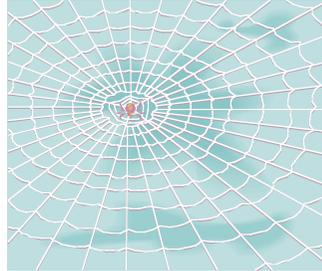


Dei problemi

Il problema è uno stato di cose di cui non siamo soddisfatti e per il quale siamo incerti sul modo, TRA QUELLI POSSIBILI, per portarlo ad una condizione migliore

Il problema si pone se e solo se c'è la volontà di risolverlo e se le azioni perseguibili sono più di una



La soluzione consiste nella scelta della linea di azione più utile ed efficace per il raggiungimento di un obiettivo.

Una soluzione accettabile si ha anche con la dimostrazione che è indifferente la scelta tra due o più linee di azione.

Modello del problema

In ogni problema ci sono degli aspetti (o fattori, o VARIABILI) che sono per noi:

- CONTROLLABILI** E' noto il loro comportamento e si può predeterminarne i valori
- NON CONTROLLABILI** E' noto il loro comportamento, ma i loro valori sono -in larga misura- imprevedibili
- SCONOSCIUTI** Se ne deve postulare l'esistenza, ma non si conosce né il loro comportamento né i loro effetti

Alcuni fattori sono **RILEVANTI** cioè hanno un ruolo anche minimo nel problema, altri sono **IRRILEVANTI**, cioè la loro assenza o presenza non altera la soluzione

Modello del problema/2

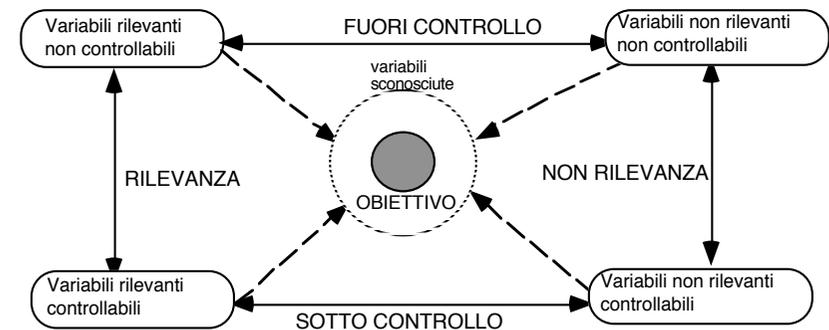
In termini formali possiamo porre la relazione

$$P = f(C; I; S) \quad \text{dove} \quad \begin{cases} P = \text{Una misura della prossimità dell'obiettivo} \\ C = \text{Insieme dei fattori sotto controllo} \\ I = \text{Insieme dei fattori incontrollabili} \\ S = \text{Insieme dei fattori sconosciuti} \\ f = \text{schema dell'interazione tra i vari fattori} \end{cases}$$

Il successo o l'insuccesso di una linea di azione dipende da come interagiscono le variabili rilevanti e da come si pongono rispetto all'obiettivo.

Dipende anche dalla circostanza fortunata che i fattori "S" assecondino o non ostacolino quella linea di azione

Modello del problema/3



Per avere un'idea delle incertezze attribuibili ai fattori "S" immaginate questo schema in tre dimensioni e pensate ai nuovi collegamenti che si possono instaurare.

The Art of Problem Solving

Il ruolo della statistica è qui essenziale perché quasi sempre risolvere un problema consiste nell'effettuare una indagine statistica

La statistica:

Ci guida nella **PRODUZIONE E RACCOLTA, CLASSIFICAZIONE E SINTESI** dei dati.

Risolve le incertezze tra fattori rilevanti e irrilevanti

Fornisce modelli e tecniche per interpretare l'influenza dei fattori sconosciuti

Aiuta nella definizione degli indici di efficienza: **COSTI/BENEFICI**

Permette il riscontro di efficacia **RISULTATI/OBIETTIVI**

The Art of Problem Solving/2

L'uso della statistica non garantisce la soluzione del problema

La combinazione dei dati di qualità ottima unita ad una volontà ferrea di ottenere una risposta non assicura che questa possa essere trovata nemmeno in forma approssimata

D'altra parte, la soluzione potrebbe essere:

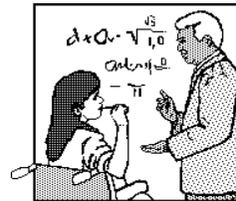
- **INSODDISFACENTE**
- **TARDIVA**
- **NON ACCETTABILE**
- **DISONESTA**
- **GENERATRICE DI ALTRI PROBLEMI PIU' COMPLESSI**

La statistica garantisce solo che i fatti non siano distorti ovvero che la distorsione avvenga in modi trasparenti e ricostruibili (almeno da chi conosce la statistica)

La statistica

La statistica è una scienza che raccoglie tutti i metodi e le tecniche che hanno come obiettivo

- **LA SCOPERTA**
- **LA NEGAZIONE**
- **L'ESTRAZIONE**



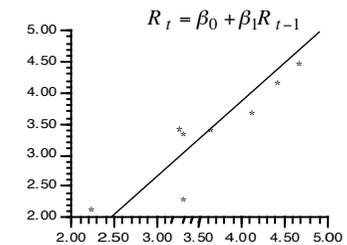
Del contenuto *informativo* di un insieme di dati

Uso della statistica

Le situazioni studiate dalla statistica sono reali ovvero sono connesse a fatti concreti

Il rendimento netto di un fondo azionario al tempo "t" indicato con R_t è legato a quello del tempo "t-1".

t	R_{t-1}	R_t
1990	2.10	2.26
1991	2.26	3.34
1992	3.34	3.29
1993	3.29	3.35
1994	3.35	3.65
1995	3.65	4.13
1996	4.13	4.42
1997	4.42	4.69



I valori mostrano un trend raffigurato da una retta. Stimati i parametri si potrà prevedere quale sarà il rendimento del prossimo anno, noto quello dell'anno attuale. Nel caso in esempio, per il 1998, si passerà da 4.69 a 4.63.

