

Quale ruolo della metrologia nell'era dei big data?

Dario Petri
Dipartimento di Ingegneria Industriale
Università di Trento



enormi quantità di **dati** estratti da:

- mondo reale (**world datafication**)
- internet, smart phones, dispositivi ICT

principale motivazione:

- **supporto alle decisioni** (data driven decision making **D³M**)
- identificazione di schemi (**patterns**)

BD è una delle tecnologie più dirompenti (**disruptive**):
sta cambiando il modo in cui viviamo, lavoriamo, pensiamo

BD: caratteristiche

volume

quantità

variety

eterogeneità
diverse forme
non strutturati

velocity

di generazione
ed elaborazione

variability

diversi tipi
sorgenti, contesti,
applicazioni



informazione o rumore ?

data quality: problema enorme

smart data: componente utile
estratta dai dati

difficile estrarre informazione

spesso usate **tecniche di I.A. (analytics)**

Smart Data: caratteristiche

VALIDITY

utilità per lo
scopo prefissato

VERACITY

incertezza

garantite applicando principi, metodi e strumenti della:

metrologia = una scienza di **data quality**

- identificazione sorgenti di **incertezza**
- affidabilità **dati** e **informazione** estratta
- valutazione e gestione **effetti incertezza** e del **rischio** di decisione errata

Una nuova ideologia: Dataismo

Dataista (forma estrema):

- percepisce l'universo come un **flusso di dati**
- decide basandosi **solo** sui **dati**
- confida nella **IA** più che nell'intelligenza umana
- vede gli organismi viventi come **sistemi di elaborazione biochimici**



Data Science (Santo Graal): scienza del tutto, unifica tutte le discipline

Ex: la quinta sinfonia di Beethoven, una bolla finanziaria, l'evoluzione di un'epidemia analizzabili con gli stessi principi, metodi e strumenti

LATO OSCURO:
via facile e veloce (seducente) per raggiungere lo scopo

Una dicotomia: la post-truth

Post-truth era: fatti subordinati a opinioni
razionalità subordinata a emotività

- **fatti** selezionati e interpretati in modo da favorire la conclusione voluta
- **fonti autorevoli denigrate**
- **negazione di evidenze scientifiche** (fumo, evoluzione, vaccini, clima, terra piatta, ...)
- **fake news**, ...
- sviluppo di **teorie del complotto**
- emotività usata per indirizzare l'**opinione pubblica**



cause: pregiudizi cognitivi e di conferma, conformità sociale, negazione di una verità oggettiva, declino dei media tradizionali a favore dei social media, ...

DM: Anatomia

Obiettivo DM: scegliere l'**alternativa migliore** in un certo **contesto**:

- in base all'**informazione** (utile e affidabile) disponibile
- seguendo una **procedura**
- soddisfacendo **vincoli** (tempo, risorse, ...)



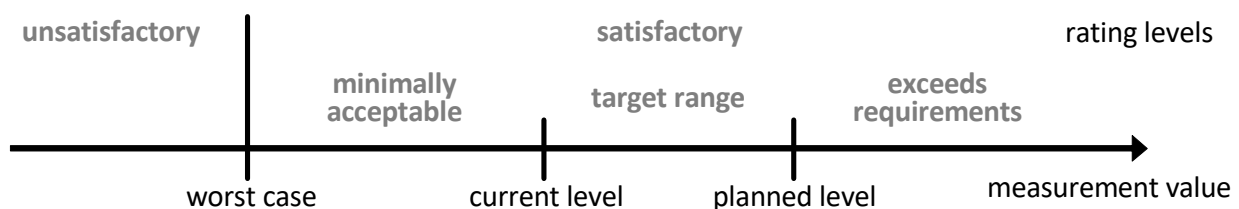
livello di fiducia della scelta è limitato da **sorgenti di incertezza**

livello di fiducia accettabile dipende da applicazione e da propensione al **rischio**

La più semplice D³M (formalizzabile)

