

Statistica

Tipi di indagini statistiche

Domenico De Stefano

a.a. 2025/2026

Oggetto dell'analisi: la popolazione

L'obiettivo ultimo è conoscere qualche caratteristica di una **popolazione**.
Una popolazione è un collettivo (un insieme di oggetti o individui)

- i componenti del collettivo sono detti **unità statistiche**, sono esempi
 - la popolazione degli italiani di sesso maschile con oltre 18 anni al 01/01/2024;
 - le famiglie italiane al 01/01/2024;
 - i 218 comuni del FVG;
 - i clienti di un negozio.
- La popolazione può essere finita (ad es. la popolazione italiana) o infinita (ad es. tutte le persone iscritte a Scienze Politiche, oggi o in futuro).

Esempi di caratteristiche

I caratteri (**variabili**) da rilevare potrebbero essere:

- per la popolazione degli italiani di sesso maschile con oltre 18 anni al 01/01/2024;
 - l'abitudine al fumo, il comportamento elettorale, colesterolemia totale;
- per le famiglie italiane al 01/01/2024;
 - il reddito familiare, la spesa per consumi, il numero di figli;
- per i 218 comuni del FVG;
 - l'ammontare della popolazione oltre i 65 anni, quota comunale gettito IMU;
- per i clienti di un negozio.
 - la spesa per acquisti, pagamento con carta di credito, età, ...

Indice

- 1 Rilevazioni statistiche
- 2 Tipi di indagini
- 3 Strategie di campionamento

Censimento

- La prima forma di raccolta dati si attua osservando tutti gli individui di una popolazione
 - Questo è un **censimento** (esattamente quello che conduce l'ISTAT ogni 10 anni sull'**intera popolazione residente in Italia** <http://www.istat.it/it/censimento-popolazione>).
- Ci sono problemi nel condurre un censimento:
 - Può essere difficile: ci sono sempre individui difficili da localizzare. **E questi individui potrebbero avere caratteristiche che li distinguono dal resto della popolazione.**
 - Le popolazioni sono in movimento.
 - Fare un censimento è costoso.

Campionamento

Quando non possiamo (vogliamo) osservare l'intera popolazione facciamo ricorso al campionamento.

Campionamento e inferenza

Osserviamo una parte della popolazione, il **campione**, e generalizziamo all'intera popolazione quanto osservato sul campione.

- Campionare è naturale: lo si fa anche in cucina.
- Immaginiamo di cucinare una zuppa: per avere un'idea della possibile riuscita, si fa un assaggio (il censimento non lascerebbe zuppa per la cena).
- Quando si assaggia un cucchiaino di zuppa e si decide che il **contenuto del cucchiaino** non è abbastanza salato, si sta facendo **analisi esplorativa** (ossia usiamo solo strumenti di statistica descrittiva).
- Se si conclude che tutta la zuppa è insipida, si fa **inferenza**.

Campionamento e rappresentatività

- Perché l'inferenza sia valida, l'assaggio deve essere **rappresentativo** dell'intera preparazione.
- Se buttiamo prima il sale, poi tutti gli ingredienti, non mescoliamo mai e assaggiamo la zuppa in superficie, probabilmente non abbiamo un assaggio “rappresentativo”.
- Se buttiamo prima il sale, poi tutti gli ingredienti, poi mescoliamo bene tutti gli ingredienti prima dell'assaggio, probabilmente la rappresentatività dell'assaggio migliorerà.

Esempio: Landon vs. Roosevelt

Un esempio storico di campione non rappresentativo:

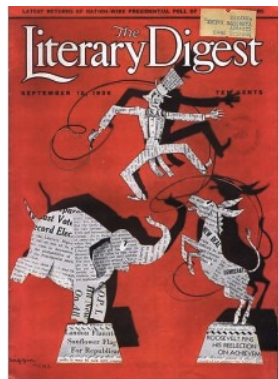


Nel 1936, Alf Landon si propose come candidato repubblicano alla presidenza opponendosi a Franklin Roosevelt, candidato democratico.



L'indagine del *The Literary Digest*

- *The Literary Digest* fu un settimanale pubblicato negli USA dal 1890 al 1938.
- È noto soprattutto per il clamoroso fallimento nel prevedere il risultato delle elezioni presidenziali del 1936.
- *The Literary Digest* contattò circa 10 milioni di americani, ricevendo 2.4 milioni di risposte.
- L'indagine predisse che Landon avrebbe stravinto e che il partito democratico avrebbe avuto solo il 43% dei voti.
- Risultati: il partito democratico vinse, con il 62% dei voti.
- Il giornale fu totalmente screditato e cessò le pubblicazioni dopo poco.



