

Gestione informatica dei dati

La costruzione del dato

Prof. Roberto Foderà

Cattedra di Gestione informatica dei dati
Dipartimento di Giurisprudenza - Palermo

A.A. 2021-2022



LUMSA
UNIVERSITÀ

La costruzione del dato

Indice

Le fasi dell'indagine scientifica (statistica)

Punto

Punto

Punto

La costruzione del dato

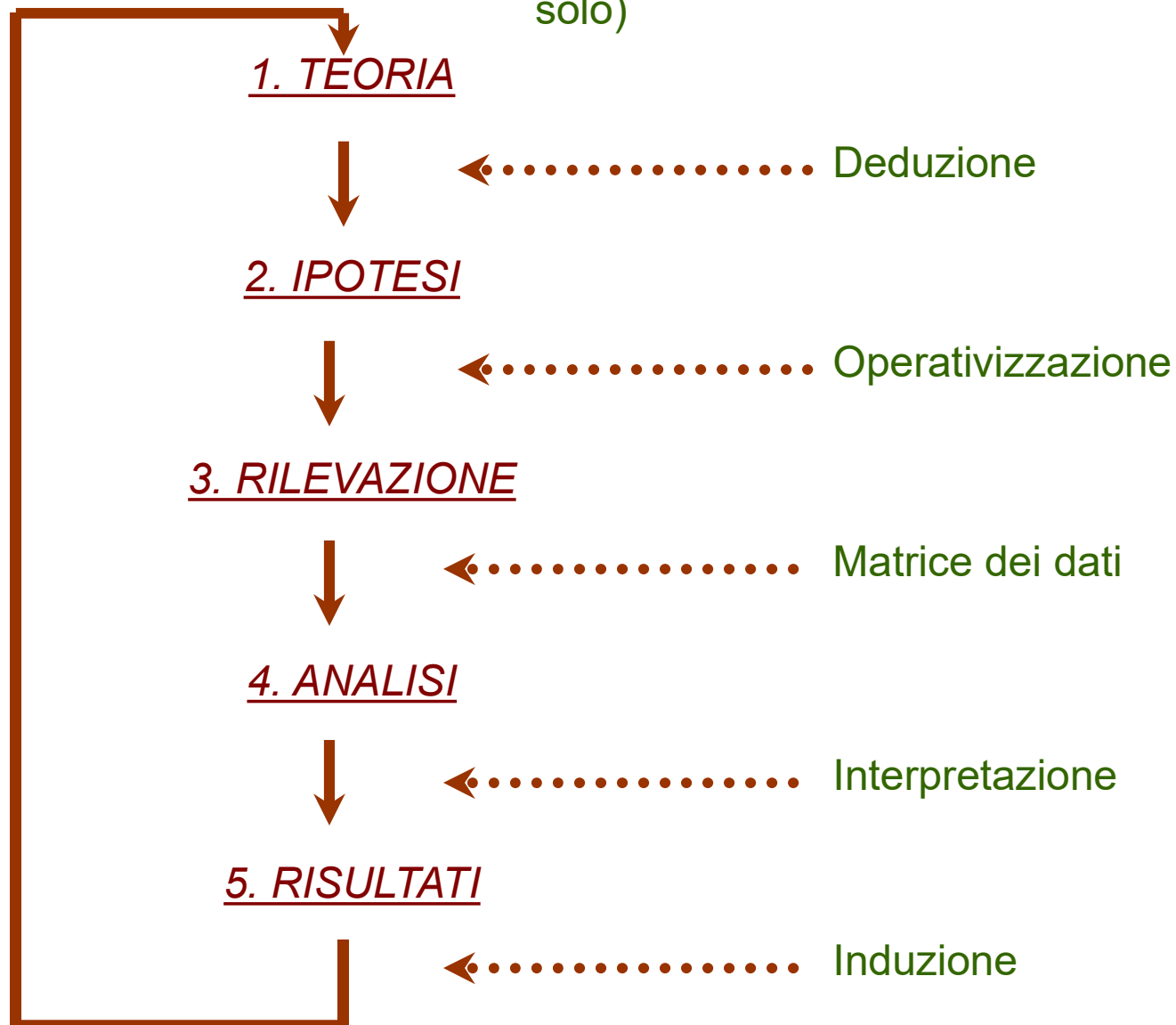
Premessa

Nella maggior parte delle ricerche sociologiche abbiamo bisogno di una adeguata miscela di parole e di numeri, di penetrazione e di tabulazione, di rappresentazione e di interpretazione, di situazioni capite e di correlazioni indiscutibili: cioè il materiale grezzo della vita e la statistica autentica della validità

(Harrison T.)

La costruzione del dato

Due modi di definire il processo di ricerca: A - Le fasi dell'indagine statistica (e non solo)



La costruzione del dato

Le fasi dell'indagine statistica (e non solo)

PRIMA FASE

Richiesta di
informazioni



Obiettivi e
finalità

SECONDA FASE

Individuazione
e delle fonti

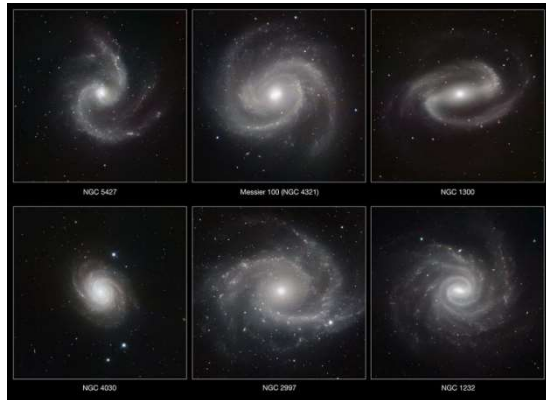


Indagine di
sfondo

La costruzione del dato

Le fasi dell'indagine statistica (e non solo)

