



Anno scolastico
2020/21

Il sorprendente ruolo della statistica nella società moderna

Dalla valutazione dei farmaci agli algoritmi di
apprendimento automatico



Leonardo Grilli

Dipartimento di Statistica, Informatica,
Applicazioni «G. Parenti»



Università degli Studi di Firenze

Oggi parliamo di ...

- Statistica ed errate percezioni
- La Statistica nelle aule dei tribunali: l'accusa di discriminazione di genere nei confronti dell'Università di Berkeley
- Statistica e sperimentazioni cliniche: la valutazione dei vaccini per il Covid-19
- Statistica e apprendimento automatico (machine learning)

La Statistica

*La Statistica è l'**arte** e la **scienza** di imparare dai **dati***

In origine era una disciplina arida che consisteva nel manipolare lunghe sequenze di numeri (senza l'aiuto di calcolatrici!)

Molti credono che sia ancora oggi così ...

... ma la Statistica è diventata una disciplina assai diversa: ai calcoli ci pensa il computer, lo statistico è un **investigatore** che usa la metodologia statistica e la tecnologia informatica per individuare degli 'andamenti', delle 'regolarità' nella complessa realtà fisica, biologica, sociale, economica ...

Come i telescopi, i microscopi, i raggi X e i radar, la Statistica moderna consente di vedere cose invisibili a occhio nudo. *David Hand*

Errate percezioni

Su molti aspetti della realtà gli esseri umani hanno errate percezioni → la Statistica è necessaria per separare i fatti dalle opinioni (questo è importante nel dibattito pubblico!)

Proviamo a rispondere a queste domande:



Ogni 100 abitanti della tua nazione, quanti veicoli registrati vi sono?

Su 100 abitanti della tua nazione, quanti sono immigrati (cioè nati all'estero)?

Domanda: ogni 100 abitanti della tua nazione, quanti veicoli registrati vi sono?

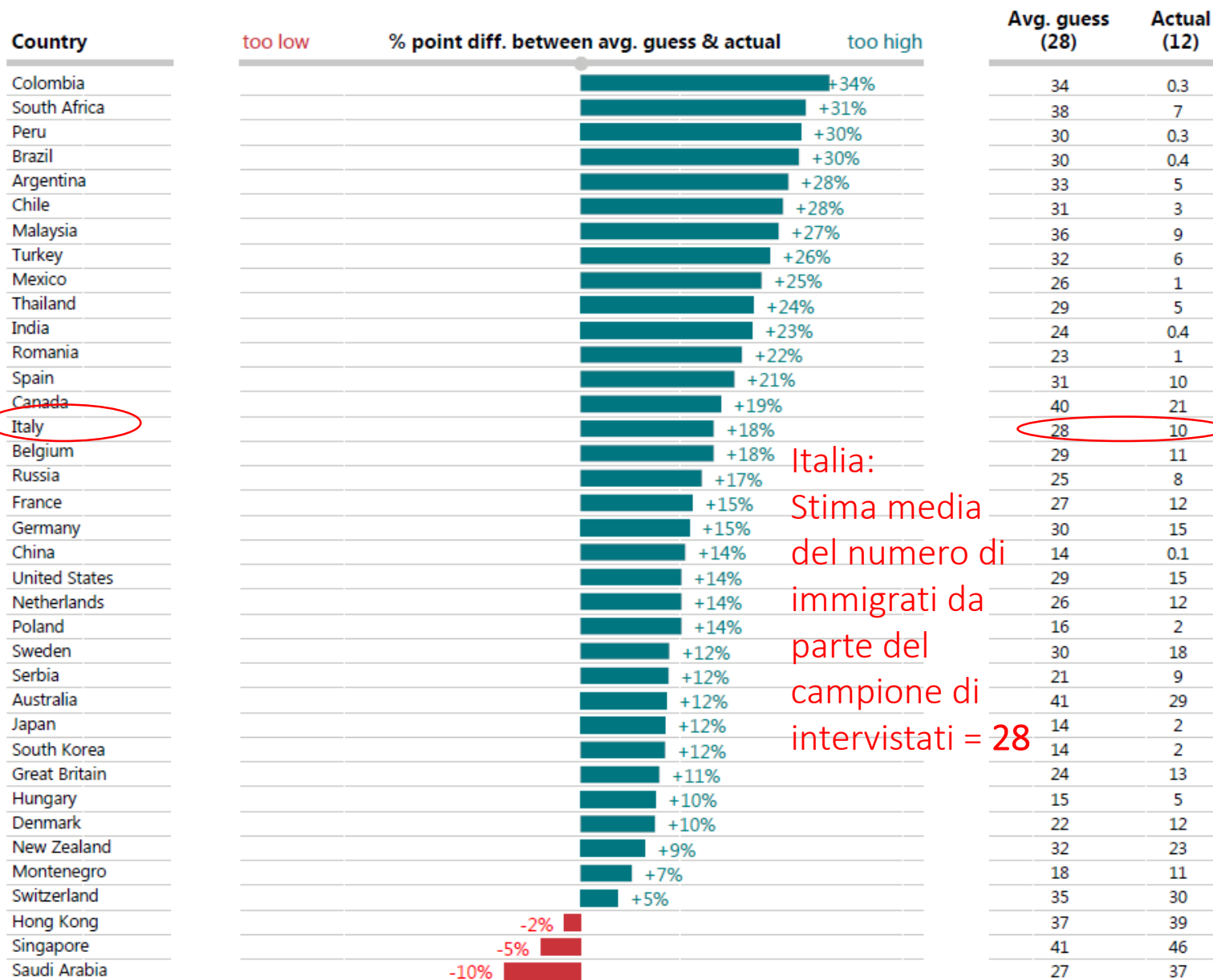
Country or district	too low	% point difference	too high	Avg. guess (66)	Actual (46)
Philippines			53%	61	8
Saudi Arabia			+51%	75	24
India			+50%	62	13
Chile			+46%	71	24
Peru			+46%	60	14
South Africa			+43%	62	19
Israel			+39%	75	35
Colombia			+39%	60	21
Montenegro			+37%	69	32
Indonesia			+36%	77	41
Singapore			+35%	53	18
Serbia			+30%	60	30
Mexico			+28%	62	34
Brazil			+27%	68	40
Hungary			+27%	64	37
Hong Kong			+27%	33	7
China			+26%	44	18
Turkey			+25%	49	24
Russia			+21%	56	35
Netherlands			+16%	73	57
Great Britain			+15%	71	56
Denmark			+15%	67	52
Belgium			+13%	75	63
France			+12%	77	65
Argentina			+12%	66	54
South Korea			+10%	56	46
Germany			+8%	73	65
Canada			+7%	71	64
Spain			+7%	76	69
New Zealand			+2%	76	74
Sweden			+2%	62	60
Australia		0		75	75
Malaysia		-1%		80	80
Norway		-4%		70	73
Poland		-5%		61	65
Italy		-6%		79	85
USA		-13%		73	85
Japan		-13%		58	72

Italia:
Stima media del
numero di veicoli
ogni 100 abitanti
da parte del
campione di
intervistati = 79

Italia:
numero di
veicoli ogni 100
abitanti = 85

Indagine Ipsos 'Perils of Perception' anno 2018. <https://perils.ipsos.com>

Domanda: su 100 abitanti della tua nazione, quanti sono immigrati?