The Literary Digest Poll – cosa andò storto?

- Il “campione” era enorme (2.4 milioni di persone),
- però il giornale contattò
 - i suoi lettori,
 - i possessori di automobili
 - gli utenti telefonici
- Questi gruppi avevano un reddito ben superiore alla media nazionale (era il periodo della grande depressione),
- cioè era un gruppo di elettori molto più probabilmente sostenitori dei repubblicani.
- In altre parole, il campione **non era rappresentativo** dell'intera popolazione.

Campioni grandi sono preferibili, ma..

- The Literary Digest aveva un campione di 2.4 milioni di persone, che è enorme, ma siccome era **distorto**, non produsse previsioni accurate.
- In termini culinari: se la zuppa non è ben mescolata, non importa quanto grande è l'assaggio....

Come dev'essere il campione

Quando diciamo che il campione sono n individui selezionati nella popolazione, questo non vuol dire che qualunque gruppo di n individui vada bene.

Campione “rappresentativo”

Un campione “rappresentativo” è un sottoinsieme della popolazione che ne riflette le caratteristiche.
(Una versione in miniatura della popolazione.)

È il fatto che il campione è rappresentativo che consente di generalizzare i risultati che si ottengono sulla base di calcoli fatti sul campione, alla popolazione.

Come NON dev'essere il campione

NON si ottiene un campione rappresentativo

- prendendo le persone presenti in quest'aula,
- prendendo gli amici/parenti/conoscenti,
- ponendo una domanda in una trasmissione televisiva e invitando il pubblico a rispondere via telefono o sms o internet.

questi gruppi di persone hanno caratteristiche peculiari, non possiamo escludere che queste siano legate alle caratteristiche che stiamo indagando, quindi introdurremmo delle distorsioni.

Per grande che sia, un campione non rappresentativo non consente generalizzazioni.

Campione autoselezionato

In particolare, riportare risultati basati sul **porre una domanda in una trasmissione televisiva e invitare il pubblico a rispondere via telefono o SMS o internet**. è abbastanza usuale.

I risultati vanno visti con molta diffidenza per varie ragioni

- risponderà più facilmente chi ha più a cuore il problema, ovvero ha un'opinione “forte” su esso;
- i rispondenti sono tutti spettatori di quella particolare trasmissione (oltreché spettatori televisivi ecc.).

Come ottengo un campione rappresentativo?

L'idea è di selezionare le unità da includere nella popolazione in modo casuale, poi ci sono diversi metodi

- Il modo più semplice è scegliere n individui in modo che **ciascun individuo della popolazione abbia la stessa probabilità di essere estratto.**
- Altre opzioni sono spesso usate allo scopo di
 - migliorare la rappresentatività,
 - semplificare la procedura (risparmiare quattrini);

tra queste

- campione stratificato,
- campione a grappoli,
- campione a più stadi;

tutti possono essere estratti, le probabilità possono variare.

Indice

- 1 Rilevazioni statistiche
- 2 Tipi di indagini**
- 3 Strategie di campionamento

Esempio 1

Per studiare un nuovo farmaco, viene organizzata una **prova clinica**.

Esempio 1

Per studiare un nuovo farmaco, viene organizzata una **prova clinica**.

Sono reclutati 50 pazienti. A 25 di questi, estratti a caso, viene somministrato il nuovo farmaco; ai rimanenti 25 un placebo (una sostanza inerte che viene somministrata per far credere di aver ricevuto un farmaco).

Esempio 1

Per studiare un nuovo farmaco, viene organizzata una **prova clinica**.

Sono reclutati 50 pazienti. A 25 di questi, estratti a caso, viene somministrato il nuovo farmaco; ai rimanenti 25 un placebo (una sostanza inerte che viene somministrata per far credere di aver ricevuto un farmaco).

Dopo un periodo di tempo, i due gruppi sono confrontati per vedere se il gruppo che ha ricevuto il trattamento mostra effetti positivi.

Esempio 2

Per studiare gli effetti dell'inquinamento atmosferico sulla salute, nel 2004 lo studio MISA (metanalisi italiana degli studi sugli effetti a breve termine dell'inquinamento atmosferico) ha studiato le relazioni tra l'inquinamento e le morti per per cause respiratorie e per cause cardiovascolari nel periodo 1996-2002 in 15 città italiane, scelte tra i principali centri urbani del paese. Complessivamente, lo studio ha coinvolto un totale di 9 milioni e centomila abitanti al censimento 2001.