- **scelta univoca:** esiste un'unica opzione ottimale
- **struttura completa:** la procedura è ben definita
- **informazione completa:** sufficiente per selezionare (con un dato livello di fiducia) l'opzione ottimale

Ex.: **verifica di conformità:** confronto tra risultati di misura e limiti di specifica per decidere se il misurando soddisfa i requisiti



Decisioni complesse

procedure automatiche

Procedure automatiche spesso usate quando:

- **opzione ottimale non chiara**
- **struttura della decisione non ben definita**
- **informazione è parziale (indicatori)** e acquisita con procedure non validate



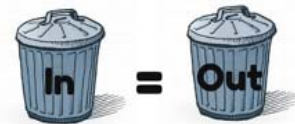
⇒ **livello di fiducia?** raramente valutabile/fornito

principio GIGO (*garbage in – garbage out*):

informazione di bassa qualità conduce a decisioni errate



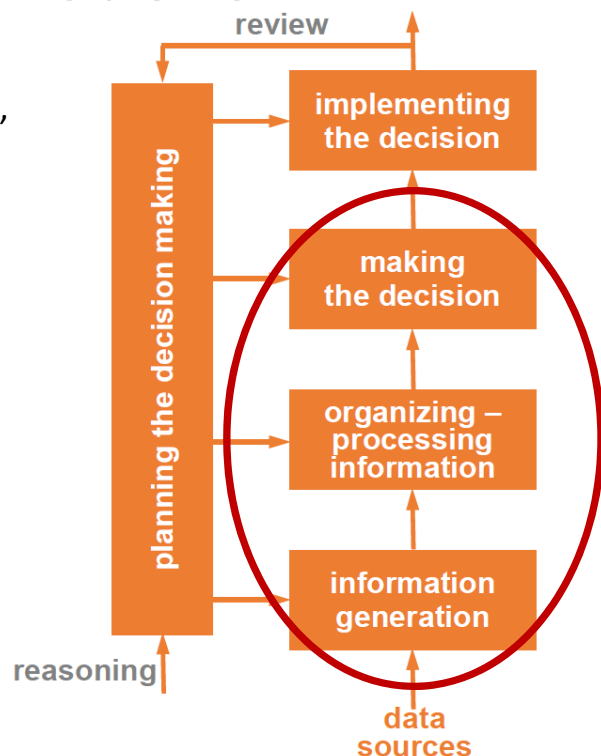
valutazione soggettiva è di norma necessaria