USO DELLA STATISTICA/2

EFFICIENZA NELLA PUBBLICA AMMINISTRAZIONE

La legge sulla trasparenza dei procedimenti amministrativi impone tra l'altro che sia predeterminato il tempo entro cui deve concludersi. Occorre perciò stabilire i TEMPI MEDI e MASSIMI di completamento

Dipendenti P.A. e var. %

Ministero	1987	1991	Variaz. Perc
Presid. Consiglio	5287	6230	+17.8%
Affari esteri	8669	7311	-9.5%
Agricoltura	8355	10172	+21.7%
Ambiente	5	279	istituendo
Beni culturali	25707	24749	-3.7%
Bilancio	263	433	64.6%
Commercio estero	538	536	-0.4%
Difesa	294482	307939	4.5%
Finanze	122439	127432	+4.1%
Giustizia	69844	77283	+10.5%
Industria	1462	1434	-1.9%
Interno	121718	146671	+20.5%
Lavori pubblici	4469	4922	+10.2%
Lavoro	15445	15924	+3.1%
Maria Mercantile	1814	2062	+13.7%
Partecip. Statali	137	129	-5.8%
Università	93702	107328	+14.5%

Altre occasione è il controllo del personale e la sua ripartizione efficienti in base alla produttività

Nell'esempio è aumentato il personale dell'agricoltura ed è diminuito quello dell'industria. Non sembra che questa fosse la tendenza giusta

USO DELLA STATISTICA/3

MARKETING

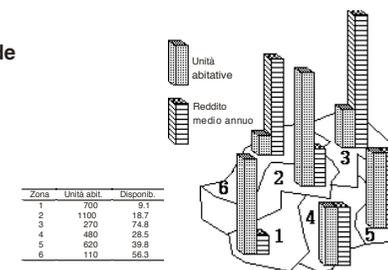
L'obiettivo del marketing è il raccordo della domanda di beni e servizi con la capacità di soddisfarle da parte delle imprese.

Lo studio delle esigenze dei clienti, delle abitudini di consumo, della pubblicità, della strategia rispetto alla concorrenza etc. impongono la conoscenza e la trattazione di informazioni quantitative

ESEMPIO:

La localizzazione di un nuovo punto vendita richiede tra l'altro la zoning di un area secondo il numero di unità abitative ed il reddito

In questo fase la statistica trova largo impiego



Dove non c'è statistica

Certamente la statistica ha poco a che vedere con gli "statisti" e con la "statica", ma ci sono anche altri casi

CASI LIMITE

Monterone (Co) è il comune più piccolo d'Italia (29 ab.) ed è una curiosità per gli studiosi di statistica.

I casi isolati o i casi singoli non interessano la statistica che infatti si presenta come scienza dei collettivi

ACCOSTAMENTI FORZATI

La gazetta dello sport ha un angolo intitolato: "Per gli amanti della statistica" dove riporta dati relativi ai precedenti incontri tra due squadre.

Non si capisce bene quale sia il collegamento se non che i due club hanno lo stesso nome

Dove non c'è statistica/2

GENERALIZZAZIONI SEMPLICISTICHE

La statistica si ritrae quando ci si avventura in estrapolazioni non suffragate da riscontri fattuali

GIORNALISMO NON SPECIALIZZATO

Le statistiche dicono che il 23% dei lettori legge il giornale in meno di 15 minuti.

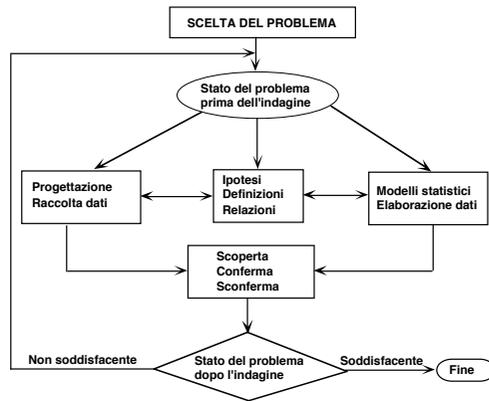
Sembra che l'autore si chiami fuori e che "le statistiche" non siano un supporto non essenziale all'articolo.

COMPILAZIONE DI TABELLE

La tabellazione come raccolta organizzata di dati è essenziale alla statistica, ma non è la statistica.

Il paradigma di lavoro

Si effettua un'indagine statistica per dare sostegno a teorie incerte



L'insieme delle conoscenze teoriche ed empiriche ed un SANO scetticismo aiutano a spiegare le variazioni tra due stati: prima e dopo l'indagine

L'indagine statistica

Se la trattazione del problema costringe a cercare nuovi dati, questi debbono essere rilevati con uno schema appropriato.

La rilevazione si articola in una sequenza ordinata di casi o repliche che hanno tanti elementi in comune da essere considerati facenti parte di un unico processo: l'indagine statistica.

Ogni indagine ha il suo piano di realizzazione legato alle peculiarità della disciplina in cui il problema è sorto.

Nuove rilevazioni

L'acquisizione di nuovi dati è dovuta al fatto che:

 La base informativa di un problema non è soddisfacente

 E' utile e praticabile realizzarne una nuova o integrare quella esistente

La rilevazione dei dati consiste nella annotazione sistematica, precisa e impersonale della modalità delle variabili riscontrate sull'unità

Le rilevazioni possono essere classificate in vario modo. Quella più rilevante è la distinzione tra TOTALI e PARZIALI:

TOTALI: coinvolgono tutti gli elementi di una popolazione

PARZIALI: la rilevazione è estesa solo ad una parte, comunque scelta, di popolazione

Le rilevazioni totali

Le RILEVAZIONI TOTALI (O CENSIMENTI) sono quelle in cui sono enumerate o misurate tutte ed indistintamente le unità della popolazione

All'interno delle totali si hanno:

 **RILEVAZIONI GENERALI:** riguardano la rilevazione di tutte le unità rispetto alle variabili di interesse (POPOLAZIONE)

Esempio: un'indagine sul voto che si rivolga a tutti gli elettori di qualsiasi sesso e regione di residenza

 **RILEVAZIONI SPECIALI:** riguardano la rilevazione delle sole unità rispondenti a certe specifiche (SOTTOPOPOLAZIONE)

Esempio: un'indagine sul voto che si rivolga a tutti, ma i soli iscritti alle camere di commercio come "artigiani"

Le rilevazioni totali/2

- Supponiamo di considerare come unità gli stabilimenti industriali attivi in Calabria per poi circoscrivere l'attenzione a quelli con più di 50 addetti (popolazione TARGET o teorica)



Quello che accomuna le due indagini è che sono enumerate **TUTTE** le unità che formano la popolazione o la sottopopolazione

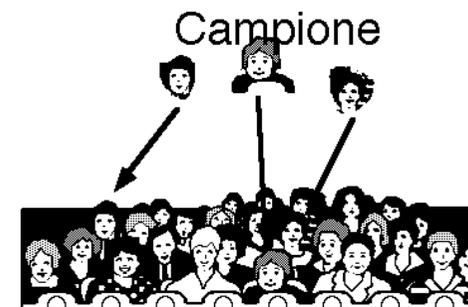
Si tratta perciò di **Popolazioni (o sottopopolazioni) FINITE E CENSIBILI** cioè la cui rilevazione può effettivamente cominciare e finire in tempi e a costi praticabili

Le rilevazioni parziali

Sono limitate solo ad una parte delle unità della popolazione (o sottopopolazione) scelta in base ad opportuni criteri. La parte esaminata si chiama **CAMPIONE**.