Definizioni

- Esempio 1 → **esperimento**.

Il ricercatore assegna un “trattamento” ad alcuni individui scelti attraverso un meccanismo casuale. Il punto cruciale è che c'è un trattamento che viene somministrato e un meccanismo casuale per somministrarlo.

- Esempio 2 → **studio osservazionale**.

Il ricercatore “osserva” semplicemente, senza intervenire sui soggetti. Può essere di tipo **retrospettivo**, condotto sulla base di documentazione raccolta in passato e, quindi, già esistente prima della decisione di iniziare lo studio.

Un altro esempio

Uno studio nelle scuole superiori ha mostrato che gli studenti che studiano uno strumento musicale hanno una media più alta di quelli che non lo fanno.

Un altro esempio

Uno studio nelle scuole superiori ha mostrato che gli studenti che studiano uno strumento musicale hanno una media più alta di quelli che non lo fanno.

Tra gli studenti che hanno studiato/studiano musica, il 16% aveva una media più alta di 8, rispetto al solo 5% tra gli studenti lontani dalla musica.

Un altro esempio

Uno studio nelle scuole superiori ha mostrato che gli studenti che studiano uno strumento musicale hanno una media più alta di quelli che non lo fanno.

Tra gli studenti che hanno studiato/studiano musica, il 16% aveva una media più alta di 8, rispetto al solo 5% tra gli studenti lontani dalla musica.

Lo studio prova che studiare musica causa l'aumento della media?

Risposta

È uno **studio osservazionale** (di tipo **retrospettivo**), basato sull'osservazione di scelte (passate) degli studenti e dei loro voti. Questo non può provare una relazione di causa-effetto. Potrebbe semplicemente essere che studenti con voti alti più verosimilmente si dedicano anche alla musica.

Risposta

È uno **studio osservazionale** (di tipo **retrospettivo**), basato sull'osservazione di scelte (passate) degli studenti e dei loro voti. Questo non può provare una relazione di causa-effetto. Potrebbe semplicemente essere che studenti con voti alti più verosimilmente si dedicano anche alla musica.

Un **esperimento** avrebbe assegnato gli studenti a caso ad uno dei due gruppi: studia musica/non studia musica. Poi, dopo un periodo di tempo, avrebbe confrontato la media dei due gruppi. Se questo tipo di studio **prospettico** mostrasse una differenza tra i due gruppi, saremmo probabilmente giustificati nel concludere che esiste una relazione di causa-effetto.

Esercizio

New study sponsored by General Mills says that eating breakfast makes girls thinner

Study: Breakfast Helps Girls Stay Slim

I love these studies....and finding out who sponsored them!

By ALEX DOMINGUEZ, Associated Press

Girls who regularly ate breakfast, particularly one that includes cereal, were slimmer than those who skipped the morning meal, according to a study that tracked nearly 2,400 girls for 10 years.

Girls who ate breakfast of any type had a lower average body mass index, a common obesity gauge, than those who said they didn't. The index was even lower for girls who said they ate cereal for breakfast, according to findings of the study conducted by the Maryland Medical Research Institute. The study received funding from the National Institutes of Health and cereal-maker General Mills.

"Not eating breakfast is the worst thing you can do, that's really the take-home message for teenage girls," said study author Bruce Barton, the Maryland institute's president and CEO.

The fiber in cereal and healthier foods that normally accompany cereal, such as milk and orange juice, may account for the lower body mass index among cereal eaters, Barton said.

The results were gleaned from a larger NIH survey of 2,379 girls in California, Ohio and Maryland who were tracked between ages 9 and 19. Results of the study appear in the September issue of the Journal of the American Dietetic Association.

Nearly one in three adolescent girls in the United States is overweight, according to the association. The problem is particularly troubling because research shows becoming overweight as a child can lead to a lifetime struggle with obesity.

As part of the survey, the girls were asked once a year what they had eaten during the previous three

Che tipo di studio è? *“Le ragazze che fanno regolarmente colazione, specie se a base di cereali, sono più magre di quelle che non fanno colazione, come riporta uno studio che ha seguito 2400 ragazze per 10 anni [...] come parte dello studio, alle ragazze veniva chiesto una volta all’anno che colazione avessero assunto nei precedenti tre giorni.”*

Quale è la conclusione dello studio?

Chi sponsorizzava lo studio?