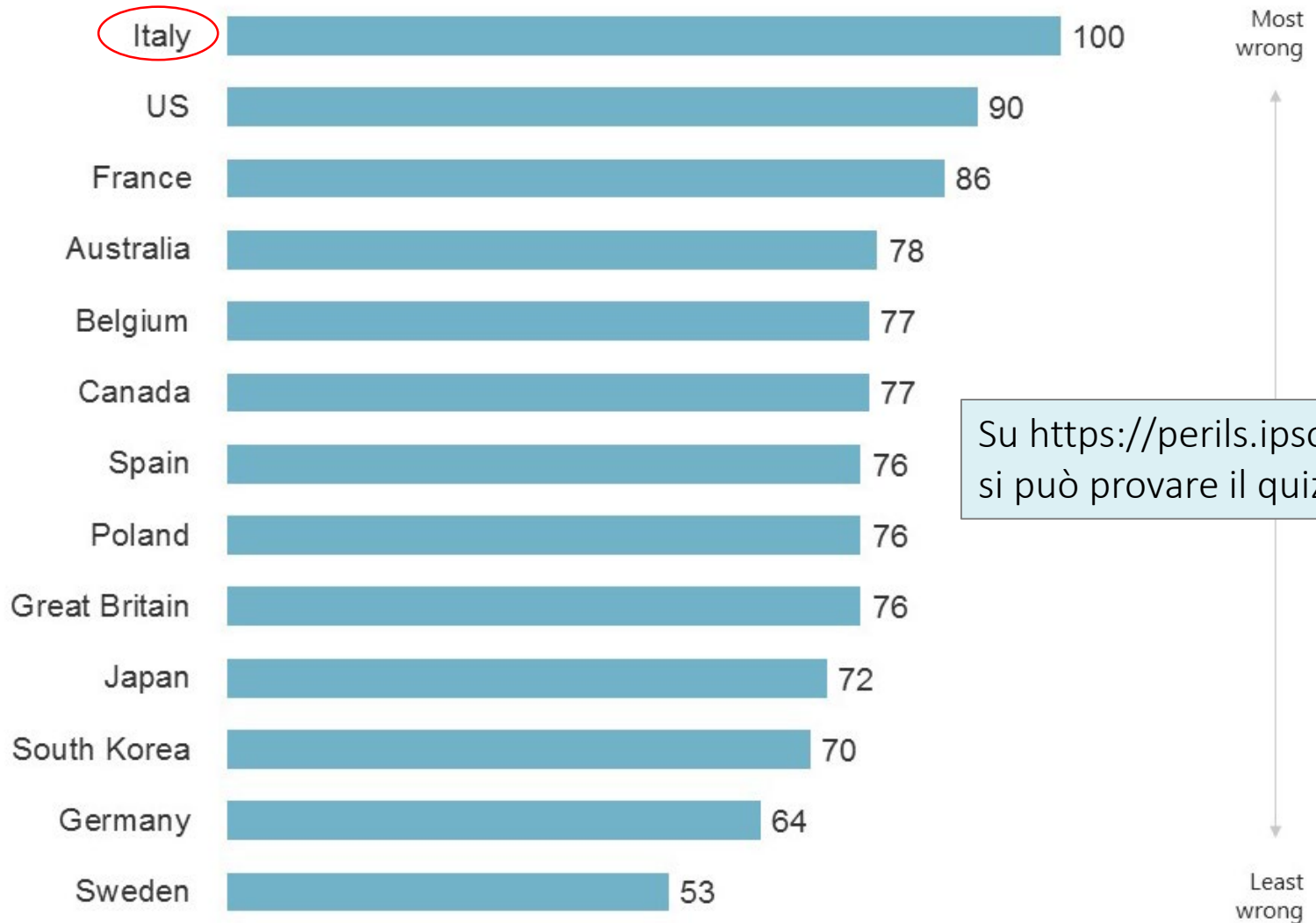


Italia:
Stima media
del numero di
immigrati da
parte del
campione di
intervistati = 28

Italia:
numero di
immigrati = 10

Indagini Ipsos 'Perirls of Perception' 2014-2017. <https://perils.ipsos.com>

Indice di errata percezione basato su oltre 50000 interviste in 13 nazioni, con domande su vari aspetti della società (lavoro, salute, immigrazione etc.). L'indice vale 100 per la nazione in cui si fanno più errori. The winner is ...



I risultati delle indagini Ipsos sulle percezioni sono discussi in questo libro di Bobby Duffy, ex direttore di Ipsos MORI, professore di Public Policy presso il King's College di Londra



P come Percezione. Dei rischi che corriamo tutti nell'interpretare la realtà. I dati ci spiegano quanto poco tutti noi sappiamo del mondo, quanto ci inganniamo su come stanno le cose.

La Statistica nelle aule dei tribunali:

l'accusa di discriminazione di genere nei confronti dell'Università di Berkeley

Statistica e giustizia



- In molti processi la decisione del giudice si basa, in parte più o meno grande, sui risultati di un'analisi statistica
- Alcuni esempi:
 - Valutazione probabilistica della **prova del DNA**
 - Nei processi ai dirigenti aziendali accusati di **inquinamento ambientale**: comparazione dei tassi di mortalità dell'area in cui ha sede lo stabilimento con i tassi di mortalità di aree simili
 - Nei processi ai datori di lavoro accusati di **discriminazione** nell'assunzione: comparazione dei tassi di assunzione in base a sesso, razza ecc.

Discriminazione di genere?



Negli anni Settanta un responsabile dell'università di Berkeley ha analizzato i dati relativi alle assunzioni nei vari dipartimenti, scoprendo che il tasso di ammissione delle femmine era sostanzialmente inferiore a quello dei maschi. L'università poteva essere accusata di discriminazione sessuale!

Per illustrare quel caso, consideriamo un esempio semplificato con 200 candidati, di cui 100 maschi e 100 femmine

	Maschi	Femmine
Assunti	55	45
Non assunti	45	55
Totale	100	100
% assunti	55.0%	45.0%

Il minore tasso di assunzione delle femmine è una prova di discriminazione di genere?

Discriminazione di genere? /cont.

	Sociologia		Fisica		Totale	
	Maschi	Femmine	Maschi	Femmine	Maschi	Femmine
Assunti	5	30	50	15	55	45
Non assunti	15	50	30	5	45	55
Totale	20	80	80	20	100	100
% assunti	25.0%	37.5%	62.5%	75.0%	55.0%	45.0%

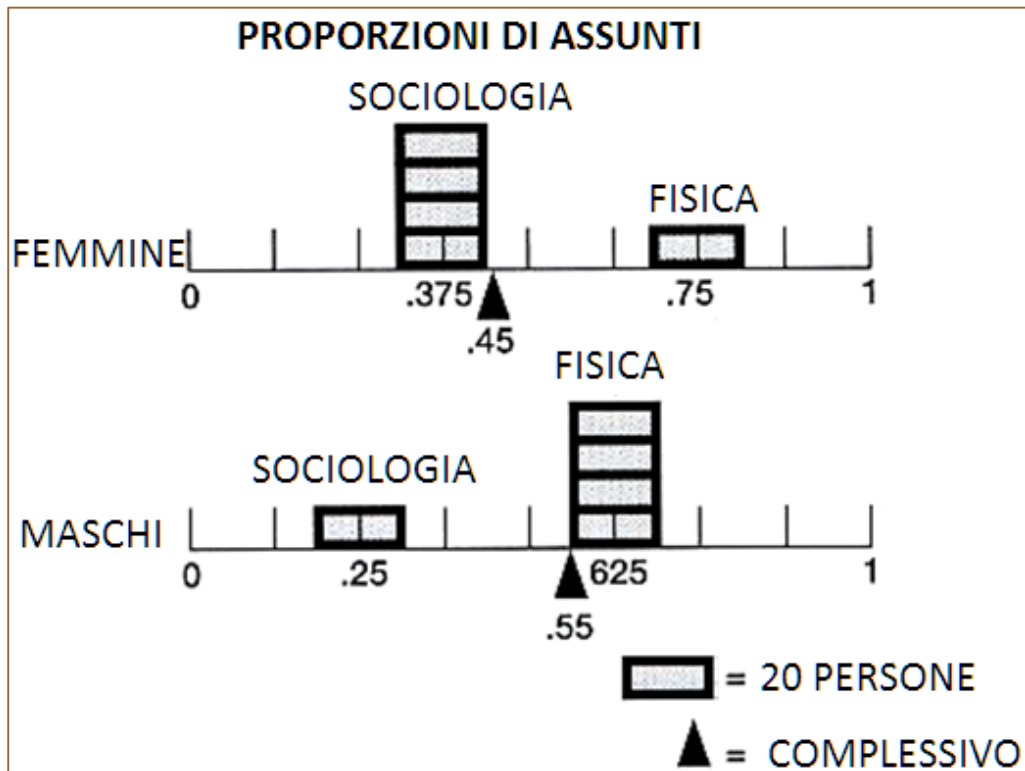
L'analisi dei dati ha mostrato che l'accusa era palesemente infondata: infatti, considerando i tassi di ammissione per dipartimento, la situazione era rovesciata poiché nella maggior parte dei dipartimenti le femmine facevano registrare un tasso di ammissione più elevato (il tasso di ammissione globale risultava inferiore perché le femmine facevano domanda soprattutto nei dipartimenti più «difficili»).

Cfr. Bickel PJ, Hjammel EA, O'Connell JW (1975) Sex Bias in Graduate Admissions: Data From Berkeley. *Science*, 187: 398–404.

E' un esempio del noto *Paradosso di Simpson*

http://it.wikipedia.org/wiki/Paradosso_di_Simpson

Discriminazione di genere? /cont.



Il tasso di successo complessivo si ottiene come media pesata dei tassi di successo nei due dipartimenti, usando pesi proporzionali al numero di partecipanti.