TERZA FASE

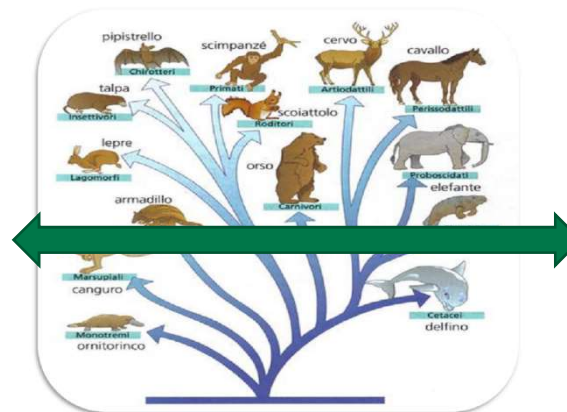


Collettivo
statistico



QUARTA FASE

Scelta delle
variabili



Classificazione

La costruzione del dato

Le fasi dell'indagine statistica (e non solo)

QUINTA FASE

Piano di
rilevazione

Piano di
elaborazione



SESTA FASE

Rilevazione

Diffusione



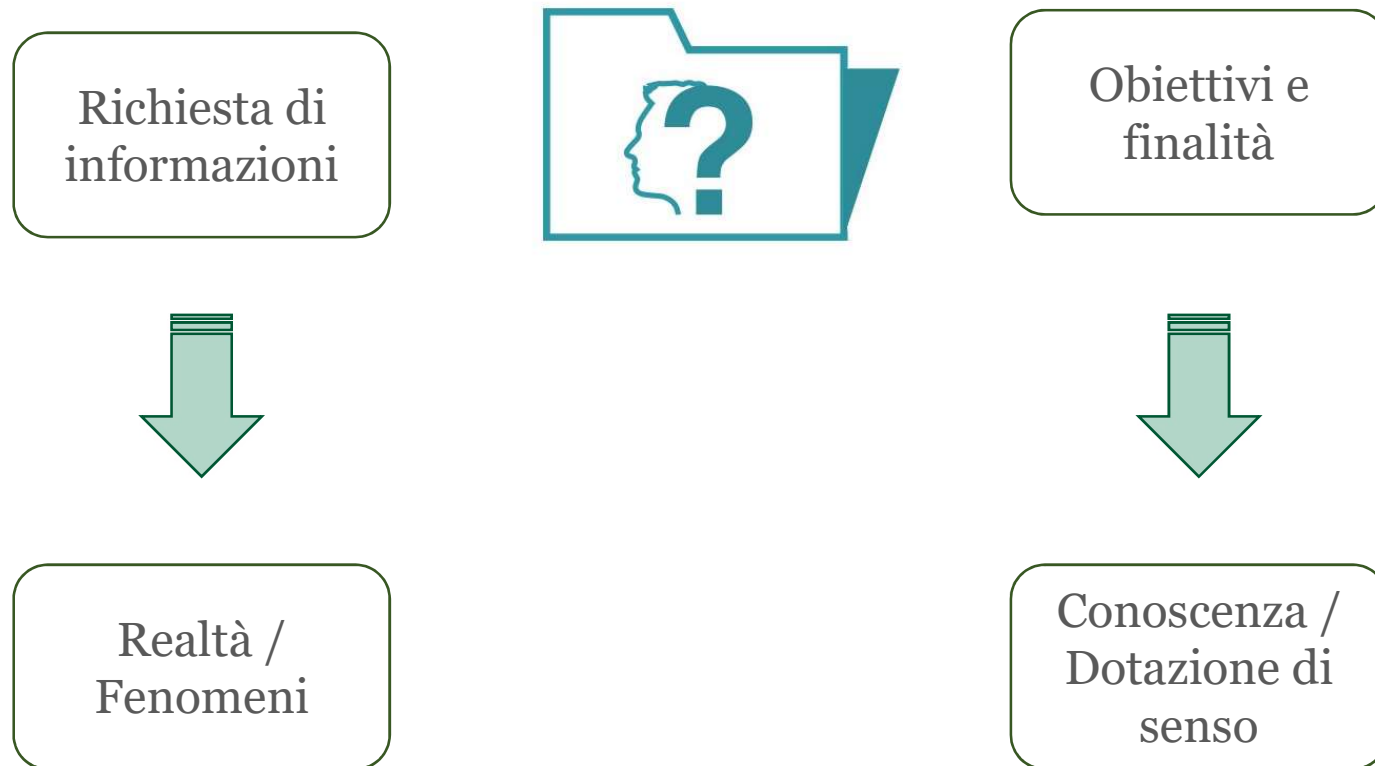
Elaborazione

I.Stat

La costruzione del dato

Le fasi dell'indagine statistica (e non solo)

PRIMA FASE



Alcuni elementi da considerare nelle analisi sociali

- **VALORI**
- **REATTIVITA'**
- **UNITA' DI ANALISI**
- **CLASSIFICAZIONI**

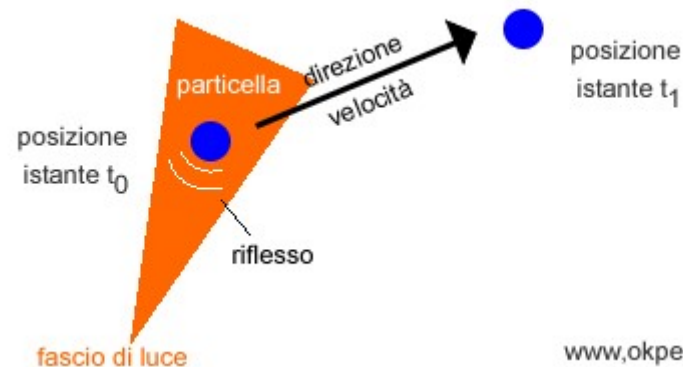
La costruzione del dato

Alcuni elementi da considerare nelle analisi sociali

• REATTIVITA'

Per reattività si intende la reazione che coloro che sono oggetto della rilevazione hanno, reazione che potrebbe influenzare e distorcere la rilevazione stessa.

