



Introduzione ai metodi statistici per le applicazioni industriali parte 1

Antonio Panico

Department of Engineering for Industrial Systems and Technologies
University of Parma

18-06-2025

Tutto il materiale della lezione è su GitHub

Slides, notebook ed esercizi sono consultabili al seguente link:

<https://github.com/APanico12/Lezione-PCT0>

Scansiona il QR code per accedere direttamente:



Dottorando in Ingegneria Industriale Università degli Studi di Parma

Antonio Panico



antonio.panico@unipr.it

Aree di Ricerca:

- Previsione dei prezzi dell'energia in Europa (ENTSO-E)
- Stima dei Marginal Emission Factors (MEFs) con approcci robusti
- Integrazione di Machine Learning e statistica industriale

Pubblicazioni principali:

- *"Interconnectedness in the Tails: Risk Contagion in European Electricity Markets"*
Premio **ENBIS Knowledge Fund** (2025)
- *Multi objective optimization of FDM 3D printing parameters set via design of experiments and machine learning algorithms*
Scientific Reports, 2025
doi.org/10.1038/s41598-025-01016-z

Cos'è la Statistica

- La statistica studia i dati: come raccoglierli, analizzarli e interpretarli.
- L'analisi dei dati dipende:
 - dai dati stessi
 - dalla conoscenza del problema
 - dalle ipotesi usate nel modello matematico
- Due approcci principali:
 - **Statistica descrittiva**: analisi e rappresentazione dei dati senza fare ipotesi esterne.
 - **Inferenza statistica**: uso della probabilità per fare stime e previsioni.
- *Nota*: Non tratteremo la statistica bayesiana, ma è importante nelle applicazioni.

Dove si applica la Statistica?

- **Disegno degli esperimenti:** progettare prove scientifiche controllate (es. trial clinici, test industriali).
- **Analisi di sopravvivenza:** stimare il tempo fino a un evento (guasto, morte, abbandono).
- **Controllo di qualità:** monitorare processi produttivi con carte di controllo.
- **Machine learning:** modelli statistici che imparano dai dati (classificazione, regressione).
- **PCA (Principal Component Analysis):** riduzione della dimensionalità per esplorare dati complessi.
- **Economia e finanza:** previsioni, analisi di rischio, studi di mercato.

Ovunque ci siano dati, la statistica è fondamentale.

L'essenza della statistica moderna

- **Origini solide:** nasce nel XVII secolo con Pascal e Fermat e la teoria della probabilità, per problemi reali (es. pensioni, astronomia).
- **Evoluzione storica:**
 - XIX secolo: collegamento con medicina, biologia e genetica.
 - XX secolo: approccio matematico e tecnico, vista spesso come “borsa di strumenti”.
- **Sfide attuali:**
 - **Big Data:** grandi quantità di dati da sensori, social media, ecc. → servono competenze in programmazione, gestione dati, conoscenza del dominio.
 - **Crisi di riproducibilità:** molte ricerche non replicabili → attenzione al disegno, all'analisi e all'interpretazione.
- **Conclusione:** *più dati \neq più verità*. Serve senso critico, rigore e consapevolezza per distinguere segnali da rumore.

Cos'è un Dato? Tipi di Dati

- **Dato:** un'informazione osservabile o misurabile su un oggetto o fenomeno.
- **Tipi di dati:**
 - **Qualitativi** (categorici): es. colore degli occhi, tipo di auto.
 - **Quantitativi** (numerici):
 - *Discreti* (es. numero di figli)
 - *Continui* (es. altezza in cm)
- La tipologia dei dati influenza il tipo di analisi statistica da applicare.



Quando un dato ha senso?

- **Ha un contesto:** un numero da solo non basta. Serve sapere dove, quando, su chi, con quale unità.
- **È stato raccolto correttamente:** un dato sbagliato o incompleto può fuorviare.
- **Risponde a una domanda reale:** i dati devono servire a capire o decidere qualcosa.
- **È interpretabile:** ha senso solo se possiamo confrontarlo con altri (media, storia, standard).
- **Rispetta la logica del fenomeno:** attenzione a valori assurdi o contraddittori.

"Un dato ha senso solo se ha contesto, qualità e scopo."



Il dato deve essere oggettivo

- **Domanda:** Quanti alberi ci sono sul pianeta?
- Prima di contare, serve una definizione precisa: *cos'è un albero?*
- Un albero è definito come una pianta con **fusto legnoso** e **diametro a petto d'uomo (DBH) ≥ 10 cm**.
- Senza una definizione condivisa:
 - Ogni osservatore conterebbe cose diverse (arbusti, palme, cespugli. . .).
 - I dati sarebbero incoerenti, e le stime inutilizzabili.
- I ricercatori hanno:
 - Campionato alberi in diversi biomi
 - Stimato densità media per km^2
 - Usato immagini satellitari e modelli per estrapolare il totale
- **Risultato:** circa **3.04 trilioni** di alberi sulla Terra.

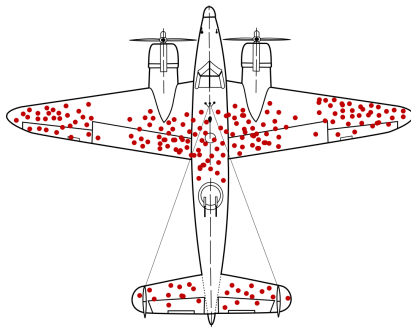
Definire bene un concetto è il primo passo per misurarlo con senso.