Ad es., le femmine hanno un tasso di successo di 0.375 a sociologia (dove partecipano in 80) e 0.75 a fisica (dove partecipano in 20) → il tasso di successo complessivo è $0.375 \cdot (80/100) + 0.75 \cdot (20/100) = 0.45$

Il disegno mostra che in entrambi i dipartimenti il tasso di successo delle femmine è superiore a quello dei maschi, ma per il tasso complessivo accade il contrario a causa dei pesi (le femmine partecipano in prevalenza alla selezione nel dipartimento più 'difficile').

**Statistica e
sperimentazioni cliniche:
la valutazione dei vaccini
per il Covid-19**

La sperimentazione clinica – perché?

- Le autorità sanitarie (es. EMA in Europa e FDA negli Stati Uniti) concedono l'autorizzazione all'uso di un vaccino in base ai dati della **sperimentazione clinica**, nella quale il vaccino viene somministrato ad un insieme di persone seguendo rigidi protocolli
- La sperimentazione clinica intende valutare se un vaccino è
 - **Sicuro**: non produce effetti collaterali gravi
 - **Efficace**: previene l'infezione o la malattia
- Ogni misura di risultato deve essere definita in modo rigoroso, es. si definisce *malato* chi soddisfa due condizioni: (1) risulta positivo ad uno specifico test diagnostico; (2) ha almeno un sintomo della malattia.
- Vi sono diverse misure di risultato, es. «infezione» e «malattia» (è possibile che un vaccino protegga parzialmente dall'infezione, ma quasi totalmente dalla malattia)



La sperimentazione clinica – come?

- Supponiamo che un vaccino venga somministrato a 10000 soggetti (**gruppo di trattamento**)
- Nell'intervallo di osservazione (es. 3 mesi), alcuni dei soggetti vaccinati sviluppano comunque la malattia: cosa si può concludere?
- Non si può pretendere che il vaccino abbia un'efficacia del 100%
- Ma come si può valutare se il risultato è soddisfacente?
- Abbiamo bisogno di un **gruppo di controllo**, cioè di un gruppo di soggetti non vaccinati, e confrontare i risultati
- Problema: come si sceglie il gruppo di controllo?
- Dovremmo cercare nella popolazione un gruppo di soggetti simili a quelli del gruppo di trattamento, simili per tutte le caratteristiche rilevanti, come *sex*, *età*, *etnia*, *indice di massa corporea*, *tipo di lavoro*, *stile di vita* ... questo elenco non ha termine, perché le caratteristiche rilevanti sono potenzialmente centinaia, molte non osservabili (es. genetiche)

Idea: gruppi con assegnazione casuale



Affinché i gruppi di trattamento e di controllo siano in media identici in tutte le caratteristiche, i soggetti vengono assegnati a caso (estraendo a sorte):

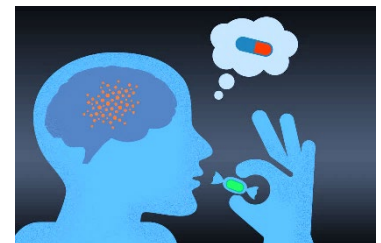
1. Si reclutano i soggetti secondo specifici criteri (es. età ≥ 16 , non avere l'infezione in corso, etc.)
2. Estraendo a sorte si assegnano i soggetti a due gruppi:
 - *Il **gruppo di trattamento** riceve il vaccino*
 - *Il **gruppo di controllo** non riceve il vaccino, ma il **placebo** (una soluzione fisiologica senza efficacia)*
3. Dopo un tempo stabilito si confrontano i due gruppi in base a una *variabile di risultato*, ad es. la percentuale di infettati oppure la percentuale di malati

Con l'assegnazione casuale i due gruppi (trattamento e controllo) sono in media identici in tutte le caratteristiche *osservate* (sesso, età, ...) e *non osservate* (malattie non diagnosticate, profilo genetico, ...). Al momento in cui si somministra il vaccino, *due gruppi diventano diversi per il fatto di ricevere o non ricevere il vaccino* e quindi la differenza di risultato è attribuibile al vaccino.

Placebo e doppio cieco

- Il **placebo** è un finto vaccino, che viene somministrato con una iniezione proprio come il vero vaccino
- Gli esperimenti clinici con placebo di solito sono in **doppio cieco**: quando un soggetto riceve l'iniezione, né il soggetto né il personale sanitario sanno se si tratta del vaccino o del placebo (questa informazione è nota solo agli organizzatori dell'esperimento)
- Il fine della procedura con placebo in doppio cieco è di *evitare distorsioni indotte dal sapere di essere trattato*: ad esempio, se un soggetto ha certezza di aver ricevuto il vaccino potrebbe cambiare il proprio comportamento, riducendo le misure di protezione contro l'infezione – in questo modo il gruppo di trattamento sarebbe diverso dal gruppo di controllo non solo per il fatto di ricevere il vaccino, ma anche per la diversa adozione delle misure di protezione

nel caso dei farmaci si parla di *effetto placebo*: alcuni soggetti mostrano un miglioramento a causa dell'effetto psicologico di sapere di essere curati



La sperimentazione Pfizer-BioNTech

Il vaccino Pfizer-BioNTech è stato il primo vaccino per il Covid-19 ad ottenere l'autorizzazione da parte dell'EMA. Nella sperimentazione i partecipanti sono stati assegnati a caso a

- **Gruppo di trattamento:** 18860 persone hanno ricevuto il vaccino, quasi tutti due dosi (a distanza di 21 giorni)
- **Gruppo di controllo:** 18846 persone hanno ricevuto il placebo, quasi tutti due dosi (a distanza di 21 giorni)

I partecipanti appartenevano a 4 nazioni (USA, Argentina, Brasile, Sud Africa) ed erano di diverse etnie.

Il 57.8% aveva tra 16 e 55 anni, il 42.2% aveva più di 55 anni → non ci sono dati relativi a bambini e adolescenti (per loro infatti EMA non ha autorizzato l'uso), invece ci sono dati a sufficienza relativi agli anziani (a differenza del vaccino AstraZeneca).

Efficacia del vaccino Pfizer-BioNTech

I casi di infezione (rilevati almeno 7 giorni dopo la somministrazione della seconda dose) sono stati

- **Gruppo di trattamento:** 8
- **Gruppo di controllo:** 162

Pertanto il vaccino è stato in grado di prevenire $162 - 8 = 154$ casi di infezione, ovvero $154/162 = 95.06\%$ (questa è la misura di efficacia)

In base alla sperimentazione, effettuata in un certo momento storico, con una popolazione di un certo tipo, questo vaccino evita circa il 95% delle infezioni

In pratica, l'efficacia reale varia in dipendenza di molti fattori (tipo di popolazione a cui si applica, varianti del virus in circolazione ...)

Effetti collaterali

Nella sperimentazione (fase 3) del vaccino Pfizer-BioNTech nel gruppo di trattamento (persone che hanno ricevuto il vaccino) si sono registrati 2 decessi

Ma sono decessi imputabili al vaccino oppure ad altre cause?
Qui è in agguato la fallacia nota come

post hoc, ergo propter hoc
(dopo questo, e quindi a causa di questo)

sofisma per il quale si afferma l'esistenza di un rapporto di causalità tra due avvenimenti, per il solo fatto che l'uno è posteriore all'altro.

Effetti collaterali /cont.

Dunque, i decessi avvenuti dopo la somministrazione del vaccino sono dovuti al vaccino oppure ad alte cause?

La risposta si fonda su due elementi

- Accertamenti medici
- Analisi statistica

L'analisi statistica si basa sul fatto che, osservando un grande gruppo di persone per alcune settimane, si registrano alcuni decessi per cause naturali.

Come facciamo a capire se 2 decessi in un gruppo di 18'000 persone sono in linea con la mortalità naturale? Abbiamo bisogno di un punto di riferimento (benchmark) ...

Effetti collaterali /cont.

Nel caso di un esperimento è facile individuare il benchmark: guardiamo cosa è accaduto nel gruppo di controllo (persone che hanno non hanno ricevuto il vaccino).

Nel gruppo di controllo Pfizer-BioNTech si sono verificati 3 decessi.

Dunque i 2 decessi osservati nel gruppo di trattamento sono in linea con la mortalità naturale e quindi non sono motivo di allarme. Inoltre, i medici hanno accertato che i 2 decessi del gruppo di trattamento non erano dovuti a cause imputabili al vaccino.

Cosa accade con la vaccinazione di massa?

In base ai dati del 2020, mediamente in Italia ogni giorno muore 1 persona ogni 30'000. Vaccinando 300'000 persone al giorno ci aspettiamo 10 decessi per cause naturali, *dopo* il vaccino, ma *non a causa* del vaccino.