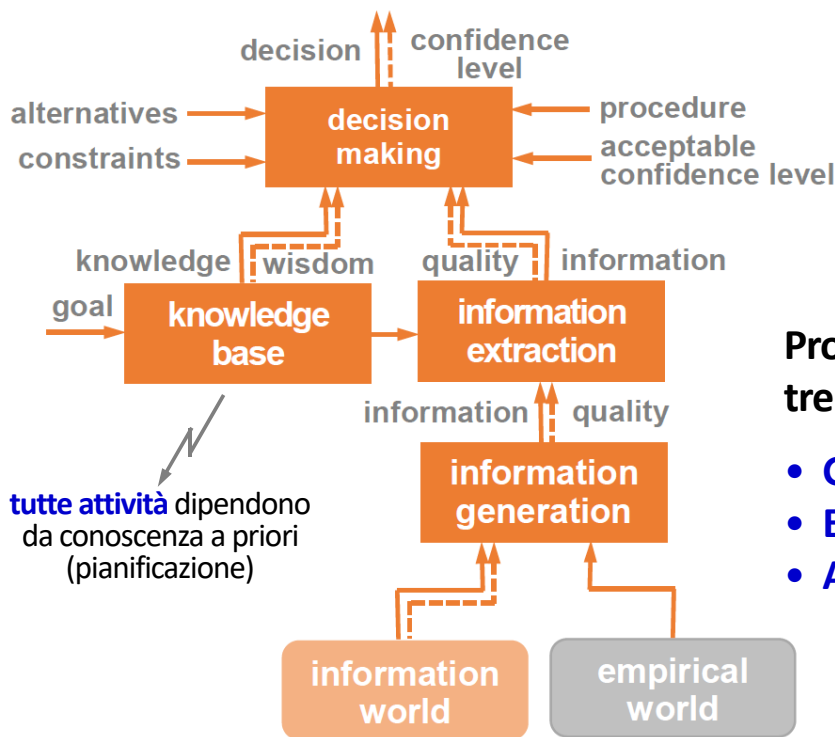
D³M razionale: un modello

Pianificazione basata su:

- **conoscenze** esplicite o tacite (sentimenti, emozioni, ...)
- **scelte soggettive** (micro-decisioni coscienti o intuitive)
- **valori**
- **mentalità**
- **esperienza** (decisioni precedenti)
- **definizione problema e obiettivi**
- **identificazione vincoli**
- **identificazione contesto e stakeholder**
- **Modello della decisione:**
 - Identificazione alternative
 - Analisi conseguenze e rischi
 - Definizione livello di fiducia
 - Definizione procedura
 - Selezione grandezze rilevanti
- **allocazione risorse**



D³M: un modello metrologico



Processo strutturato in tre livelli gerarchici:

- Generazione informazione
- Estrazione informazione
- Attività decisionale

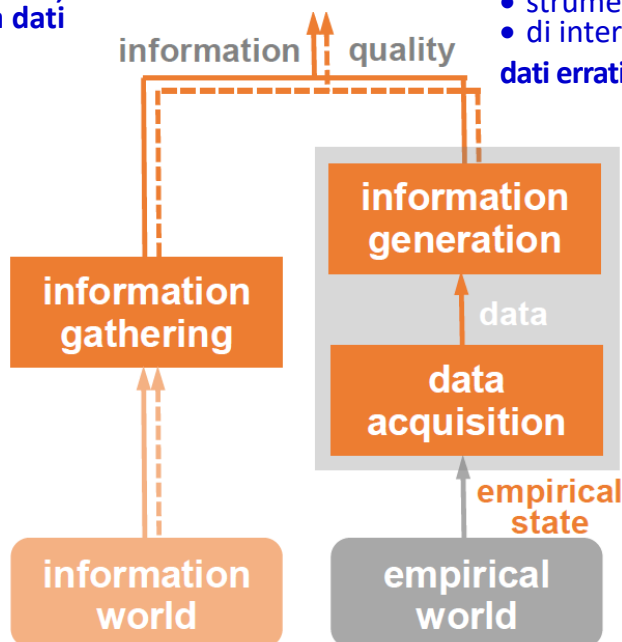
sorgenti di incertezza ?

D³M: generazione dell'informazione

- incertezza software di lettura
- copertura, attendibilità banca dati

- incertezza di misura
- di definizione
- strumentale
- di interazione
- dati errati/omessi

C'è dell'altro?



dati:

- **acquisiti** dal mondo empirico (ex. con sensori)
- **estratti** da banche dati (ex. in internet)

D³M: qualità dell'informazione di misura



grado in cui l'informazione ...

SINTATTICA: ... soddisfa predefinite regole di **integrità**:

integrità dell'oggetto (*sensore*), del riferimento, del dominio, ...