La riduzione delle unità propria del metodo **CAMPIONARIO** è valida solo se permette il raggiungimento di risultati molto prossimi di quelli ottenibili con la **TOTALE**.

TOTALE/PARZIALE NON E' UNA COTRAPPOSIZIONE, MA UNA COMPLEMENTARITA'



Esperienze consolidate in molti paesi e in molte discipline dimostrano che si può dare pieno affidamento ai campioni purché scelti con accuratezza.

Il Campione

L'analisi del campione è meno costosa, più precisa, più asettica, più controllabile e più rapida dell'esame della rilevazione totale.

I censimenti generali si limitano alle variabili fondamentali lasciando ai campioni il compito di scendere nei dettagli.



Il discorso statistico ha due piani: a livello di popolazione e a livello di campione. Il primo è la nave che quando naviga lascia vedere solo la parte che galleggia: il campione.

Osservando e analizzando la parte visibile si conoscerà anche la parte che è sotto l'acqua.

Le ragioni del campione

Nel corso di un'indagine ci si può accorgere che la **RILEVAZIONE TOTALE** non è praticabile perché:

- **HA UN COSTO ECCESSIVO O RICHIEDE GRANDI ORGANIZZAZIONI**

Esempio: il censimento generale si realizza ogni 10 anni

- **RICHIEDE TROPPO TEMPO**

Esempio: l'intervista di tutti i lavoratori dipendenti richiederebbe tanti anni che una volta finita la popolazione attuale sia molto diversa dalla censita

- **E' TEORICA: PARTE DELLE SUE UNITA' NON ESISTE ANCORA o NON ESISTE PIU'**

Esempi: il controllo della qualità dovrebbe riguardare anche le unità non ancora prodotte.
Le vestigia di antiche civiltà

Le ragioni del campione/2

E' SUPERFLUO

Esempio: per valutare l'inquinamento di un tratto di mare non bisogna esaminarne tutta l'acqua

E' RISCHIOSO

Esempio: prima di somministrare un nuovo vaccino all'intera popolazione è bene sperimentarne su poche unità la tossicità e l'efficacia

E' DISTRUTTIVO

Esempio: per stimare la durata di una batteria di 1.5 volt si deve per forza farla esaurire, ma se si esauriscono tutte, che cosa si vende

E' RAPIDO

Esempio: per conoscere i risultati delle elezioni non si possono aspettare gli esiti ufficiali. Più immediati e precisi possono essere gli EXIT POLL

Le ragioni del campione/3

LA POPOLAZIONE NON E' PIU' SONDABILE

Esempio: si pensi alle registrazioni delle nascite e morti presso le curie del nel '600 in gran parte andate perdute.

PROVOCA TURBATIVE NELLA POPOLAZIONE (E' INVASIVO)

Esempio: la rilevazione alle frontiere di tutti i forestieri causerebbe code ed i ritardi indurrebbero le unità a rinviare o rinunciare all'entrata

NON E' ACCURATA (POCHI DATI SI CONTROLLANO MEGLIO)

Esempio: una certa misurazione è stata registrata con due cifre decimali che sono poi risultate insufficienti. Occorrerebbe ripetere tutto e non sempre le rilevazioni sono semplici

APPLICAZIONI DEL CAMPIONE

il campione è una parte della popolazione utilizzata per conoscere l'intera popolazione

Tecniche di campionamento si impiegano ad esempio in ...

-  Sondaggi elettorali; gradimento delle amministrazioni locali; consenso alle scelte politiche governative.
-  Ricerche di mercato: accettazione di un nuovo prodotto; apprezzamento della modifica di un prodotto conosciuto; desiderio di un nuovo servizio.
-  Controllo della qualità: aderenza agli standard di un item; verifica della integrità di una fornitura; certificazione della composizione di un prodotto.
-  Indagini di laboratorio: efficacia di un fertilizzante; pericolosità di un farmaco; validità di terapie comportamentali; tolleranza ad un prodotto.
-  Imprenditoria: pagamento di *royalties*; diffusione di quotidiani e settimanali; *audience* televisiva; revisione dei conti.

Tipologia di osservazione

La rilevazione è il rapporto che si instaura tra chi -consapevolmente- osserva ed i soggetti osservati.

Una prima utile distinzione tra i diversi tipi di rilevazione passa per il legame che può intercorrere tra agenti attivi e passivi della rilevazione.

-  **ESTERNA** si annotano i fatti come i sensi li percepiscono o i sensori li avvertono, senza tentare intervenire sulle unità (tecniche non invasive).
-  **PARTECIPANTE** l'osservatore vive nella collettività che studia e interagisce con i soggetti che ne fanno parte modificandone in qualche modo le tendenze naturali (effetto Hawthorne).

Rilevazione isolata e ripetibile

Un'altro utile distinguo è tra manifestazione



ISOLATA

si effettuano in relazione ad un fatto -volontario o involontario- che non può riaccadere oppure che è costoso o vietato provocare.

Questi di solito non interessano la Statistica



RIPETIBILE

Ogni fatto è unico ed è impossibile replicarlo, ma alcuni elementi essenziali rimangono intatti nelle varie manifestazioni:

tutte le volte che si configura un insieme di circostanze determinate si possono osservare certe conseguenze.

Survey ed esperimenti

All'interno delle rilevazioni ripetibili è importante la differenza tra:



SURVEY

si analizzano eventi che non si possono provocare a volontà, monitorati man mano che si verificano secondo cadenze predeterminate. Fenomeni metereologici, economici, demografici, finanziari



ESPERIMENTO

si creano situazioni di studio artificiali programmate in modo che i risultati possano rispondere a precise domande del tipo causa/effetto.