Il dibattito sui vaccini

- I vaccini costituiscono uno dei maggiori **contributi della medicina al miglioramento della qualità e dell'aspettativa di vita**
- Ad es. l'Organizzazione Mondiale della Sanità attesta che la vaccinazione contro il **morbillo** ha realizzato un calo dell'80% delle morti per morbillo tra il 2000 e il 2017 in tutto il mondo, prevenendo circa 21 milioni di morti (il morbillo è molto contagioso ed ha un tasso di mortalità nell'ordine di 1 su 1000, che diventa molto maggiore in caso di malnutrizione)
- Nonostante questo, *negli ultimi anni è aumentato il numero di coloro che sono dubbiosi o ostili nei confronti dei vaccini* – i motivi sono molti, dalle teorie complottiste all'illusione che certe malattie siano scomparse (mentre sono sempre in agguato, come dimostrato dalle periodiche epidemie di morbillo)

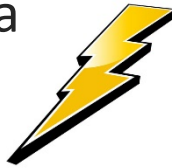
Il dibattito sui vaccini /cont.

- I vaccini hanno tipicamente **effetti collaterali modesti** (dolore muscolare, mal di testa), mentre gli effetti collaterali gravi sono rarissimi
- Per contro, i vaccini garantiscono **benefici enormi** perché consentono di evitare migliaia di morti dovuti alle conseguenze delle malattie – anche una malattia ‘comune’ quale l’influenza stagionale causa ogni anno molti decessi per via delle complicanze (in Italia si stima che le complicanze dell’influenza provochino da 4000 a 10000 morti all’anno)
- I mass media e i social non aiutano a valutare correttamente costi e benefici: una persona che ha avuto un effetto collaterale grave fa più notizia di migliaia di vite salvate

Stimare costi e benefici



- **Effetti collaterali:** al 22/03/21 per il vaccino AstraZeneca sono stati segnalati 86 casi di trombosi (di cui 18 fatali) su 25 milioni di vaccinati (il nesso causale è plausibile, ma non dimostrato) → il tasso è **3.44 per milione**: la probabilità è molto piccola, inferiore a quella che una persona durante la vita sia colpita da un fulmine (12.5 per milione).
- **Benefici:** nel primo anno di pandemia si sono registrati 100'000 morti da Covid su 60 milioni di italiani → il tasso è **1666.7 per milione** e il vaccino evita quasi del tutto i casi mortali
- Il rapporto tra i due tassi è $1666.7/3.44 = 485$, quindi molto favorevole



Questa stima si basa su varie ipotesi su cui potremmo discutere a lungo (es. accuratezza dei dati) e fa riferimento alla mortalità passata (quella futura sarà, auspicabilmente, più bassa), ma è utile per avere un ordine di grandezza costi-benefici. *Attenzione: si tratta di una stima per l'intera popolazione, i valori cambiano se consideriamo dei particolari gruppi, ad es. per fascia d'età.*

Ancora sugli effetti collaterali

- Consideriamo questi effetti collaterali:
 - prolungamento del tempo di sanguinamento
 - disturbi gastrici
 - mal di testa
 - ronzio nell'orecchio
 - sindrome asmatica
 - distress cardiorespiratorio (grave e acuta insufficienza respiratoria)
 - alterazione della funzione renale
 - sindrome di Reye, una malattia acuta a carico del cervello e del fegato, potenzialmente fatale (effetto raro, che colpisce quasi esclusivamente i bambini)

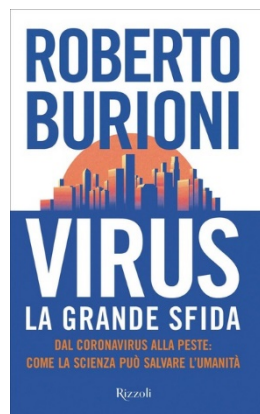
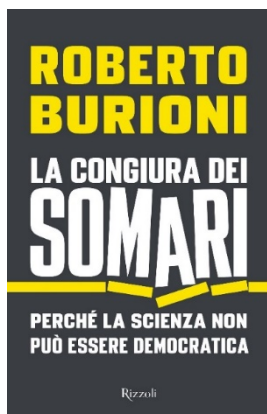
Ancora sugli effetti collaterali /cont.

- Gli effetti collaterali appena elencati sono relativi all'*acido acetilsalicidico*, il principio attivo dell'Aspirina[®], un farmaco che non necessita di prescrizione medica
- Questo è solo un esempio relativo ad un farmaco molto comune: tutti i farmaci, come i vaccini hanno dei potenziali effetti collaterali
- Le *segnalazioni di sospette reazioni avverse ai farmaci* sono raccolte in una banca dati europea, utilizzata per il monitoraggio di tutti i prodotti autorizzati – la banca dati è liberamente consultabile <http://www.adrreports.eu/it/index.html>
- L'impiego di farmaci e vaccini si base sulla valutazione complessiva di potenziali rischi e benefici – ***il rischio zero non esiste!***

ZERO
RISCHIO?

Vaccini e società

- Per tenere sotto controllo una malattia infettiva occorre vaccinare gran parte della popolazione (70-80%, la cosiddetta ‘immunità di gregge’).
- Le istituzioni pubbliche e gli scienziati devono comunicare in modo corretto ed efficace i costi e i benefici della vaccinazione.
- Per saperne di più si possono leggere i libri del virologo Roberto Burioni, che gestisce il sito web *Medical Facts* www.medicalfacts.it



Nota a margine. L'*effetto Dunning-Kruger* è una distorsione cognitiva a causa della quale individui poco esperti in un campo tendono a sopravvalutare le proprie abilità autovalutandosi, a torto, esperti in quel campo. Come corollario di questa teoria, spesso gli incompetenti si dimostrano estremamente supponenti.

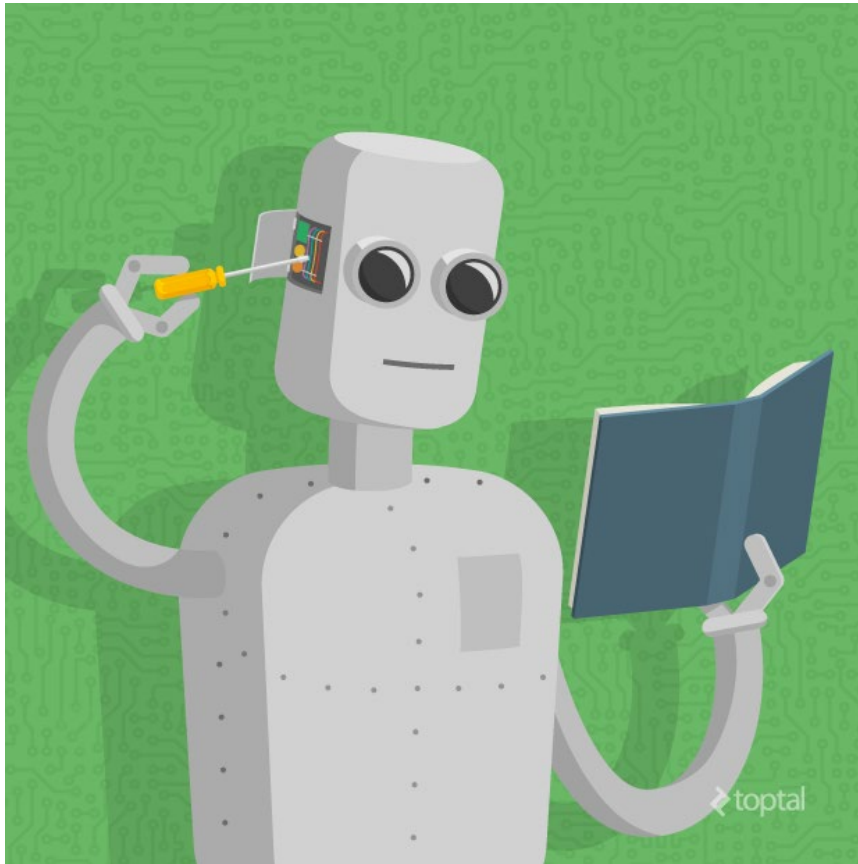
https://it.wikipedia.org/wiki/Effetto_Dunning-Kruger

Distorsioni cognitive

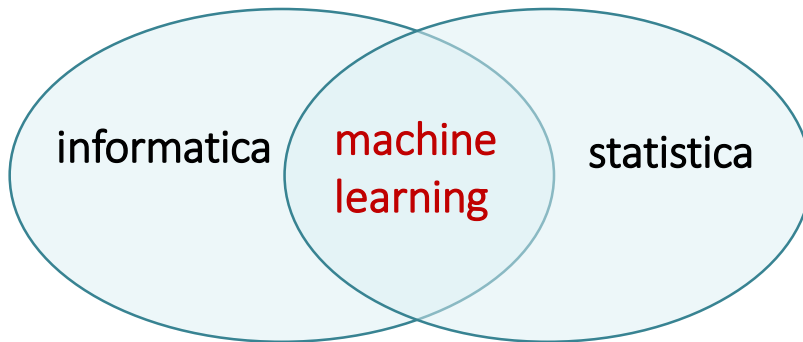


Astutillo    
Smeriglia

Statistica e apprendimento automatico (machine learning)



L'apprendimento automatico (o **machine learning**) è una branca dell'*intelligenza artificiale* che «fornisce ai computer l'abilità di apprendere senza essere stati esplicitamente programmati» (Arthur Samuel, 1959). Viene impiegato in quei campi dell'informatica nei quali progettare e programmare algoritmi espliciti è impraticabile.