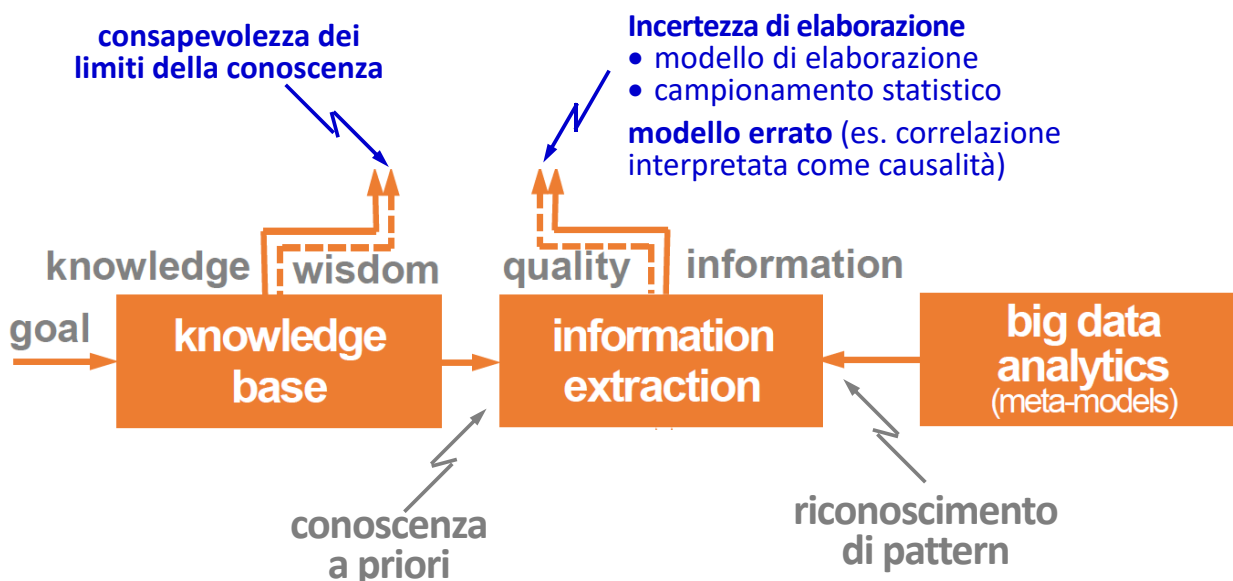
SEMANTICA: ... fornisce una **rappresentazione attendibile** del misurando (**significato** che si intende comunicare)

consistenza, correttezza (*incertezza*), risoluzione, concisione, contestuale

PRAMMATICA: ... soddisfa i **requisiti del DM** in un dato contesto (dipende dalla **conoscenza** dell'utilizzatore)

significatività, completezza, validità, non ambiguità, comprensibilità, tempestività, accessibilità, usabilità, confidenzialità, sicurezza

D³M: estrazione dell'informazione

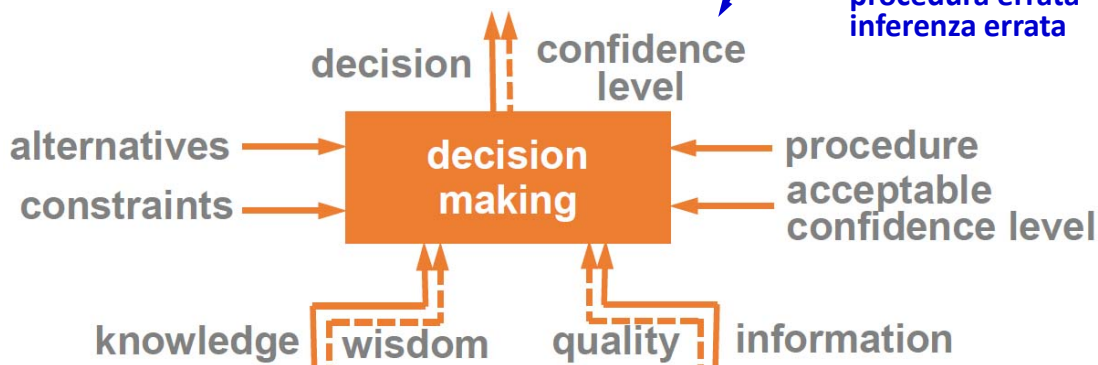


D³M: decisione

Incertezza di decisione:

- alternative ambigue
- procedura ambigua
- condizioni contrastanti
- pregiudizi cognitivi
- significatività informazione
- inferenza incerta

procedura errata
inferenza errata

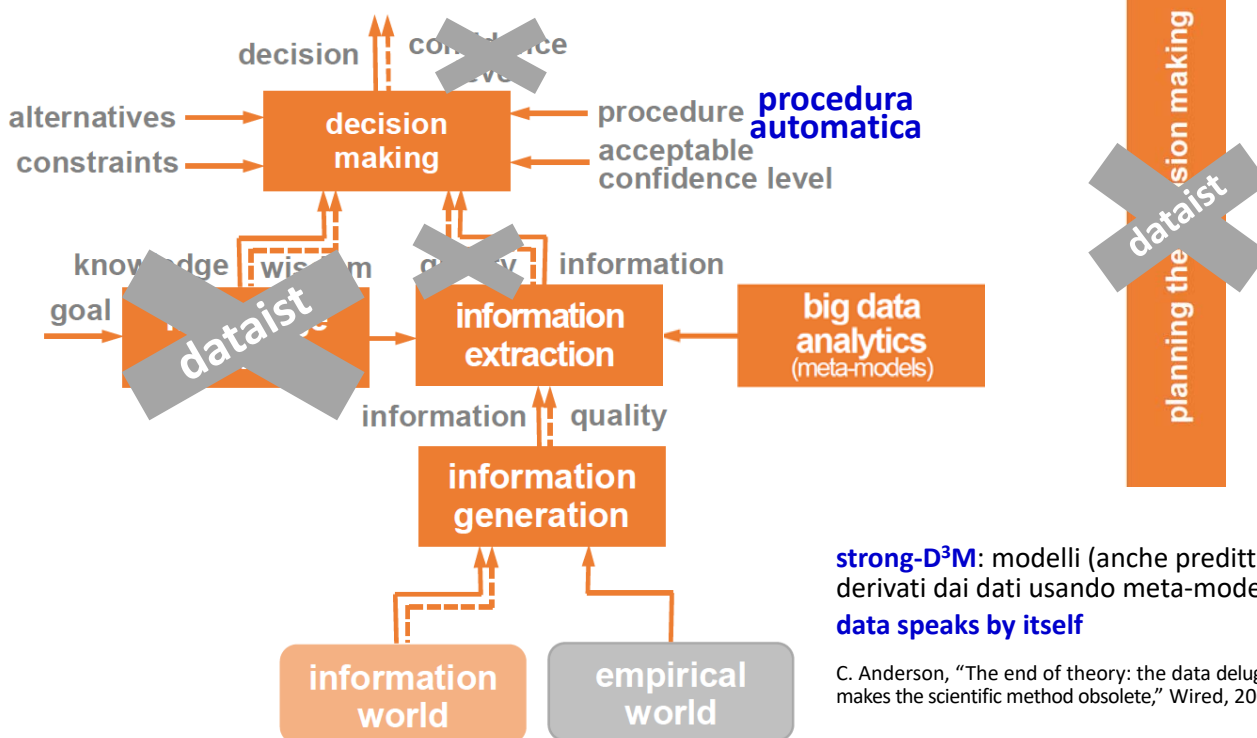


patologie:

Es.: errata procedura decisionale

Es.: inferenza errata a causa pregiudizi

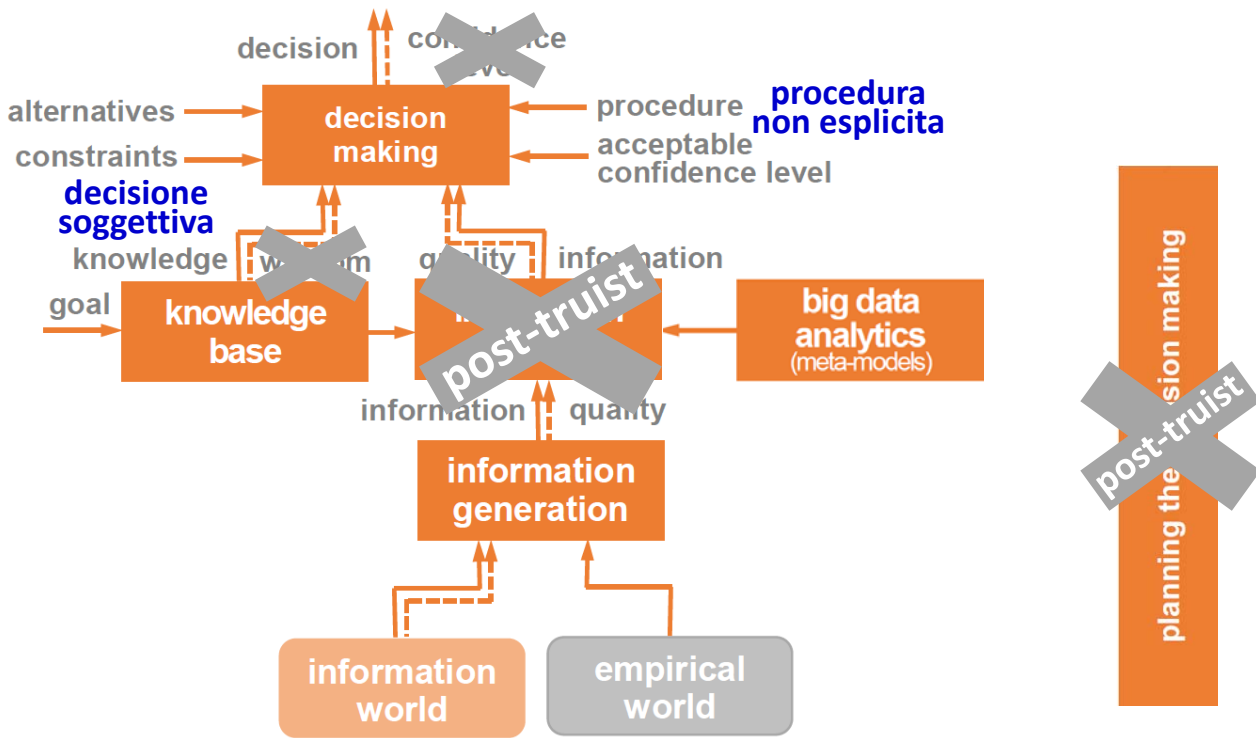
DM: approccio dataista



strong-D³M: modelli (anche predittivi) derivati dai dati usando meta-modelli
data speaks by itself

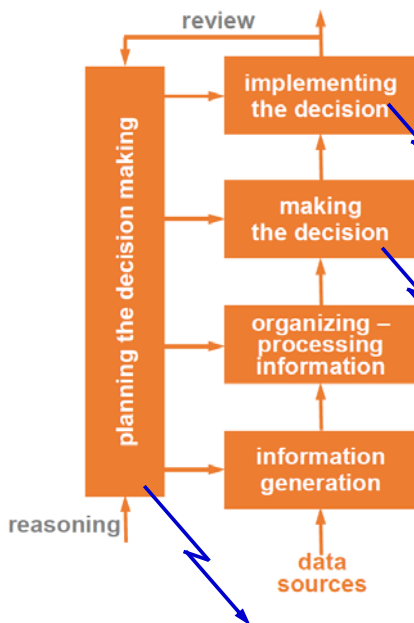
C. Anderson, "The end of theory: the data deluge makes the scientific method obsolete," Wired, 2008