L'analisi osservativa

Ci interessano rilevazioni **ESTERNE** su fenomeni che si **RIPETONO** spontaneamente o che seguono flussi regolari. Una parte ha il ruolo dei fattori controllabili dell'analisi sperimentale

L'analisi osservativa o **INDAGINE STATISTICA** è una aggregazione di entità elementari dette **OSSERVAZIONI**.

L'osservazione è composta da **DATI**: "ciascuno degli elementi di fatto (notizia, comunicazione, messaggio, rilevazione strumentale) utilizzabile per la soluzione di un problema"

Esperimenti in senso statistico



Casi unici

Fatti che non possono riaccadere perché speciali e isolati o perché le condizioni loro antecedenti non possono essere ripristinate.



Accadimenti ripetibili

Fatti che possono replicarsi spontaneamente oppure essere indotti artificialmente (fatte salve certe condizioni e garanzie).

La Statistica tende ad ignorare i primi e ad occuparsi dei secondi per quelle parti soggette a variazioni sensibili nell'evoluzione del fenomeno

Elementi costitutivi del Dato

La statistica è centrata sul dato che studiamo nei suoi elementi costitutivi:

- L'UNITA'
- LA VARIABILE
- LA SCALA DI MISURAZIONE
- IL CRITERIO ORGANIZZATIVO

ESEMPIO

Nell'idea che i disavanzi delle aziende pubbliche si concentrino in particolari regioni a fianco c'è la tabella che li riporta, in milioni, per alcune regioni.

La caratterizzazione dei dati è ora: {Regione, Disavanzo, Milioni di lire, Ordinamento alfabetico};

Regioni	Disavanzo
Abruzzi	110558
Calabria	49991
Campania	2189901
Emilia R.	478704
Lazio	2739464
Liguria	378193
Lombardia	1111113
Marche	83445
Piemonte	342798
Puglia	360113
Toscana	562888
Umbria	143723
Veneto	600062
Totale	9150955

L'unità statistica

L'unità è il soggetto elementare cui l'indagine si rivolge: una persona fisica oggetto, azienda, o un gruppo di entità che, dal punto di vista dell'indagine, formino un tutt'uno.

Le unità devono essere obiettivamente distinguibili e deve pure essere stabilito quali siano quelle che interessa rilevare e quali debbano invece tralasciarsi.

ESEMPI

- a) Interessi maturati su di un conto corrente (Il conto corrente)
- b) Tipo di riscaldamento di un appartamento (L'appartamento)
- c) Numero di testi consigliati in un corso (Il corso)
- d) Emissione di gas tossici da un automobile (L'automobile)
- f) Numero di arresti per agente di polizia (L'agente)

Problemi di definizione

INDAGINE SULLE FAMIGLIE

Come si considerano i "single", le coabitazioni, le comunità?

PUBBLICITA' TURISTICA

Non è raro leggere o sentire messaggi promozionali del tipo: 30 giorni di sole nel mese X. Il problema è capire cosa si intende per "giornata di sole": ad esempio nelle ore diurne una sequenza di almeno otto ore di sereno e senza nebbia.

SONDAGGI PRE-ELETTORALI

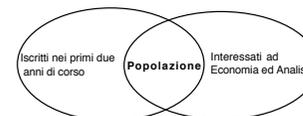
Un'intervista telefonica agli abbonati di "La Gazzetta del Sud" può solo indicare come la pensano gli abbonati che hanno risposto alle telefonate.

La popolazione

Dicesi popolazione o UNIVERSO l'insieme di tutte e solo le unità che si è interessati ad osservare in una certa indagine.

ESEMPIO:

Alcuni studenti intendono finanziare le spese di frequenza universitaria avviando un programma di ripetizioni ben fatte ed a basso costo. Quale sarà la popolazione?



E' chiaro che non possono essere tutti gli studenti iscritti. Ci si può limitare agli studenti dei primi due anni.

Occorre poi determinare le materie per cui esistono le competenze: diciamo i corsi fondamentali di economia e matematica.

La delimitazione dell'universo è ora chiara: studenti del biennio che non hanno sostenuto economia e/o analisi.

La popolazione/2

NON è un gruppo di persone che risiedono in una certa zona.

il termine POPOLAZIONE ha una accezione più ampia e più astratta : tutte e solo le unità che hanno in comune una o più proprietà rilevanti per il problema.

La caratteristica unificante deve essere evidente cosicché il riconoscimento avvenga con il minimo di incertezza tenuto conto delle difficoltà create da unità congiunte o sfocate.

Se disponiamo di un elenco del fatturato di 500 imprese edili ciò che studiamo non è la popolazione delle imprese, ma la popolazione dei fatturati.

Tipologia delle popolazioni

La popolazione è un insieme e come tale può essere:



FINITA

Se include oggetti che possono essere contati ed il conteggio, ad un certo punto si interrompe.

Esempi: le pagine di un libro, i diplomati di una scuola



ENUMERABILE

Le unità sono contabili, ma il conteggio non si interrompe mai

Esempi: i numeri naturali, i lanci di un dado



INFINITA

Ogni sottoinsieme di popolazione contiene lo stesso numero di entità contenute nella popolazione.

Esempi: le frazioni tra zero ed uno, le nuances di un colore,

Popolazioni indeterminate ed elusive

Non sempre è nota o determinabile la numerosità della popolazione



INDETERMINATE

L'insieme dei soggetti è finito in quanto esiste un limite fisico non valicabile alla sua crescita, ma le unità sono sparse o rare al punto da rendere impossibile il loro materiale censimento.

Esempi: animali selvatici, tifosi di una squadra, gruppi etnici o religiosi particolari



ELUSIVE

Composte da unità che hanno buone ragioni per non farsi censire. Per queste non si potrà mai essere sicuri che le unità individuate siano tutte o solo una parte perché le altre rimangono nascoste

Esempi: extracomunitari senza documenti, tossicodipendenti, affetti da malattie infettive, affiliati alla onorata società, vincitori di grossi premi alle lotterie, gli idraulici nel mese di agosto.

Popolazioni censibili e virtuali

E' censibile la popolazione le cui unità possono essere esaminate in tempi e costi "ragionevoli"

E' virtuale la popolazione infinita e la enumerabile. Anche la finita lo può essere se la disamina delle unità è costosa, difficile, superflua, impossibile.

Esse esistono solo in via teorica e debbono essere censite comunque in poche unità (campione).