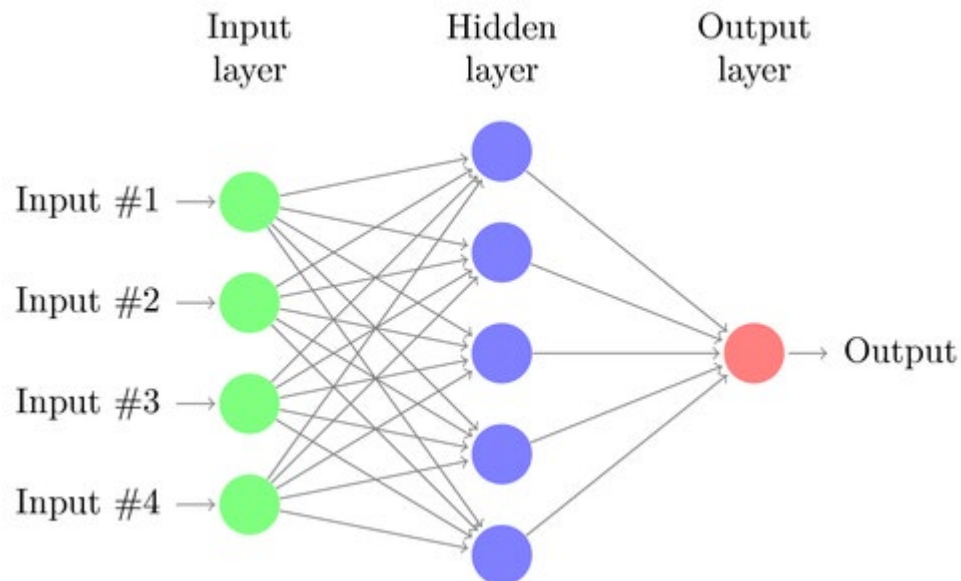
Il machine learning si basa su metodi di **analisi statistica multivariata**

Come apprendono le macchine?

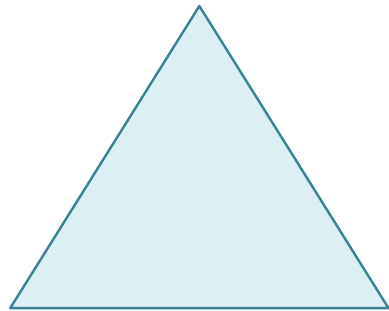
Ci sono numerosi metodi di *machine learning*: la scelta del metodo dipende dal tipo di applicazione.

Alcuni metodi mimano il funzionamento del cervello umano: si tratta delle **reti neurali** (con i recenti sviluppi del **deep learning**), che sono *modelli matematici del funzionamento del cervello* (estremamente semplificati, dato che il cervello umano ha una complessità ben lungi dall'essere riproducibile su una macchina).

cerchi \Leftrightarrow neuroni
freccette \Leftrightarrow sinapsi



Trovare una regola. Facile...

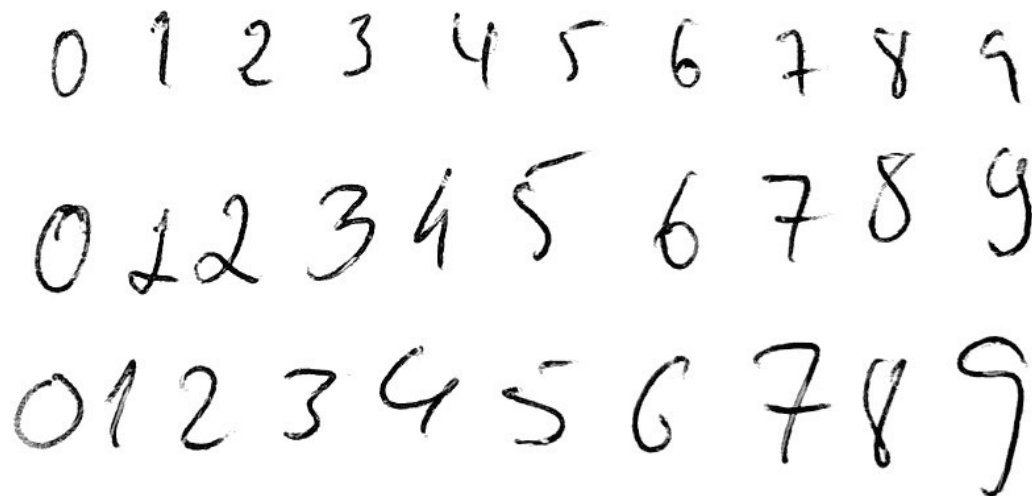


Per distinguere un triangolo da un quadrato c'è una semplice regola:

«se ha 3 lati è un triangolo, se ha 4 lati è un quadrato»

Trovare una regola. Non facile...

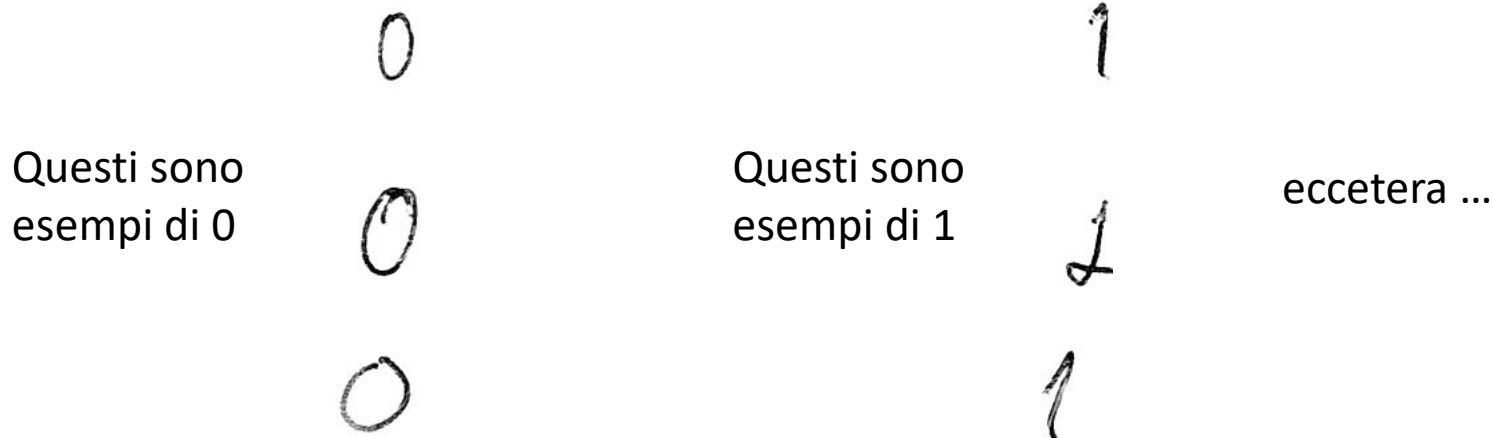
Non esistono semplici regole per riconoscere i caratteri scritti a mano. Ognuno di noi sa riconoscere i caratteri, ma non riesce a tradurre questa capacità in un insieme di regole.



Una delle prime applicazioni del machine learning è stato il ***riconoscimento automatico dei caratteri*** (OCR: Optical Character Recognition) – adesso disponibile su uno smartphone!