Anche la condizione dello scienziato sociali dell'essere parte della società che sta studiando (condividendo i valori e la cultura implicita di quella) può essere elemento di orientamento (distorsione) della ricerca.



La costruzione del dato

Alcuni elementi da considerare nelle analisi sociali

- **VALORI**

C'è una grande differenza tra affermare che si vede un fenomeno e poi lo si interpreta e affermare che si vedono fenomeni interpretati. Quest'ultima opinione sembra quella corretta.

(McCain G., Segal E.M)



La costruzione del dato

I presupposti logici dell'indagine statistica per le scienze sociali



La costruzione del dato

I presupposti logici dell'indagine statistica per le scienze sociali

Concetto A

- ☑ *parola breve*
- ☑ *utilità comunicativa*
- ☑ *parsimonia*
- ☑ *non verificabilità*

La costruzione del dato

I presupposti logici dell'indagine statistica per le scienze sociali

NEVE

*** *Precipitazione caratteristica di
climi freddi*

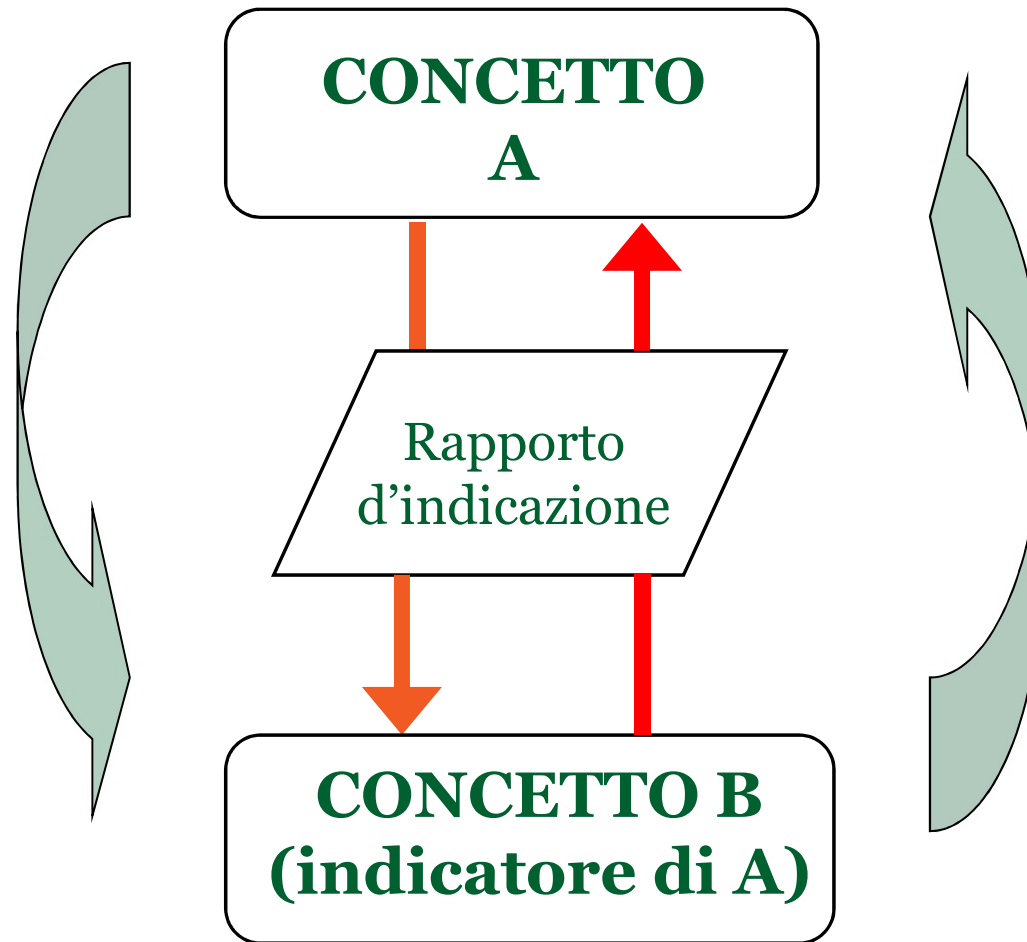
*** *neve*

*** *neve gelata*

*** *quel tipo di neve gelata che va bene
per fabbricare le pareti di un igloo*

La costruzione del dato

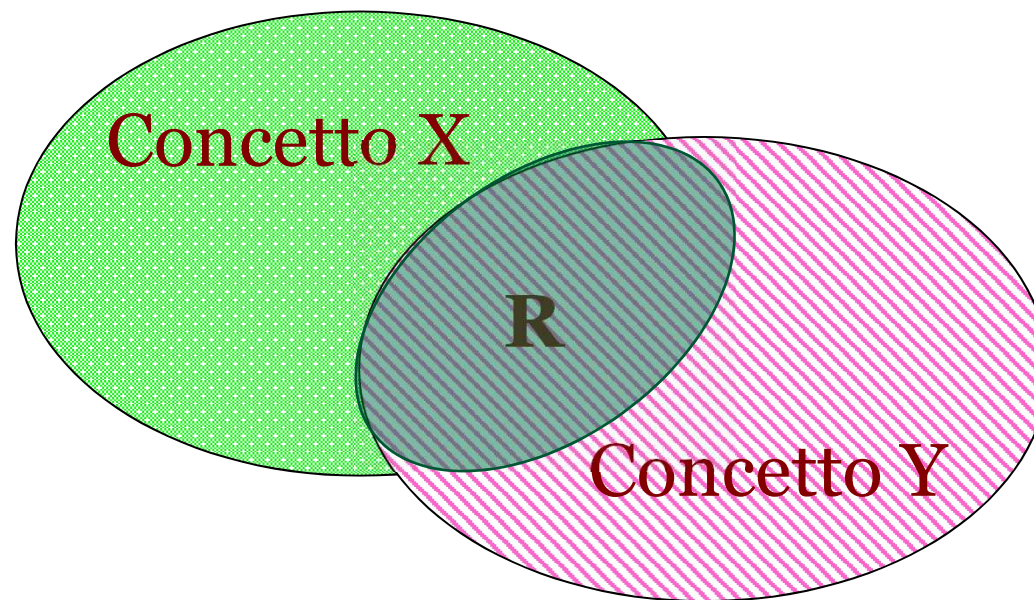
I presupposti logici dell'indagine statistica per le scienze sociali



Scala di generalità

La costruzione del dato

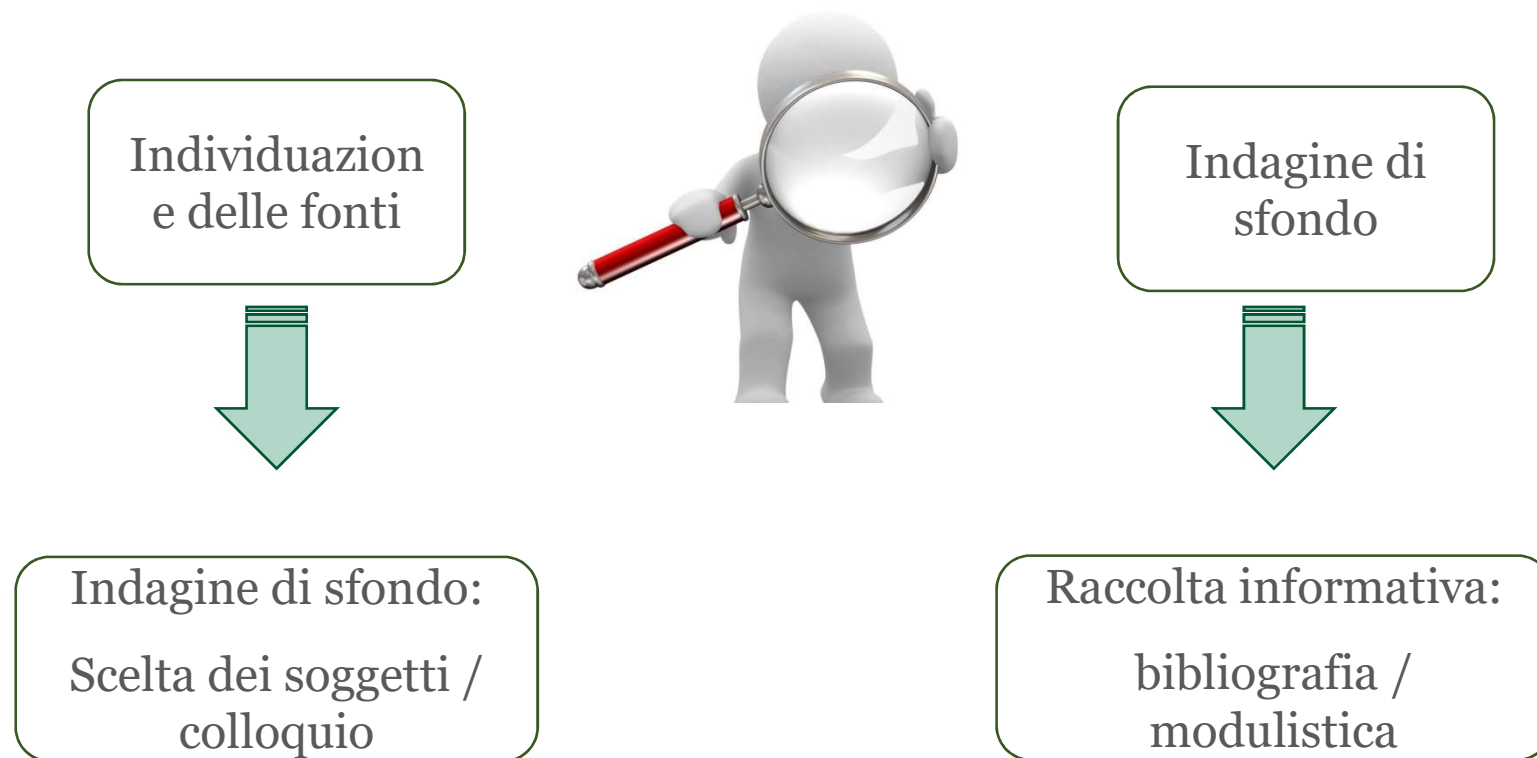
I presupposti logici dell'indagine statistica per le scienze sociali