- 1) *Le popolazioni preistoriche possono essere analizzate solo attraverso i pochi resti che gli scavi portano alla luce.*
- 2) *Il controllo di qualità non riguarda solo quello che si è già prodotto, ma anche quello che si produrrà, che però non è ancora censibile.*
- 3) *I risultati di ogni esperimento sono in realtà solo una parte delle infinite repliche che si potrebbero effettuare.*

Microdati e macrodati

L'unità per cui si cercano i dati (unità di rilevazione) non sempre coincide con quella oggetto di studio (unità di indagine)

Esempio:

La rilevazione delle scuole materne può essere effettuata per comuni, ma essere poi elaborata per province

I microdati sono i valori riferiti all'unità elementare che non può essere ulteriormente scomposta.

I macrodati sono i valori ottenuti o direttamente o dalla aggregazione di più dati elementari.

I microdati sono un sistema di rilevazione comodo quando non si è sicuri della scala di aggregazione che poi potrà servire

La variabile

E' l'aspetto si intende studiare nel dato.

Può essere una distanza, una numerosità, una forma, un atteggiamento, un grado od anche una composizione di caratteristiche da trattare in modo aggregato.

I simboli più diffusi sono:

X, Y, W, Z

Che sono la codifica della variabile

La codifica è l'espressione abbreviata con cui le informazioni sulle variabili acquisite dalle unità sono trasferite sui supporti di elaborazione o nei ragionamenti astratti

La variabile/2

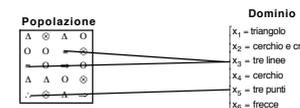
Perché una generica qualità o quantità sia definita "variabile" occorre...

- **ATTINENZA** con la realtà di interesse la cui comprensione aumenta (anche di poco) per la disponibilità di dati sulla variabile.
- **ESSERE SOGGETTA A VARIAZIONI**: cioè possa presentarsi con almeno due valori o categorie distinte nell'ambito della popolazione.
- **ESSERE ACCERTABILE** e cioè essere rilevabile strumentalmente senza ambiguità

Si presuppone inoltre che la variabile possa essere osservata/misurata in modo separato da altre variabili che pure incidono sull'unità.

il dominio della variabile

Individuata la variabile occorre definire l'insieme di tutti e solo i valori o modalità della variabile X (il dominio) riscontrabili nella popolazione:



Ad ogni unità della popolazione sarà associata una ed una sola modalità del dominio.

In questo caso, una delle sei diverse forme presenti. Unità diverse possono presentare la stessa modalità

il dominio della variabile è un insieme di "k" elementi con "k" finito od infinito

$$S = \{X_1, X_2, \dots, X_k\}$$

L'abbinamento unità/modalità si effettua confrontando ciascuna delle unità è con il dominio "S" ed associandola ad una delle X_i in base ad una regola di classificazione o misurazione.

il dominio della variabile/2

Perché non insorgano ambiguità è necessario che le modalità siano

UNIVOCHÉ: sia possibile osservarne una sola per ogni unità e sia subito chiaro quale

ESAUSTIVE: non sia possibile osservarne di diverse da quelle già in S

RIPRODUCIBILI: la rilevazione dovrà dar luogo sempre allo stesso schema di attribuzione.

a) *Incompatibilità:* $X_i \neq X_j$ per ogni $i \neq j \in X_i, X_j \in S$

b) *Esaustività:* per ogni $u \in P X(u) \in S$;

c) *Riproducibilità:* $X = X_i$ se e solo se $X(u) = X_i$

Dominio chiuso o aperto

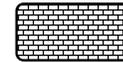
L'insieme dei valori ammissibili "S" può essere



APERTO

Quando il fenomeno descritto non ha un limite minimo e/o massimo ben definito prima che sia completata la rilevazione

Esempio: Reddito (che può anche essere negativo)



CHIUSO

Quando le sue modalità sono definite e note in anticipo e non possono cambiare durante la rilevazione

Esempio: Stato civile

il dominio aperto comporta problemi di elaborazione. Quello chiuso consente dei controlli di validità dei dati

Analisi univariata e multivariata

Ogni problema è una ragnatela: se si tocca un filo tutti gli altri vibrano. Lo stesso succede per le variabili.

Lo studio univariato ha solo scopo didattico. Nella pratica i dati sono sempre multivariati

ESEMPIO: dove vanno gli studenti

Studio/Residenza	Nord		Centro		Sud		Totale	
	numero	%	numero	%	numero	%	numero	%
-----	286999	83,6	178642	50,7	250887	74,7	719.124	81,6
-----	18783	5,5	1926	0,8	8378	2,5	28687	3,3
-----	27308	8,0	4749	2,4	11312	3,3	43369	4,9
-----	9149	2,7	3396	4,8	38800	11,4	47345	5,5
-----	929	0,3	2796	1,4	27296	8,0	30981	3,5
-----	6.164	1,8	1842	0,5	36786	10,3	46032	5,3
-----	34224	100,0	197119	100,0	339663	100,0	879206	100,0

La lettura della tabella non è difficile. Lo è la generalizzazione dei risultati

Gli studi multidimensionali sono al momento rinviati. Faremo invece solo studi univariati.

Col presupposto che si possa avere l'idea di un concetto multilaterale studiandone separatamente le componenti

La definizione operativa

E' l'insieme di regole con cui classificare un concetto, determinarne la misura o, in generale, per agganciarlo alla realtà osservabile.

ciò impone la definizione operativa solo con variabili di cui sia possibile seguire con facilità il meccanismo di conversione di una proprietà delle unità in una categoria o valore del dominio.

ESEMPIO

il concetto di interconnessione tra due centri abitati, diciamo "A" e "B", è misurato con la semisomma degli automezzi che si sono spostati da "A" a "B" e quelli che da "B" sono andati ad "A".



Classificazione e misurazione

L'acquisizione dei dati può avvenire classificando in categorie distinte la proprietà di cui l'unità è portatrice oppure misurandola in base ad una determinata unità di misura.

con la CLASSIFICAZIONE si identificano l'unità (e le modalità numeriche del dominio sono equivalenti ad ogni altro insieme di simboli);

con la MISURAZIONE si quantifica una proprietà posseduta ed i numeri sono utilizzati in quanto inseriti in un sistema di numerazione.



UN ATTEGGIAMENTO NON SELETTIVO

Classificazione e misurazione/2

La classificazione e la misurazione possono scaturire da due procedure di assegnazione dei valori: enumerazione delle unità rispetto alla proprietà posseduta

oppure comparazione della proprietà studiata rispetto ad un ventaglio di possibilità che, identico per tutte le unità, non dipende né dal numero.