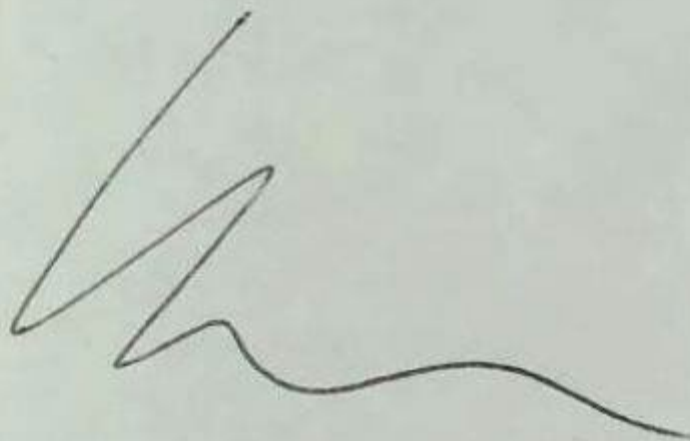
Apprendimento supervisionato

- Il riconoscimento dei caratteri è un esempio di **apprendimento supervisionato** perché si addestra il sistema fornendo una serie di casi di cui è nota la classificazione (*training set*):



- Pian piano il sistema impara a riconoscere i caratteri e quindi diventa in grado di classificare i caratteri scritti da persone che non fanno parte del training set. Naturalmente, il sistema può sbagliare (infatti fornisce un *tasso di errore*) – d'altra parte anche noi umani a volte non riusciamo a interpretare la scrittura a mano.

**QUANDO IL DOTTORE
TI DICE "PRENDA QUESTO"**



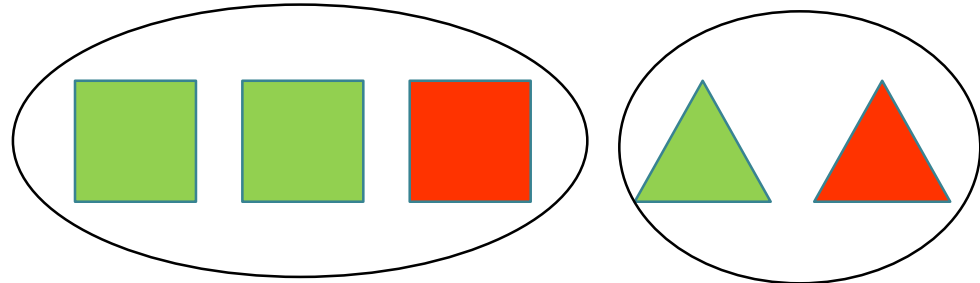
E IL FARMACISTA TE LO DÀ

Apprendimento non supervisionato

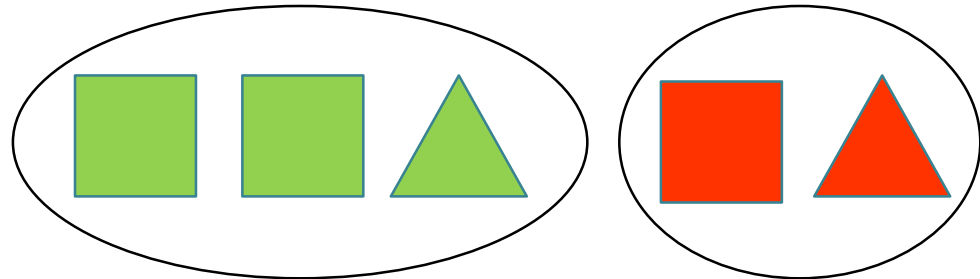
In alcune situazioni vi sono molti modi di fare la classificazione e a priori non si sa quale sia il migliore per raggiungere l'obiettivo.



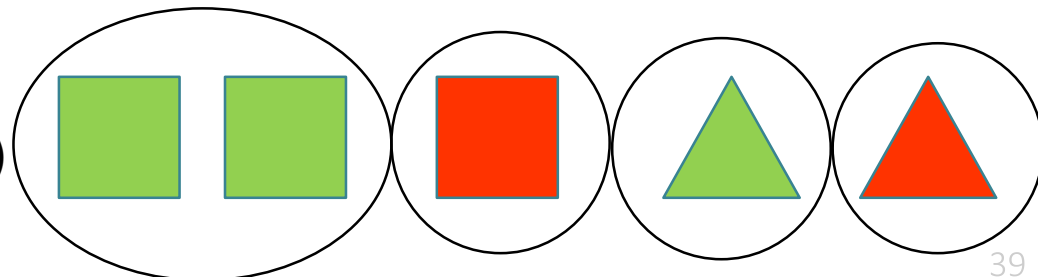
Classificazione n. 1: in base alla figura geometrica



Classificazione n. 2: in base al colore



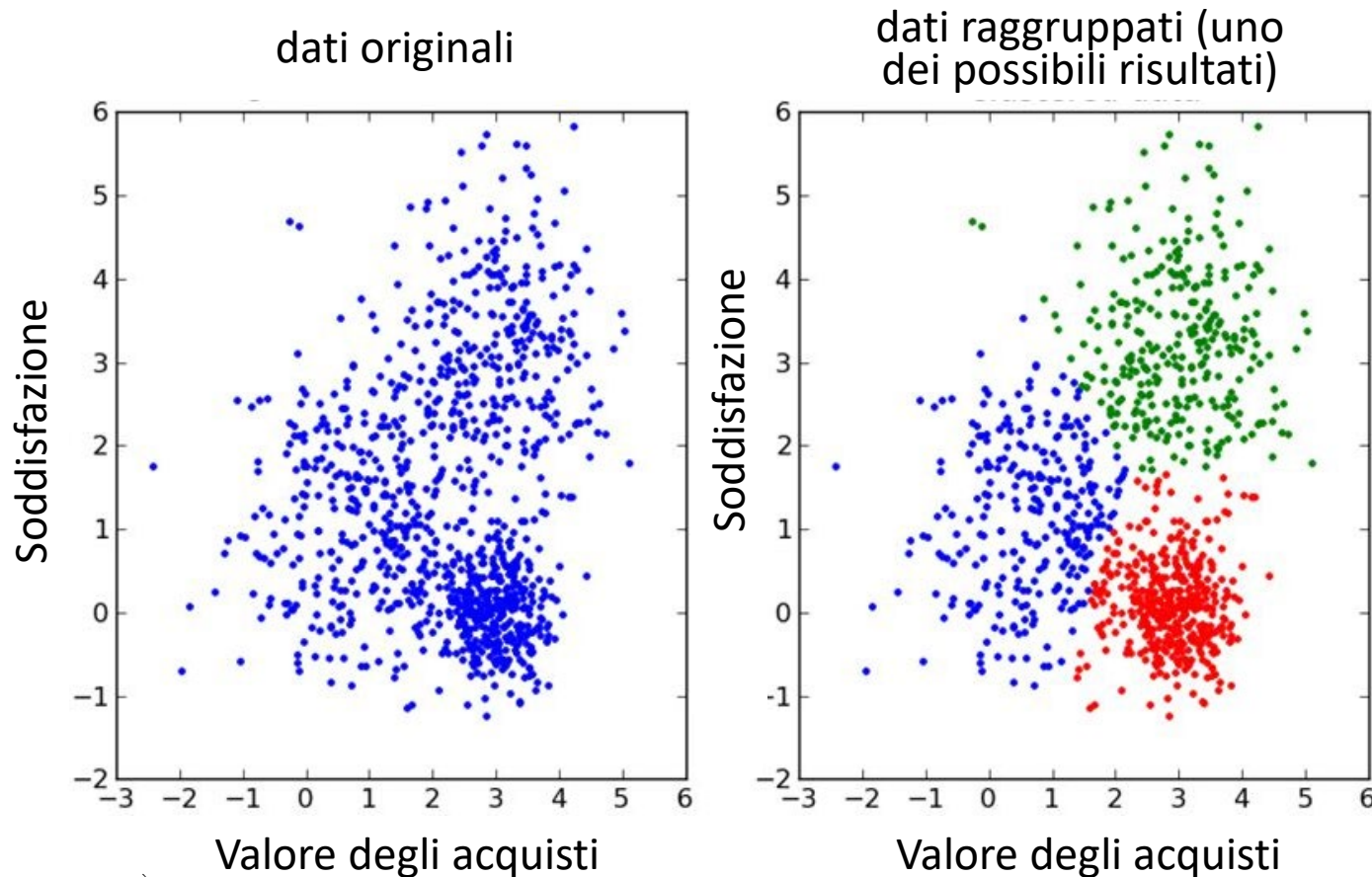
Classificazione n. 3: in base alla figura geometrica e al colore (congiuntamente)



Apprendimento non supervisionato

La metodologia statistica più usata per l'**apprendimento non supervisionato** è l'**analisi dei gruppi** (*cluster analysis*).

Esempio. Un'azienda vende prodotti sul web. Per ogni cliente registrato si hanno molti dati, tra cui il valore degli acquisti nell'ultimo anno e il livello di soddisfazione.