La costruzione del dato

Le fasi dell'indagine statistica – seconda fase

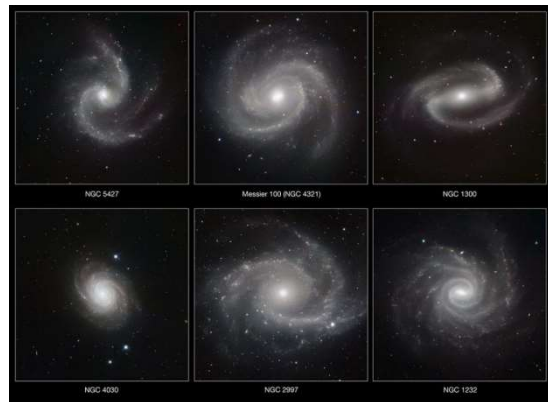
SECONDA FASE



La costruzione del dato

Le fasi dell'indagine statistica – terza fase

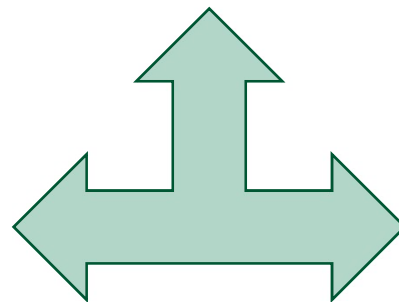
TERZA FASE



Collettivo
statistico



Unità di rilevazione



Unità di analisi

semplice

composta

La costruzione del dato

Le fasi dell'indagine statistica – terza fase

• UNITA' DI RILEVAZIONE

E' l'elemento strumentale presso il quale vengono assunte le informazioni elementari relative alle unità elementari. Non sempre corrisponde all'unità di analisi e potrebbero essere più di una in una stessa rilevazione. La sua funzione si esaurisce una volta completata l'attività di rilevazione. Ha però forti impatti sull'organizzazione dell'indagine e sulle possibilità di ottenere stime accurate.

Unità di rilevazione possono essere la famiglia, l'impresa, l'azienda agricola, la classe, ma anche l'edificio.

La costruzione del dato

Le fasi dell'indagine statistica – terza fase

• UNITA' DI ANALISI

Sono i singoli elementi base sui quali vengono rilevate le variabili e operata la misurazione. Su di esse si fonda tutta l'attività successiva alla fase di rilevazione.

Fanno riferimento normalmente ad un'unica unità di rilevazione mentre può avvenire che ad una unità di rilevazione corrispondano più unità di analisi.

Unità di analisi di tipo semplice (o elementare) rappresentano unità cui la misura ha riscontro diretto sul questionario. Si parla di microdati.

Unità di analisi composte (o aggregate) pur essendo a tutti gli effetti delle vere e proprie unità di analisi devono essere costruite a partire dai microdati.

La costruzione del dato

Le fasi dell'indagine statistica – terza fase

Questionario indagine Multiscopo sulle famiglie
Condizioni di salute e ricorso ai servizi sanitari

Domande poste alla persona di
riferimento dell'**unità di rilevazione**

QUESTIONARIO FAMILIARE - PARTE SECONDA	
1. L'ABITAZIONE IN CUI VIVE LA FAMIGLIA	
<i>(Con riferimento all'abitazione principale specificare)</i>	
1.1 A che titolo la famiglia occupa l'abitazione?	1.10 Si verificano uno o più dei seguenti problemi? <i>(una risposta per ogni riga)</i>
Affitto e subaffitto1 <input type="checkbox"/>	Le spese per l'abitazione sono troppo alte.....1 <input type="checkbox"/> NO SÌ
Proprietà2 <input type="checkbox"/>	L'abitazione è troppo piccola.....3 <input type="checkbox"/> 2 <input type="checkbox"/>
Usufrutto3 <input type="checkbox"/>	L'abitazione è troppo distante da altri familiari.....5 <input type="checkbox"/> 4 <input type="checkbox"/>
Titolo gratuito4 <input type="checkbox"/>	Irregolarità nell'erogazione dell'acqua.....7 <input type="checkbox"/> 6 <input type="checkbox"/>
Altro5 <input type="checkbox"/>	Nelle stanze da letto dell'abitazione sono presenti macchie di umidità oppure muffe o funghi.....1 <input type="checkbox"/> 8 <input type="checkbox"/>
1.2 Di quante stanze si compone l'abitazione? <i>(comprendere anche la cucina se ha caratteristiche di stanza)</i>	L'abitazione è in cattive condizioni.....3 <input type="checkbox"/> 4 <input type="checkbox"/>
N. di stanze..... <input type="text"/>	
1.3 L'abitazione dispone di bagno con vasca e/o doccia?	
NO.....1 <input type="checkbox"/>	
Sì.....2 <input type="checkbox"/> → N. di bagni..... <input type="text"/>	
1.4 L'abitazione dispone di telefono fisso?	
NO.....1 <input type="checkbox"/>	
Sì.....2 <input type="checkbox"/> → N. di linee telefoniche..... <input type="text"/>	