		ASSEGNAZIONE VALORI	
		Enumerazione	Comparazione
RILEVAZIONE DEL DATO	Classificazione	Nominazioni	Scala nominale
	Misurazione	Graduatorie	Scale ordinali semplici Scale ordinali graduate Scale intervallari Scale proporzionali

Nominazioni e Variabili nominali

Le modalità di queste variabili esprimono categorie, qualità, status: le $\{X_i\}$ in "S" hanno la sola funzione di etichettare le unità per formarne un elenco o per raggrupparle in classi omogenee.

ESEMPI:

Nominazione: La variabile "Regione" si manifesta con le usuali 20 modalità $S=\{\text{Calabria, Sicilia, ..., Val d'Aosta, Piemonte}\}$.

variabile nominale: Un'impresa può ricadere nel settore {agricoltura, industria, altre attività}.

Le differenze possono essere accertate, ma non ordinate né misurate: si possono scambiare di posto senza che ciò influisca sulla validità della classificazione

Uso dei numeri

la codifica delle modalità porta ad usare dei numeri. Questo però non significa che siano lecite delle operazioni aritmetiche:

i ruoli di una squadra di calcio sono indicati con dei numeri, ma non si può dire che l'ala sinistra ("11") sia maggiore dello stopper ("5") o che l'unità di misura "1" dei calciatori sia il portiere;

ESEMPI

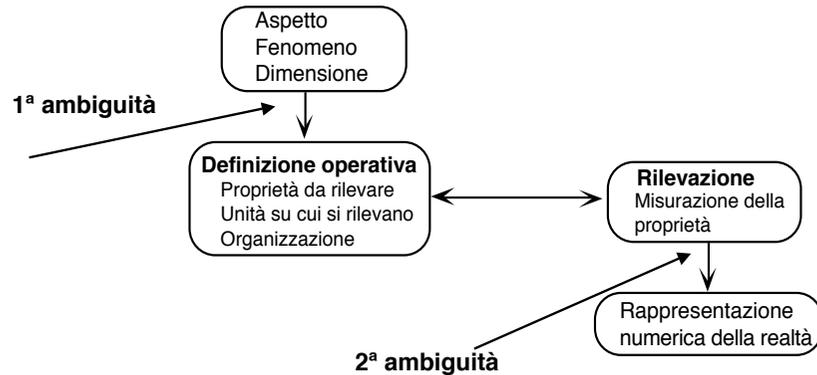
il numero civico delle abitazioni:



Non ha significato la eventuale progressione delle modalità;

Tecniche di misurazione

il concetto di misurazione è uno dei più controversi tanto che oggi, dopo più di 50 anni, il dibattito è sempre aperto.



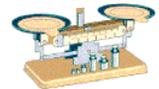
La scala di misurazione

Qui esiste una sovrapposibilità tra una categoria e la successiva che, oltre a contenere quella che la precede, vi aggiunge un quantum di proprietà che la differenzia dalla prima senza cancellarla, anzi inglobandola

ciò che distingue le scale di misurazione è il diverso grado di formalizzazione che si può dare al meccanismo dell'aggiunta del quanto di proprietà.

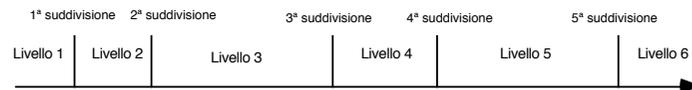
1. Ordinamento tra valori senza distinguibilità degli scarti;
2. Ordinamento tra valori con ordinamento degli scarti;
3. Quantificazione dei valori con parità tra scarti: $7-5=3-1$;
4. Quantificazione dei valori con parità tra rapporti: $8:4=6:3$.

Scala=bilanciamento



Continuo percettivo

L'intensità con cui si avverte una sensazione varia in una successione continua di stati: al tessuto uniforme del concetto si sovrappone una griglia più o meno regolare



una unità che sia X_i in una rilevazione ed X_j in una successiva con $X_i < X_j$ sarà passata per tutti gli stati intermedi tra X_i ed X_j .

Le suddivisioni non sono però oggettive: osservatori diversi scelgono divisioni diverse ovvero lo stesso punto di separazione ha senso diverso.

N.B. Talvolta la proprietà studiata ha natura discontinua: si modifica con una scansione non frazionabile per un numero finito di stati che sono i soli a poter essere osservati.

Ordinamenti

Il termine "scala" ha senso se tra le modalità di "S" sono possibili degli ordinamenti.

- 1) $X_i < X_j$ oppure $X_i > X_j$ per ogni $i \neq j$
- 2) $X_i < X_j \Rightarrow X_i \neq X_j$
- 3) $X_i < X_j$ e $X_j < X_k \Rightarrow X_i < X_k$ per ogni $i < j < k$

Maggiore è il contenuto di "fenomeno" maggiore è la modalità che la rappresenta; esiste perciò una disposizione delle modalità che non può essere alterata senza che ne risulti modificata la rilevazione.

Il dominio si esprime con interi consecutivi:

$$S = \{a, a+1, a+2, \dots, a+k-1\}$$

Graduatorie

Le modalità "S" sono i ranghi corrispondenti alle posizioni in graduatorie delle unità per i valori possibili sono dati dalla numerosità della rilevazione.

$$S = \{1^a, 2^a, \dots, k^a\}$$

Il processo di misurazione è ad un livello molto superficiale, con possibilità elaborative limitate, essenzialmente basate su confronto e sintesi delle posizioni che le unità occupano rispetto a variabili diverse.

ESEMPIO

Per stare in testa occorrono buone posizioni su entrambe le graduatorie

Stud.	Grad. Scritto	Grad. Orale	Totale
A	3	1	4
B	2	3	5
C	1	7	8
D	7	2	9
E	5	4	9
F	4	6	10
G	6	5	11

Variabili ordinali

i ranghi sono dei voti che esprimono la stima della proprietà posseduta: ogni unità è confrontata con una linea di valutazione che incasella l'unità in una data categoria di valore a prescindere da quello che succede alle altre unità.

Spesso, le modalità di una variabile ordinale esprimono soglie di vicinanza ad un ideale che fungerebbe da "metro" o "campione" di misurazione del concetto.