Che tipo di studio è? *“Le ragazze che fanno regolarmente colazione, specie se a base di cereali, sono più magre di quelle che non fanno colazione, come riporta uno studio che ha seguito 2400 ragazze per 10 anni [...] come parte dello studio, alle ragazze veniva chiesto una volta all’anno che colazione avessero assunto nei precedenti tre giorni.”*

Studio osservazionale, perchè i ricercatori semplicemente osservano il comportamento delle ragazze, anziché imporre un trattamento.

Quale è la conclusione dello studio?

Chi sponsorizzava lo studio?

Che tipo di studio è? *“Le ragazze che fanno regolarmente colazione, specie se a base di cereali, sono più magre di quelle che non fanno colazione, come riporta uno studio che ha seguito 2400 ragazze per 10 anni [...] come parte dello studio, alle ragazze veniva chiesto una volta all’anno che colazione avessero assunto nei precedenti tre giorni.”*

Studio osservazionale, perchè i ricercatori semplicemente osservano il comportamento delle ragazze, anziché imporre un trattamento.

Quale è la conclusione dello studio?

C'è **associazione** tra le ragazze che fanno colazione e il peso.

Chi sponsorizzava lo studio?

Che tipo di studio è? *“Le ragazze che fanno regolarmente colazione, specie se a base di cereali, sono più magre di quelle che non fanno colazione, come riporta uno studio che ha seguito 2400 ragazze per 10 anni [...] come parte dello studio, alle ragazze veniva chiesto una volta all’anno che colazione avessero assunto nei precedenti tre giorni.”*

Studio osservazionale, perchè i ricercatori semplicemente osservano il comportamento delle ragazze, anziché imporre un trattamento.

Quale è la conclusione dello studio?

C'è **associazione** tra le ragazze che fanno colazione e il peso.

Chi sponsorizzava lo studio?

General Mills (multinazionale statunitense del settore alimentare).

Tre possibili spiegazioni

Tre possibili spiegazioni

- 1 Fare colazione rende più magri.



Tre possibili spiegazioni

- ① Fare colazione rende più magri.

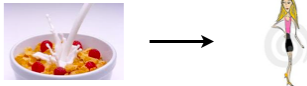


- ② Essere magri spinge a fare colazione.

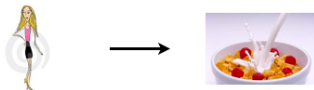


Tre possibili spiegazioni

- 1 Fare colazione rende più magri.

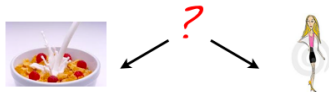


- 2 Essere magri spinge a fare colazione.



- 3 Una terza variabile determina il legame tra fare colazione ed essere magri?

Se l'associazione tra fare colazione ed essere magri è spiegata dal fatto che sia il fare colazione che l'essere magri sono associate ad una terza variabile, siamo in presenza di un **confondente**.



Images from: <http://www.appforhealth.com/wp-content/uploads/2011/08/ipn-cerealfrijo-300x135.jpg>,

<http://www.dreamstime.com/stock-photography-too-thin-woman-anorexia-model-image2814892>.

Riassumendo...

Tipi di indagine

- Censimento.
- Analisi di dati già raccolti per altri fini (es. dati amministrativi).
- Indagini campionarie.

Tipologia del dato

- Studi osservazionali.
- Esperimenti (si interviene sulle u.s.).

Lista non esaustiva!

Ogni tipo di indagine differisce in termini di ammontare di risorse richieste e “forza” dell’inferenza che può essere condotta.

Indice

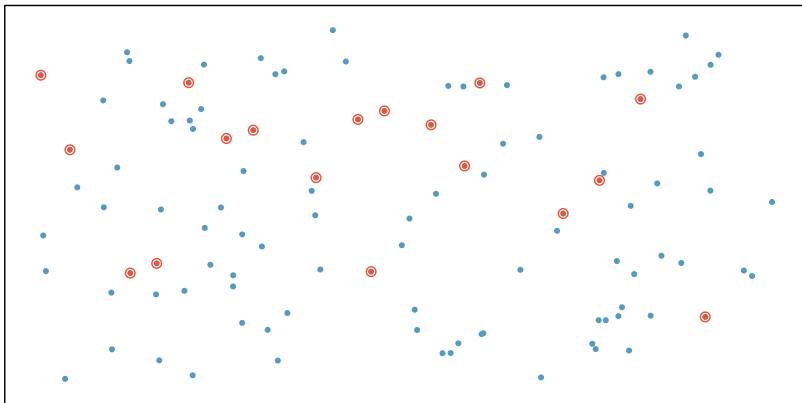
- 1 Rilevazioni statistiche
- 2 Tipi di indagini
- 3 Strategie di campionamento**

Ottenere buoni campioni

- Quasi tutti i metodi statistici sono basati sull'idea di casualità
- Anche negli studi osservazionali si cerca di utilizzare campioni casuali
- Le tecniche più comuni di campionamento sono il campionamento casuale **semplice**, **stratificato**, e **a grappolo**.