INDAGINE STATISTICA MULTISCOPO SULLE FAMIGLIE	
MOD. ISTAT/IMF-8/A.12-13	
Istat SISTEMA STATISTICO NAZIONALE ISTITUTO NAZIONALE DI STATISTICA	
A MESE <input type="text"/>	
1 Provincia <input type="text"/>	
Comune <input type="text"/>	
Sezione di Censimento <input type="text"/>	
2 <i>(a cura del Comune)</i> Numero progressivo della famiglia intervistata Da 0001 al totale delle famiglie intervistate nel trimestre <input type="text"/>	
3 Numero d'ordine della famiglia nell'elenco Mod. ISTAT/IMF/2 Elenco Da 00001 al totale delle famiglie nell'elenco nel trimestre <input type="text"/>	
6 CODICE INTERVISTATORE <input type="text"/>	
7 Data di consegna all'ufficio del Comune <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Giorno Mese Anno L'INTERVISTATORE (Cognome e nome leggibili) Visto per la revisione IL CAPO DELL'UFFICIO ADDETTO ALLA RILEVAZIONE	
2. AIUTI E SITUAZIONE ECONOMICA DELLA FAMIGLIA	
2.1 La famiglia si avvale del servizio a pagamento di: <i>(una risposta per ogni riga)</i>	
5 Numero dei componenti la famiglia attuale <input type="text"/>	

La costruzione del dato

Le fasi dell'indagine statistica – terza fase

Questionario indagine Multiscopo sulle famiglie
Condizioni di salute e ricorso ai servizi sanitari

Domande poste alle singole persone
componenti la famiglia: **unità di analisi**

7. ACCERTAMENTI DIAGNOSTICI (PER TUTTI)

7.1 Nelle ultime 4 settimane si è sottoposto ad accertamenti diagnostici (escludendo quelli fatti durante un eventuale ricovero ospedaliero o in day hospital)?

NO.....1 ☐ → andare a dom. 7.6
SI.....2 ☐ → N. di volte

(Se si è sottoposto ad accertamenti diagnostici)

7.2 Nelle ultime 4 settimane ha fatto esami del sangue?

NO.....1 ☐ → andare a dom. 7.4
SI.....2 ☐ → N. di volte

17. PREVENZIONE

(PER LE PERSONE DI 15 ANNI E PIÙ)

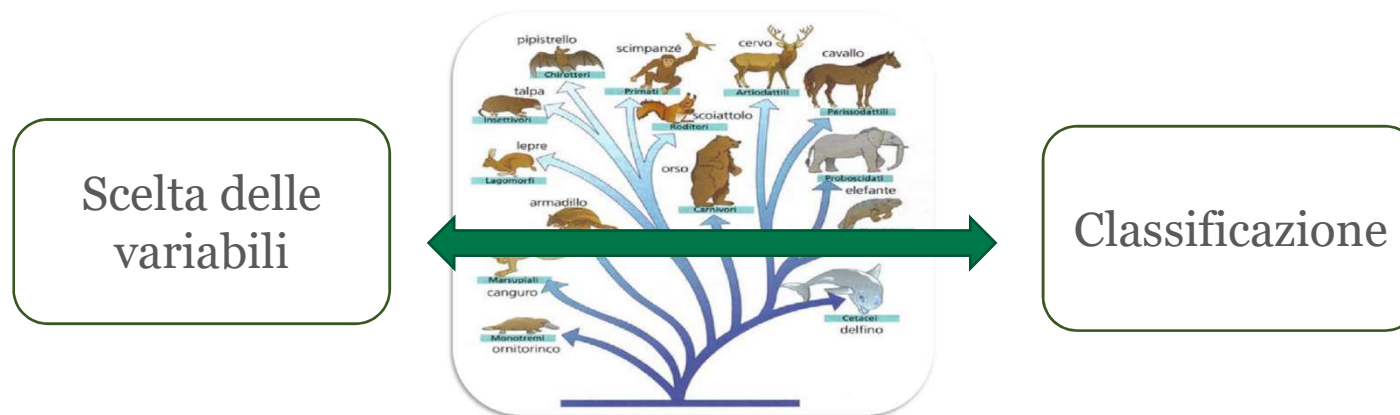
17.1 Con quale frequenza controlla il colesterolo, la glicemia e la pressione arteriosa?
(una risposta per ogni colonna)

	Cole-sterolo	Glicemia (zucchero nel sangue)	Pressione arteriosa
Almeno 1 volta al mese	1 <input type="checkbox"/>	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Ogni 2-6 mesi	2 <input type="checkbox"/>	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Una volta l'anno	3 <input type="checkbox"/>	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Ogni 2-4 anni	4 <input type="checkbox"/>	4 <input type="checkbox"/>	4 <input type="checkbox"/>
Ogni 5 anni	5 <input type="checkbox"/>	5 <input type="checkbox"/>	5 <input type="checkbox"/>
Oltre i 5 anni	6 <input type="checkbox"/>	6 <input type="checkbox"/>	6 <input type="checkbox"/>
Non ha mai fatto controlli	7 <input type="checkbox"/>	7 <input type="checkbox"/>	7 <input type="checkbox"/>

La costruzione del dato

Le fasi dell'indagine statistica – quarta fase

QUARTA FASE



Scegliamo d'ignorare molte delle differenze percettive che rendono unico ciascun oggetto, e in larga misura lo facciamo al momento di dargli un nome. (Tyler)

• CLASSIFICAZIONI

Scegliamo d'ignorare molte delle differenze percettive che rendono unico ciascun oggetto, e in larga misura lo facciamo al momento di dargli un nome. Dando un nome classifichiamo gli oggetti che per noi sono simili nella stessa categoria, anche se siamo in grado di percepire differenze fra loro ... [in tal modo] l'infinita variabilità del mondo si riduce a dimensioni tollerabili e manipolabili.

I nostri criteri di classificazione sono interamente arbitrari e soggettivi. Non c'è alcunché nel mondo esterno che domandi che certe cose siano riunite sotto la stessa etichetta ed altre no.

