il primo anno di studi fornisce gli **strumenti** essenziali per poter affrontare il corso.
- Il secondo e terzo anno costituiscono il cuore degli studi statistici ed economici.

■ Primo anno

- Algebra lineare
- Analisi Matematica
- Calcolo delle probabilità
- Microeconomia
- Statistica I
- Statistica economica I
- Laboratorio di informatica
- Informatica
- Lingua straniera

■ Secondo e terzo anno

- Analisi Matematica II
- Econometria
- Macroeconomia
- Serie storiche economiche
- Analisi statistica multivariata
- Statistica II
- Statistica III
- Statistica economica II
- Basi di dati

- ...e poi? È possibile scegliere tra **tre diversi percorsi**.

Il corso di laurea - Tre diversi percorsi

- Oltre agli esami comuni, lo studente può **personalizzare** il proprio piano di studi scegliendo uno dei seguenti **percorsi**:

■ Mercati Finanziari

- Economia dei mercati monetari e finanziari
- Matematica finanziaria
- Finanza Aziendale

■ Sistemi informativi aziendali

- Organizzazione dei sistemi informativi
- Data Mining
- Statistica Computazionale

■ Ricerche di mercato

- Analisi di mercato
- Statistica Aziendale
- Economia Industriale



Statistica descrittiva ed inferenziale

- La statistica viene spesso suddivisa in due macro-argomenti.

Statistica descrittiva (Statistica I)

- **Descrive** e **riassume** la realtà.
 - Creazione di indici, tra cui: media aritmetica, mediana, moda, varianza, etc.
 - Visualizzazione dei dati tramite ad esempio istogrammi, boxplot, etc.
- **Quantifica** l'incertezza tramite strumenti algebrici, geometrici e grafici.
- **Esplora** i dati.

Statistica inferenziale (Statistica II)

- **Quantifica** l'incertezza tramite il **calcolo delle probabilità**.
- **Stima** le quantità ignote di un modello statistico.
- **Prevede** fenomeni futuri fornendo margini di incertezza.

Contenuto del corso

- Statistica I è un corso di **statistica descrittiva**. Non faremo quindi (quasi) nessuna menzione alla probabilità.
- La statistica descrittiva e quella inferenziale **non sono argomenti separati**. Molti concetti sono comuni, sebbene vengano a volte presentati in modi diversi.
- Durante il corso i concetti teorici di statistica descrittiva verranno sempre accompagnati da un'**applicazione**.
- I contesti applicativi saranno molto diversi tra loro, a sottolineare il fatto che la statistica sia una **scienza trasversale**.

Il ruolo dello statistico

- Ogni problema ha caratteristiche peculiari, con conoscenze pregresse diverse. In ogni problema statistico ci sono **scelte** da fare.
- Uno **statistico** è quindi come un **vecchio artigiano**: tutti i pezzi sono unici, i risultati dipendono dalla sua capacità. Successi ed insuccessi sono frutto delle sue scelte.
- **Non esistono ricette** e procedure da applicare in automatico. Guardate ai contenuti dei corsi (di Statistica) solo come ad esempi di quello che si può fare.
- Del resto, perché dovrebbero esistere delle lauree in Statistica se gli statistici fossero sostituibili con delle macchine?
- Questo corso mostra idee, tecniche e strategie che si sono rivelate utili in una gamma di problemi reali. Ma i problemi che voi affronterete in futuro non saranno gli stessi!

Docente

- **Nome:** Tommaso Rigon
- **Email:** `tommaso.rigon@unimib.it`
- **Pagina personale:** <https://tommasorigon.github.io>
- **Ricevimento:** Martedì alle 17.30. Si prega di contattare il docente in anticipo.

Organizzazione del corso

- **Pagine e-learning:** <https://elearning.unimib.it/course/view.php?id=43448>
- **Pagina web del corso:** <https://tommasorigon.github.io/StatI/>
- Nella pagina web del corso potete trovare:
 - Ultimi avvisi
 - Materiale didattico da scaricare (slide, esercizi, etc.)
 - Modalità d'esame
 - Altro

Testo di riferimento

- Piccolo, D. (2010), *Statistica*, Terza edizione, Il Mulino.
- Cicchitelli, G., D'Urso P. e Minozzo, M. (2022). *Statistica. Principi e metodi*, Quarta edizione, Pearson.



Testi di consultazione

- Spiegelhalter, D. (2020). *L'arte della statistica*. Einaudi
- Piccolo, D. (2020). *Statistica per le decisioni*, Terza edizione, Il Mulino.