DM: approccio post-verista



D³M: livello di fiducia

otto tipi possibili di errori decisionali

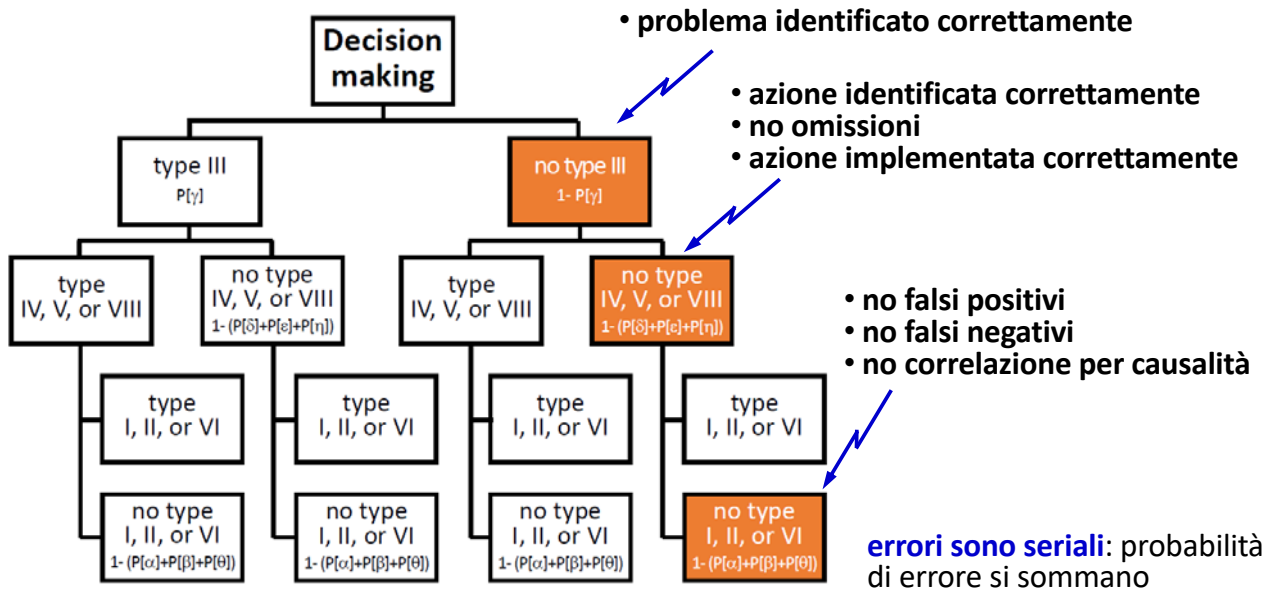


tipo	descrizione	esempio in ambito medico
VIII η	problema e azione corretti, ma azione implementata è involontariamente errata	mancata assunzione del medicinale prescritto nei tempi stabiliti

tipo	descrizione	esempio in ambito medico
I α	rifiutare H_0 quando è vera (falso positivo)	un esame rileva una patologia che la persona non ha
II β	non rifiutare H_0 quando è falsa (falso negativo)	un esame non rileva una patologia che la persona ha
IV δ	problema correttamente identificato, ma azione è errata (azione errata)	diagnosi e terapia corrette, ma effetti collaterali peggiori della patologia
V ε	decisione errata di non fare nulla (omissione)	
VI θ	decidere che il problema deriva da una causa in presenza di sola correlazione	
VII ζ	combinazione di diversi tipi di errore, risultante in un problema peggiore	patologia erroneamente diagnosticata (γ) e non trattata (ε)

tipo	descrizione	esempio in ambito medico
III γ	errori nella fase di pianificazione (identificazione del problema, delle alternative, dl contesto, ...)	diagnosi errata

D³M: livello di fiducia



$$P[\text{decisione corretta}] = 1 - \{ (1 - P[\gamma]) [1 - (P[\delta] + P[\epsilon] + P[\eta])] [1 - (P[\alpha] + P[\beta] + P[\theta])] \}$$

$P[\alpha]$ e $P[\beta]$ calcolabili con metodi statistici
 altre $P[\cdot]$ non sono facili da determinare

Decisioni complesse: conclusioni

solo un **giudizio soggettivo** basato su una solida **cultura metrologica**,

oltre a esperienza, conoscenza contesto, analisi critica dell'informazione usata, onestà intellettuale possono **garantire conclusioni attendibili e gestirne il livello di fiducia**