Tyler

• CLASSIFICAZIONI

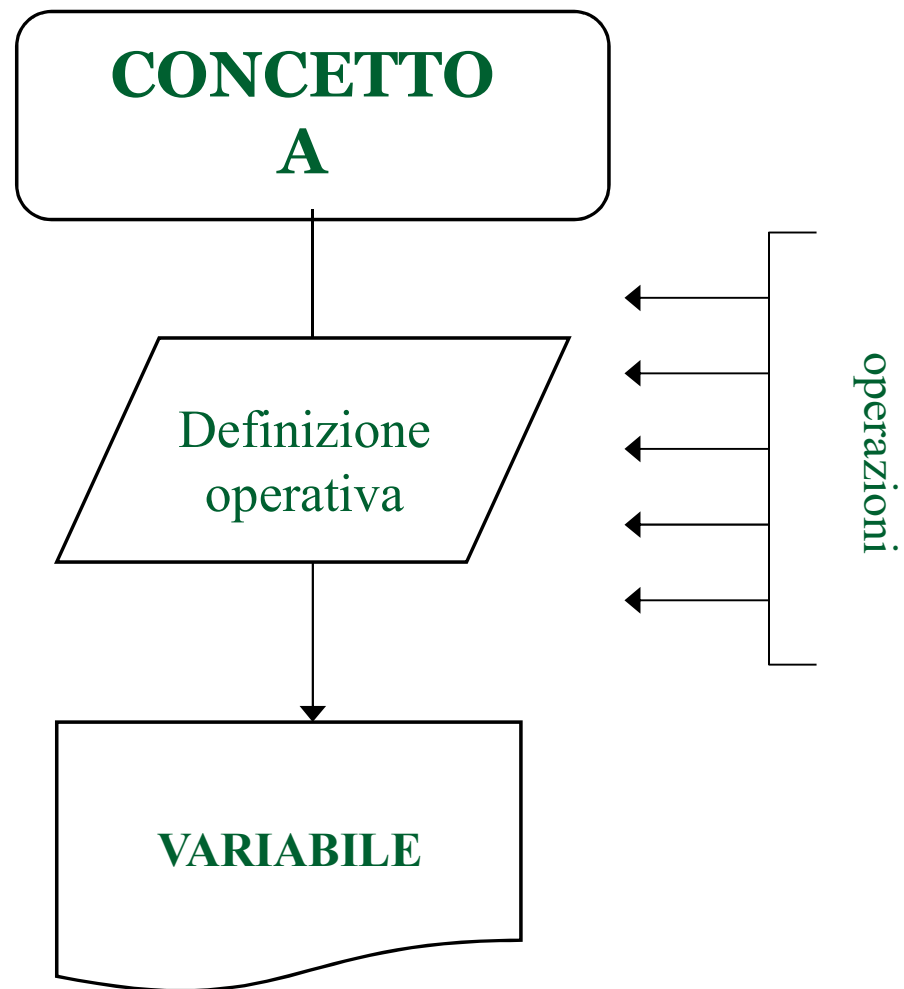
” ... d'altronde questi concetti e principi sono creazioni libere dello spirito umano, che non si possono giustificare a priori né con la natura dello spirito umano né in altro modo qualsiasi”