ESEMPI:

a) Voti di un giudice: $S = \{0, 1, 2, \dots, 10\}$;

b) Ammontare di punti da ripartire: $\{0 - 100\}$;

c) Voti grafici: $\{++, +0, 0+, 00, -0, 0-, --\}$;

d) Quantificatore verbale: $\{ \text{pianura, collina, montagna} \}$

Invarianza rispetto a trasformazioni monotone

$$f(X_i) < f(X_j) \text{ se } X_i < X_j$$

Numero di modalità e Posizione

Non esiste un numero ottimale di livelli: $k=7 \pm 2$ o $k=6$ sono considerati uno standard nelle ricerche di mercato (Kinnear e Taylor, 1979, p. 30, Malhotra 1996, p. 298).

3 o 4 gradini comportano risultati confusi per l'accorpamento di giudizi eterogenei; d'altra più di sei è utile solo per acquisire variazioni di quantità molto piccole di cui non sempre si ha bisogno.

Anche la disposizione deve essere equilibrata:

ESEMPIO:

Quale delle tre seguenti moltiplicazioni

P1. $9 * 7 * 8 * 6 * 5 * 4 * 3$

P2. $3 * 4 * 5 * 6 * 7 * 8 * 9$

P3. $7 * 3 * 8 * 4 * 6 * 5 * 9$

Effetto
posizione

darà il risultato più alto?

Differenziale semantico

Per attenuare le ambiguità derivanti delle scale ordinali si possono usare delle scale bipolari in cui sono inserite solo le valutazioni più opposte dell'aspetto indagato collocando tra di esse, ad opportune interdistanze, una serie di riquadri

Chi risponde dovrà poi individuare il punto più prossimo al suo giudizio ovvero indicare quale descrizione numerica o verbale si adatti al proprio sentire.

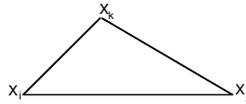
ESEMPIO:

"Come giudicate l'operato dei rappresentanti degli studenti nel Senato accademico integrato?"



Scale metriche

Date tre qualsiasi modalità di "S" allora



$$d(X_i, X_j) = 0 \text{ se e solo se } X_i = X_j; \quad \text{Identità}$$

$$d(X_i, X_j) > 0 \text{ se } X_i \neq X_j; \quad \text{Positività}$$

$$d(X_i, X_j) = d(X_j, X_i); \quad \text{Simmetria}$$

$$d(X_i, X_k) + d(X_k, X_j) \geq d(X_i, X_j); \quad \text{Disuguaglianza triangolare}$$

Se il dominio della "X" verifica le quattro condizioni allora su di esso si applicano, sia pure con qualche distinguo, tutte le procedure statistiche.

Scale intervallari

Sivaluta ciò che succede al fenomeno ponendolo in relazione con un movimento lungo un'asta graduata.

Le tacche sono regolari e separate -al livello minimo- da una unità convenzionale che può essere variata senza interferire con ciò che si misura.

L'origine o punto-zero della scala ha un ruolo marginale dato che agisce solo come riferimento e può essere sostituita, senza alcuna conseguenza sull'esito della rilevazione.

Un incremento assoluto tra due misurazioni ha lo stesso significato qualunque sia il livello da cui si calcola l'incremento.



Scale intervallari/2

La differenza tra 40C e 30C gradi è la stessa di quella tra 30C e 20C, ma non si può dire che ad 40C faccia due volte più caldo che a 20C. Se le temperature sono convertite in gradi Fahrenheit si avrà:

$$30C \rightarrow \frac{9 * 30}{5} + 32 = 86F; \quad 40C \rightarrow \frac{9 * 40}{5} + 32 = 104F; \quad 20C \rightarrow \frac{9 * 20}{5} + 32 = 68F;$$

La differenza tra due temperature ha lo stesso significato qualunque sia il livello, ma nessuna asserzione può farsi sul loro rapporto dato che C=0 o F=0 non significa "totale assenza di calore".

Se la "X" è misurata su scala intervallare è lecito -se preferibile- usare in sua vece la variabile ottenuta come trasformazione lineare.

$$Y = a + bx; \quad b > 0$$

Scale proporzionali

Ad un incremento relativo nella misura, corrisponde un incremento relativo di eguale entità in ciò che si misura.

ESEMPIO

La misura di due centimetri di un segmento è -senza incertezze- il doppio di uno con lunghezza pari ad un centimetro;

se la misura aumenta del 50% anche il segmento si allunga di una estensione pari alla sua metà.

E' ammessa ogni trasformazione del tipo:

$$\text{Se } X_j \neq 0 \quad \frac{X_i}{X_j} = c \left[\frac{f(X_i)}{f(X_j)} \right]; \quad \text{per } f(X_j) \neq 0$$

Variabili discrete

Derivano da un processo di conteggio o di numerazione:

riviste per numero di abbonati
partiti politici per numero di iscritti
catene commerciali per numero di punti vendita affiliati,

Le modalità sono presentate usualmente, ma non sempre, in ordine crescente:

$$X_1 < X_2 < \dots < X_k$$

Il simbolo "<" ha qui il significato aritmetico di "minore".

La differenza tra due modalità ha significato costante, ma nulla si può dire sul rapporto tra di esse.

In modo alternativo si può dire che le modalità della variabile discreta possono essere contate ovvero poste in corrispondenza biunivoca con l'insieme dei numeri naturali.

Discrete frazionarie e dense

Una variabile può essere discreta, ma espressa con dei numeri decimali

ESEMPIO: lancio di due dadi. Modalità= semisomma dei punti sulle facce superiori.
L'uscita di un "6" e di un "5" o di un doppio "6" danno luogo alle modalità

$$\frac{6+5}{2} = 5.5; \quad \frac{6+6}{2} = 6$$

La variabile è discreta perchè tra "5.5" e "6" la variabile non può assumere alcun valore. Le sue modalità sono tutte ISOLATE: E' sempre possibile trovare un intervallo, per quanto piccolo, che contiene una sola modalità.