Attenzione: i dati possono ingannare

- I dati osservati non sempre raccontano tutta la verità.
- Le conclusioni possono essere sbagliate se ignoriamo i dati "nascosti".

Attenzione: i dati possono ingannare

- I dati osservati non sempre raccontano tutta la verità.
- Le conclusioni possono essere sbagliate se ignoriamo i dati "nascosti".



Domanda: quali parti dell'aereo rafforzeresti?

Lo scopo della Statistica: dare senso ai dati

- I **dati** sono solo numeri, raccolti da osservazioni o misurazioni.
- Ma senza analisi, i **dati non parlano**. Sono solo "rumore grezzo".
- La **statistica** dà significato ai dati:
 - **Li descrive** per capire cosa sta succedendo.
 - **Li interpreta** per fare inferenze su fenomeni più ampi.
 - **Li trasforma** in informazione utile per decidere.
- **Obiettivo finale:** ridurre l'incertezza e guidare azioni basate su evidenze.

"Dati + Statistica = Conoscenza utile."



EDA – Exploratory Data Analysis

Cos'è l'EDA?

- È la fase iniziale dell'analisi statistica.
- Serve a capire la struttura, la qualità e i pattern presenti nei dati.
- Aiuta a formulare ipotesi e scegliere modelli adeguati.

Cosa si fa durante l'EDA?

- Controllo dei dati mancanti o anomali
- Calcolo di misure sintetiche: media, mediana, varianza...
- Visualizzazioni: istogrammi, boxplot, scatter plot, heatmap
- Individuazione di relazioni tra variabili

“L'EDA è come conoscere i tuoi dati prima di farti domande complesse”

William Bratton: un caso di Strategia Vincente...



"Non più risorse, ma visione e dati intelligenti."

Contesto (anni '90):

- Criminalità elevata a New York City.
- Polizia percepita come inefficace e distaccata.

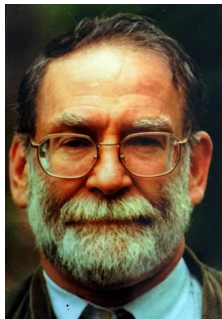
Mossa strategica:

- Introduzione del sistema **CompStat**.
- Uso sistematico di dati per decisioni operative.

Risultati:

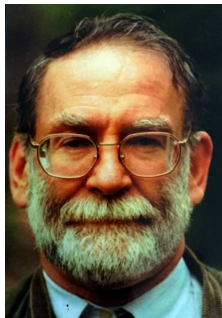
- -60% di criminalità in pochi anni.
- Rinnovata fiducia dei cittadini.
- Modello replicato a livello nazionale.

Il caso Harold Shipman [1]



Harold Shipman (1946–2004)

Il caso Harold Shipman [1]



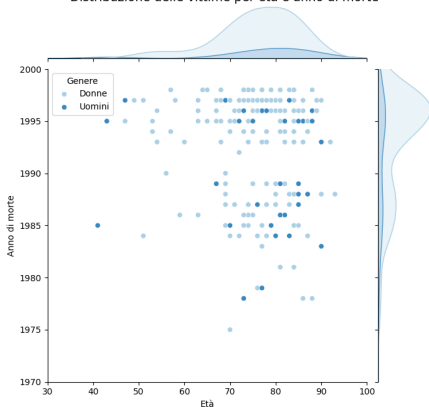
Harold Shipman (1946–2004)

Dati principali

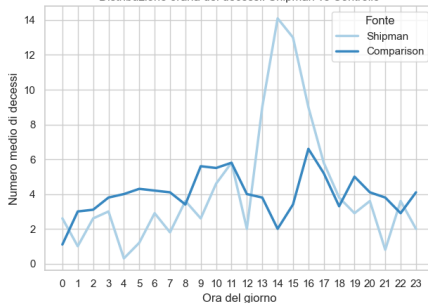
- **Scoperta:** falsificazione testamento
- **Periodo attivo:** 1975 – 1998
- **Vittime accertate:** 215 (fino a 260 sospette)
- **Profilo vittime:** donne, età 70–80 anni
- **Metodo:** overdose di diamorfina
- **variabili usate:** età, sesso, data di morte, ora decesso

Il caso Harold Shipman

Distribuzione delle vittime per età e anno di morte



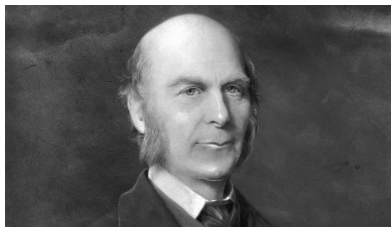
Distribuzione oraria dei decessi: Shipman vs Controllo



Francis Galton e la Saggezza della Folla (1907)

L'esperimento

- Galton assiste a una fiera del bestiame a Plymouth.
- 787 persone pagano per indovinare il peso di un bue macellato.
- Calcola la **mediana** delle stime: **1.207 libbre**.
- Il peso reale? **1.198 libbre**.



Francis Galton (1822–1911)

Conclusione

- Il risultato collettivo fu sorprendentemente accurato.
- Nasce il concetto di **saggezza della folla**.

Reference

- [1] David Spiegelhalter. *The Art of Statistics: How to Learn from Data*. Basic Books, 2019. DOI: [10.5038/1936-4660.13.1](https://doi.org/10.5038/1936-4660.13.1). URL: <https://doi.org/10.5038/1936-4660.13.1>.