Questa è una situazione semplice perché ci sono solo 2 variabili (si può fare il grafico in 2 dimensioni), nella realtà si usano centinaia di variabili

Apprendimento non supervisionato

- E' un esempio di *segmentazione di mercato*: l'azienda è interessata a classificare i propri clienti in gruppi sufficientemente omogenei in modo da fare campagne pubblicitarie specifiche per ogni gruppo.
- Ad es., l'azienda dovrebbe prendere in seria considerazione il gruppo di punti rossi, che corrisponde a clienti che acquistano molto e sono poco soddisfatti
- Questo problema ha molte possibili soluzioni: si possono fare 3 gruppi, oppure 4, oppure 5 ... ma soprattutto l'assegnazione dei clienti ai gruppi può essere fatta in molti modi perché le variabili a disposizione sono molte (età, sesso, stato civile, residenza, preferenze dichiarate, abitudini di acquisto ecc.)

Applicazioni del machine learning

- **Credit scoring:** attribuire ad una persona che chiede un prestito la probabilità che lo restituisca
- **Spam filtering:** classificare le email in spam vs non-spam
- **Diagnosi medica:** classificare un paziente usando i dati di anamnesi
- **Riconoscimento dei volti umani** <https://www.faceplusplus.com>
- **Suggerimenti di acquisto personalizzati** (Amazon, Netflix)
- **Riconoscimento vocale e assistente digitale** (es. Siri)
- **Traduzione linguistica** (es. Google Translate)
- **Analisi dei testi** (annunci pubblicitari, tweet, referti medici ...)
- **Automobili a guida automatica:**

Il futuro dell'intelligenza artificiale



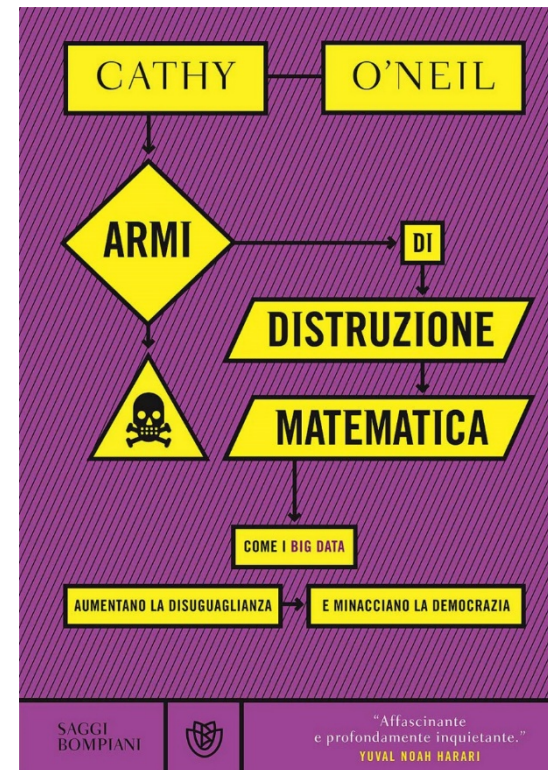
- In una intervista su [esquire.com](https://www.esquire.com) del 10/10/2019 (*Il futuro dell'intelligenza artificiale, secondo l'uomo che l'ha inventata*), Yann LeCun fa il punto della situazione: «Vorrei fornire alle macchine una qualche forma di buon senso, visto che oggi sono veramente stupide»
- «Anche se in alcune aree è capace di performance superumane, l'apprendimento supervisionato ha dei limiti: bisogna usare un sacco di dati, bisogna indicare alla macchina quando sta sbagliando e ci saranno comunque dei punti ciechi che la porteranno a commettere degli errori. [...] Al momento, l'incapacità delle macchine di imparare senza una supervisione è il più grande limite del deep learning [un tipo di apprendimento basato su reti neurali]»
- Sarà il deep learning che ci farà raggiungere l'intelligenza artificiale di tipo umano? «Sono convinto che sarà parte della soluzione, anche se ovviamente ci saranno nuovi paradigmi di apprendimento. Stiamo lavorando proprio a questo, ma non possiamo dire quanto tempo ci vorrà. Molto dipende dalle svolte che si compiono grazie ad alcune improvvise scoperte, che non possiamo predire se e quando avverranno»

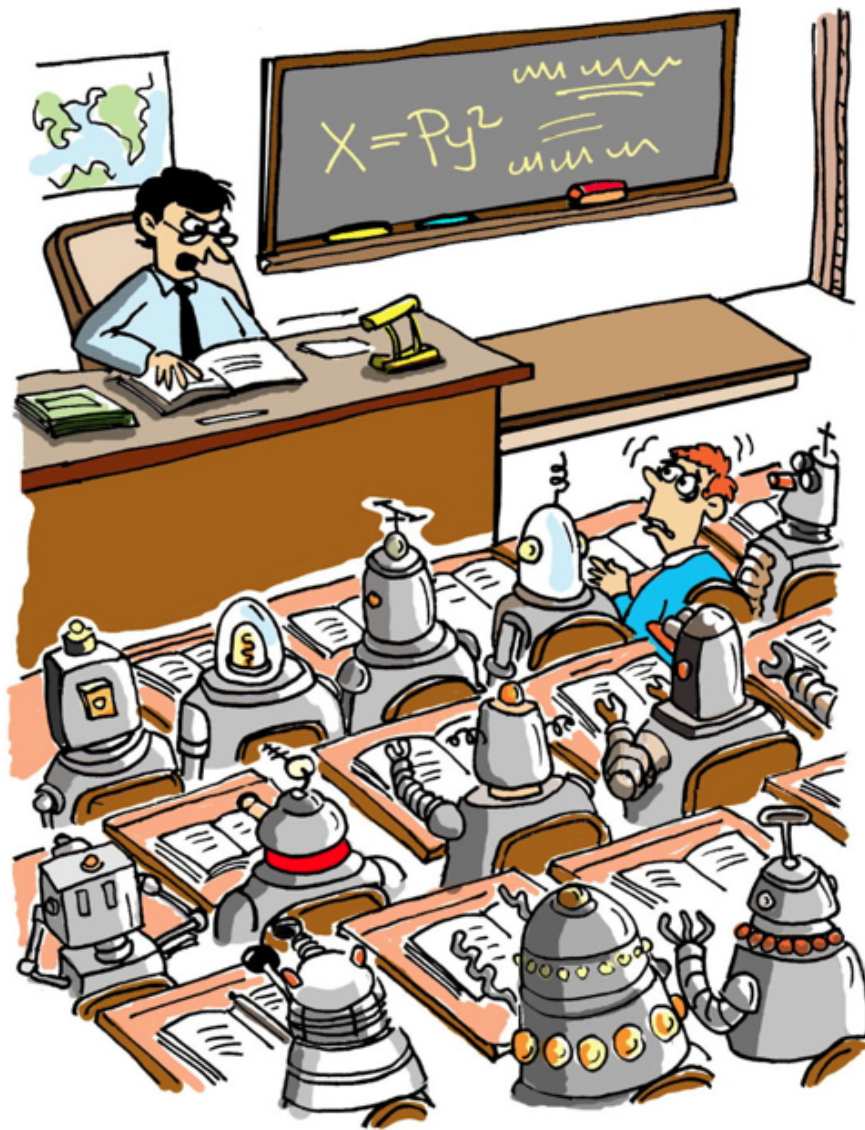
Intelligenza artificiale etica

Gli algoritmi di apprendimento automatico vengono sempre più usati a **supporto di decisioni**, es. per decidere se

- assumere un candidato (in base al suo curriculum)
- licenziare un insegnante (in base ai risultati dei suoi studenti)
- concedere la libertà condizionata ad un carcerato (in base al rischio di recidiva)

Tutto questo pone seri **problemi etici**, soprattutto nei casi in cui la decisione dipende in tutto o in gran parte dal responso dell'algoritmo.





Signor Rossi, è sicuro di trovarsi nell'aula giusta? Questo è il corso di Machine Learning.

Studiare statistica

La statistica nella società moderna

- **GOVERNO**: le *politiche efficaci* si basano sull'analisi statistica dei dati relativi alla società e all'economia
- **AZIENDE**: le decisioni su *cosa e quanto produrre* si basano su ricerche di mercato e analisi di dati economici; i *rischi di banche e assicurazioni* sono valutati tramite modelli statistici (es. sistemi di *credit scoring* per valutare preventivamente la rischiosità di un prestito)

la statistica è fondamentale per lo sviluppo dei sistemi di apprendimento automatico (**machine learning**), ad esempio: *filtri anti-spam* per l'email, riconoscimento automatico dei *caratteri scritti a mano*, riconoscimento automatico delle *immagini* (come i volti umani), personalizzazione dei *suggerimenti di acquisto* (come fanno Amazon, Netflix etc.)