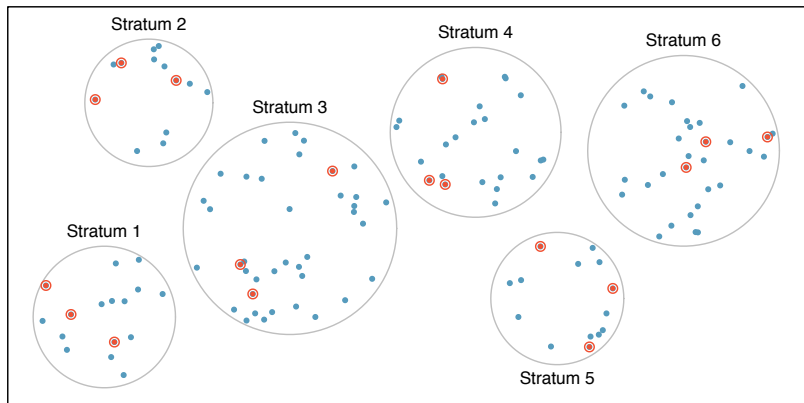
Campionamento casuale semplice (ccs)

Si selezionano le unità statistiche completamente a caso.



Campionamento stratificato

La popolazione è divisa in **strati** di unità statistiche simili. Si prende un ccs da ogni strato.

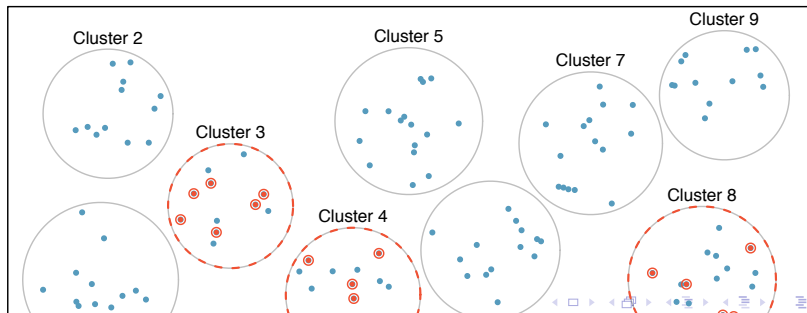


Campionamento a grappolo

La popolazione è divisa in **grappoli** (o gruppi o clusters) e l'unità di campionamento è ora un grappolo di unità della popolazione. Ossia si prende un ccs di grappoli e poi tutte le unità all'interno del grappolo.

Esempi di grappoli sono: le classi scolastiche, i reparti di lavoro, le sezioni elettorali, le famiglie.

Nel **campionamento a stadi**, prima si prende un ccs di grappoli, **poi** un ccs da ciascun grappolo.



Esercizio: l'agente arancio

- L'agente arancio è una mistura erbicida per distruggere la foresta e la vegetazione di superficie. L'agente arancio, chiamato così per il colore dei container in cui viene stoccato, contiene una diossina impura altamente tossica.
- L'agente arancio è stato collegato a patologie cancerose in molti studi epidemiologici.

Nella *Operation Ranch Hand mission* compiuta durante la guerra del Vietnam vengono “spruzzati” oltre 20 milioni di galloni di agente arancio sopra il territorio vietnamita.



Lo studio

- Circa 3 milioni di americani servirono nell'esercito in Vietnam durante la guerra.
- Per studiare l'assorbimento della diossina, nel 1987 le concentrazioni di diossina (in parti per trilione) vennero misurate nel plasma di veterani (soldati di terra).
- Il campione era così composto
 - campione (non casuale) di veterani del Vietnam che servirono nel 1967-1968
 - campione (non casuale) di veterani che servirono in USA e Germania nel 1965-1971

Di che tipo di studio si tratta?

- a) Campione casuale semplice
- b) Studio osservazionale
- c) Esperimento semplice
- d) Esperimento retrospettivo