La variabile DENSE è discreta per natura, ma ha una unità di misura è molto piccola rispetto all'ordine di grandezza con cui si manifesta

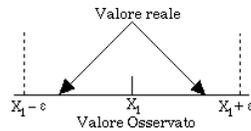
reddito in lire;
circolazione di vetture per numero di auto;
nazioni per numero di abitanti;

La trattazione dei caratteri densi è simile a quella dei caratteri continui

Variabili continue

Non possono essere rilevate puntualmente; il valore assunto è il centro dell'intervallo

$$[X_1 - \varepsilon, X_1 + \varepsilon]$$



Dire che $X = X_1$ significa dire che $|X - X_1| < \varepsilon$ cioè che si è osservato uno qualsiasi degli infiniti valori compresi in

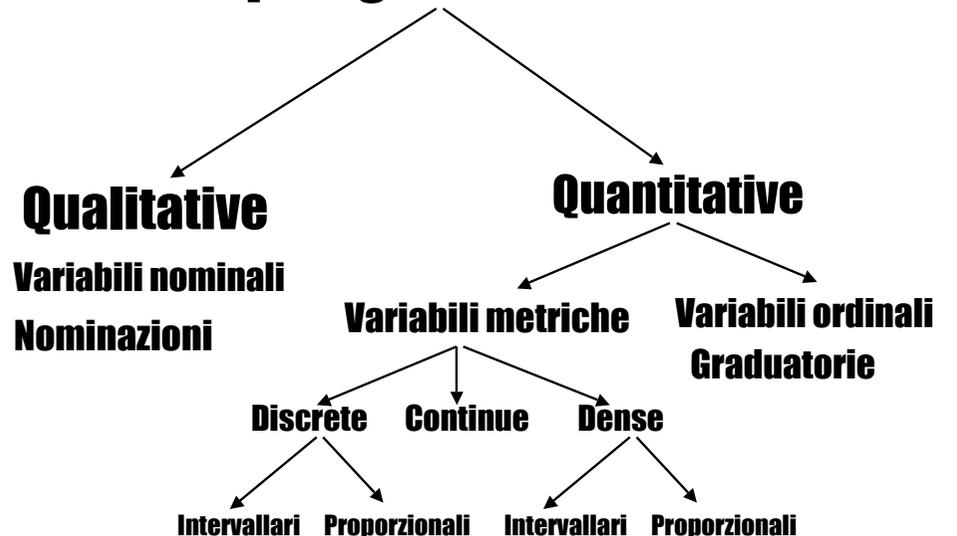
$$[X_1 - \varepsilon, X_1 + \varepsilon]$$

L'ampiezza del sottointervallo dipende dalla precisione degli strumenti di rilevazione. (Questo è un limite degli strumenti di misurazione non della variabile misurata)

Talvolta le modalità sono presentate come interi. Per distinguerle da quelle di una variabile discreta basta ricordare che:

Tra due modalità discrete non ve ne può essere un'altra, tra due modalità continue se ne trovano infinite

Tipologia delle variabili



Criterio organizzativo

Ogni unità si inserisce in un contesto in cui si distingue e che consente di attribuirle la modalità corretta.

Le tecniche statistiche sono anonime e trascurano la localizzazione dell'unità rispetto alle altre.

In alcune analisi è necessario che l'unità sia ben collocata -nel tempo o nello spazio- ed il suo esame prima o dopo di un'altra è rilevante.

Gli ordinamenti possibili sono diversi, ma noi consideriamo solo

SERIE SPAZIALI (ordinamento geografico)

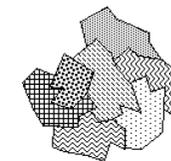
SERIE STORICHE (ordinamento temporale)

Le serie territoriali

Si ritiene che la modalità o intensità raggiunta dipenda dalla sua posizione topografica. Qui conta il tipo di unità considerata



UNITA' AREALI: rappresentata da una poligolone chiusa
 entità fisiche: isola, lago, continente, etc.
 entità amministrative: comuni, regioni, nazioni
 entità funzionali: distretti sanitari, telefonici, scolastici



Le unità si considerano omogenee al loro interno anche se la rilevazione del carattere si effettua in più punti

Paese	%
Belgio	20.2
Danimarca	29.8
Germania	15.4
Grecia	10.4
Francia	22.8
Spagna	14.3
Regno Unito	19.5
Irlanda	17.9
Italia	17.4
Lussemburgo	11.6
Paesi bassi	15.1
Portogallo	14.1

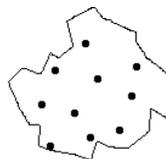
Esempio:
 Percentuale di dipendenti pubblici sul totale occupati

Altri tipi di unità (territoriali)



UNITA' PUNTUALI: costituiscono i nodi di una maglia più o meno fitta di punti che coprono un dato territorio

misurazioni atmosferiche e idrogeologiche,
 censimenti della popolazione
 rilevazione della forza lavoro



Le unità puntuali hanno il grande pregio di visualizzare l'ubicazione delle modalità o intensità rivelandone la disseminazione o la concentrazione nel territorio

Esempio: Consumi di acqua per uso domestico

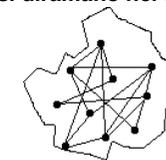
Città	Consumo
Bruxelles	108
Amburgo	146
Copenaghen	194
Londra	132
Parigi	147
Roma	220
Lussemburgo	171
Amsterdam	159
Madrid	158

Altri tipi di unità/2



UNITA' RETICOLARI (NETWORK): sono unità che si diramano nel territorio

fiumi, strade, gallerie
 direttrici di sviluppo
 rotte di navigazione
 reti di distribuzione



La rilevazione dei caratteri sui network avviene per punti, in analogia alla osservazione di un flusso che percorre il reticolo (il flusso è spesso la variabile che si intende rilevare)

Esempio:
 itinerari turistici calabresi per numero di pro-loco coinvolte

Itinerari	Pro-loco
Catanzaro-S.S. Bruno	5
Catanzaro-Capo Vaticano	7
Catanzaro-Laghi	3
Costa tirrenica	14
Sila e laghi	5
Costa Jonica	14
Aspromonte	5
Costa Viola	13
Magna Grecia	17

Le serie storiche

Spesso si studiano variabili nel tempo e la progressione cronologica dei valori è essenziale per comprendere il comportamento della variabile



VARIABILI DI FLUSSO: procedono in modo continuo. Se il flusso è regolare non importa l'unità di tempo. Se il flusso è erratico mette conto sapere che si rileva per settimane, mesi, trimestri, anni, etc.

Esempio:

Spesa dell'amministrazione statale per la cultura
te utto conto che ogni anno non arretra rispetto al
periodo precedente

Anno	Spesa
1984	1 055
1985	1 206
1986	1 322
1987	1 954
1988	1 393
1989	1 064
1990	988
1991	947



VARIABILI DI STOCK: si manifestano in un dato istante per poi ripetersi più o meno regolarmente. Ricominciano da zero

Esempio:

Voti validi nelle consultazioni politiche

Anni	Totale
1948	26268912
1953	27092743
1958	29563633
1963	30758031
1968	31803253
1972	33414779
1976	36727273
1979	36671308
1983	36906005
1987	38592383