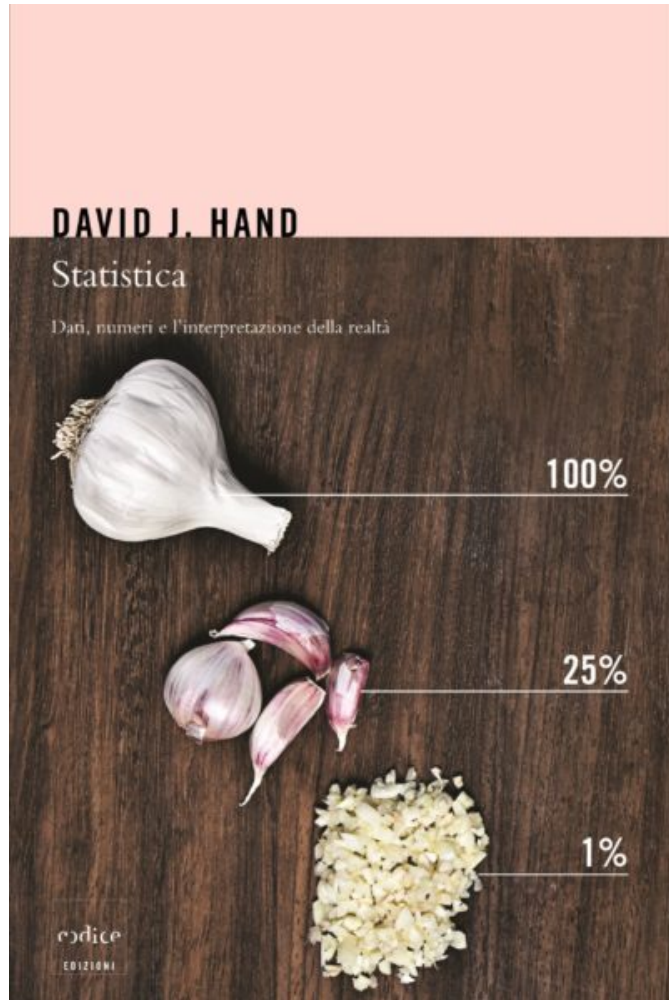
La statistica nella società moderna

- **AMBIENTE:** l'*inquinamento* viene monitorato da sistemi geo-statistici; il *cambiamento climatico* viene valutato analizzando dati di varia natura, dai carotaggi nei ghiacci alle radiazioni misurate dai satelliti
- **MEDICINA:** l'efficacia di *farmaci* e *vaccini* è testata tramite esperimenti clinici pianificati con criteri statistici; le *malattie* (fattori di rischio, diffusione) sono studiate con metodi statistici (epidemiologia)

L'epidemia di Covid-19 ha reso evidente l'importanza di un sistema statistico per il monitoraggio e la previsione. Le decisioni del Comitato Tecnico-Scientifico sono basate sull'analisi di dati.

In questo ambito un ottimo esempio di sistema statistico è quello messo a punto da StatGroup-19 (anche su Facebook) <https://statgroup19.shinyapps.io/Covid19App>

Due libri per capire la statistica moderna



Big Data

- Oggi sono disponibili dataset grandi e complessi (Big Data):
 - Dati prodotti dall'attività su siti web, Internet of Things, sensori, ..., es. dati GPS sui movimenti di persone o automobili
 - Immagini da satellite, da risonanza magnetica ...
 - Testi prodotti dall'attività sui social network
- Big Data: definiti tramite le tre V
 - Volume (molte unità statistiche, molte variabili)
 - Velocità (si aggiornano rapidamente)
 - Varietà (misurazioni, testi, coordinate geografiche, immagini ...)
- Le sfide poste dai Big Data richiedono competenze sia di Statistica che di Informatica (si parla di Data Science)

La professione dello statistico

Nelle classifiche sulla qualità del lavoro stilate dal sito americano CareerCast ai primi posti troviamo le professioni di **Statistico**, e quelle affini di **Data Scientist**, **Matematico** e **Attuario**

<http://www.careercast.com/jobs-rated/2018-jobs-rated-report>

Sulle prospettive di carriera si veda <http://thisisstatistics.org>

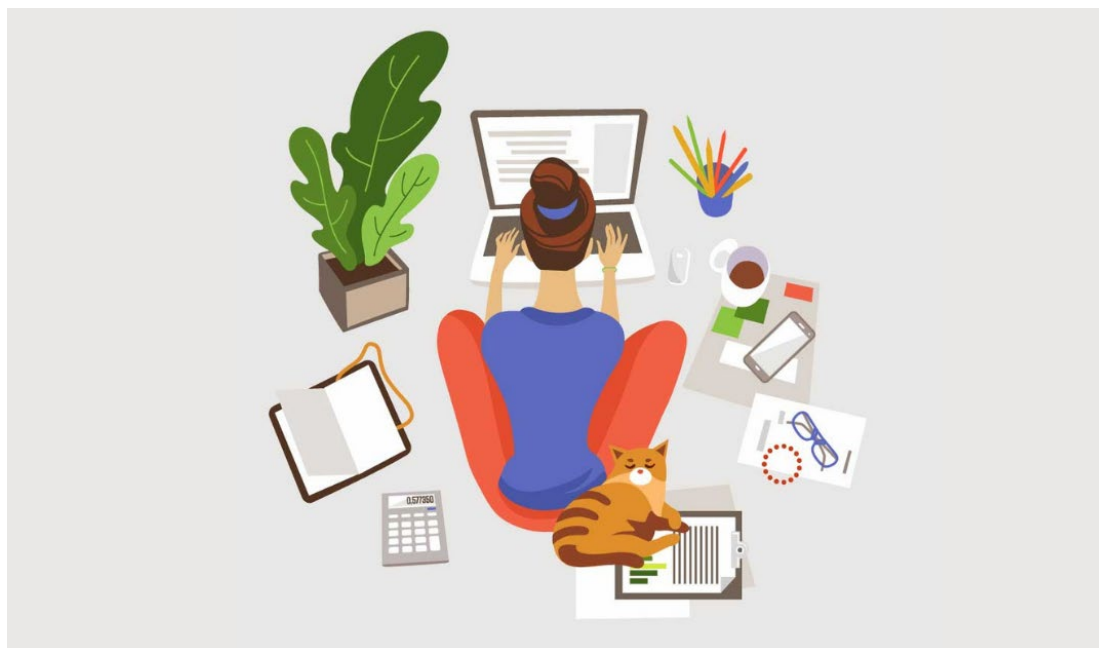
«La professione più attraente della prossima decade sarà quella dello statistico. L'abilità di gestire di dati – di capirli, elaborarli, estrarne valore, visualizzarli, comunicarli – sta diventando una abilità di enorme importanza nelle prossime decenni, non solo a livello professionale, ma perfino a livello di istruzione nella scuola primaria, nella scuola superiore, all'università. Perché ora noi disponiamo di dati gratuiti su quasi ogni argomento. Per cui ***il fattore scarso adesso è l'abilità di capire quei dati ed estrarne valore.***»

Hal Varian, Professore di economia e scienza dell'informazione presso l'Università della California a Berkeley e capo economista di Google (intervista su McKinsey Quarterly, Gennaio 2009)

Studiare statistica

Per studiare statistica con successo è importante avere

- Attitudine all'analisi quantitativa (matematica)
- Attitudine all'uso degli strumenti informatici
- Capacità di risolvere problemi
- Curiosità



Presso l'**Università di Firenze** è attivo un corso di **laurea triennale in Statistica** (l'unico in Toscana) caratterizzato dalla molteplicità dei campi applicativi (economia, demografia, industria, medicina), dai molti contenuti informatici e computazionali, e dal tirocinio curriculare presso aziende o enti pubblici.



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di Economia
e Management

corso di laurea triennale
Statistica

Trova persona, argomento...

[albo ufficiale](#)

[english version](#)

entra in
biblioteca

corso di studio



didattica



docenti



orario e calendari



[matricole](#)

[studenti](#)

[laureati](#)

[aziende](#)

"Il bello di essere statistici è che si può giocare nel giardino di tutti gli altri"

John Tukey

<http://www.statistica.unifi.it>

Presidente: Prof.ssa Emanuela Dreassi

Il corso di laurea in Statistica è attivato dal DiSIA (**Dipartimento di Statistica, Informatica, Applicazioni 'G. Parenti'**) – recentemente riconosciuto dal Ministero come **dipartimento d'eccellenza**