Einstein



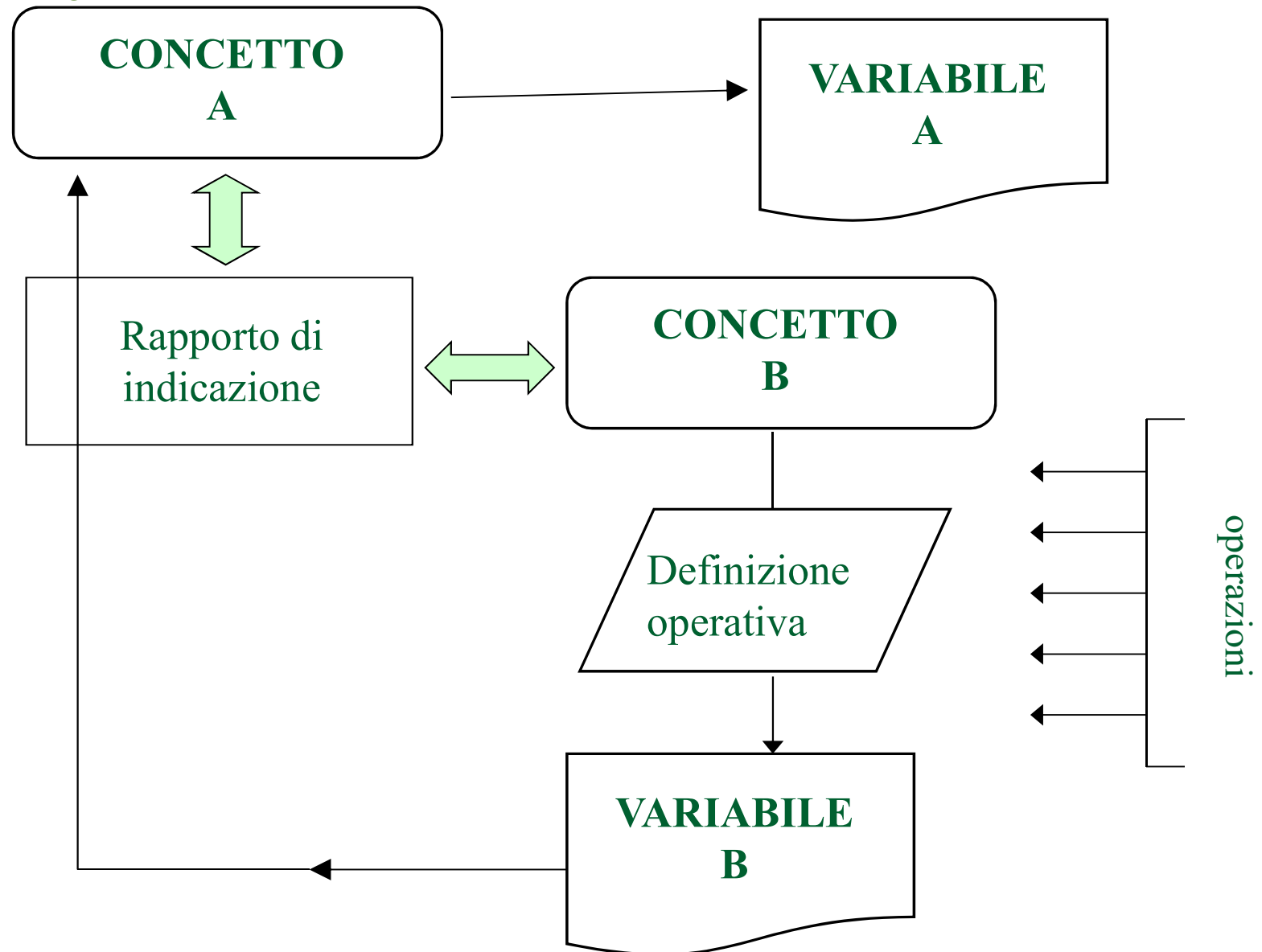
La costruzione del dato

Le fasi dell'indagine statistica – quarta fase

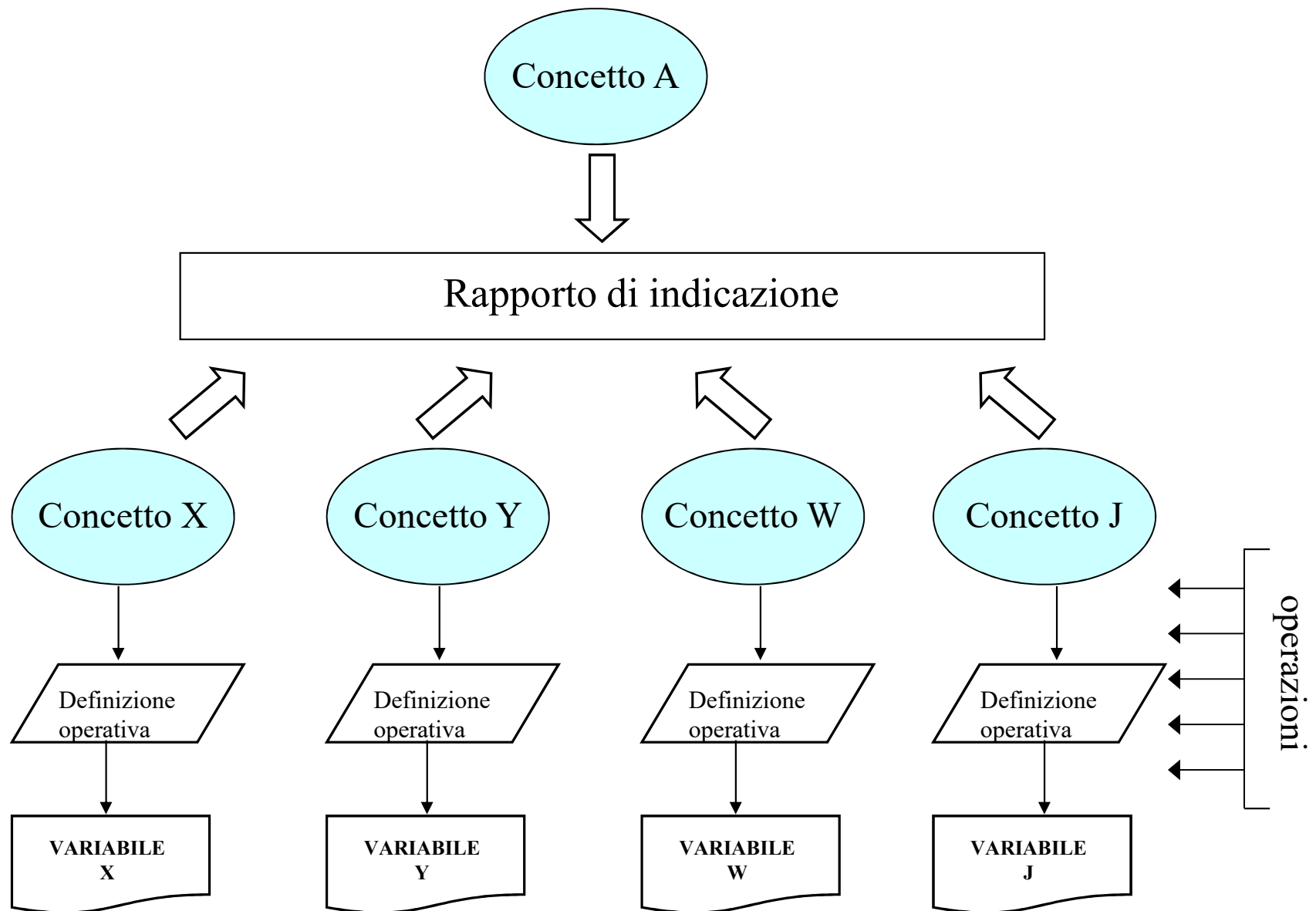


La costruzione del dato

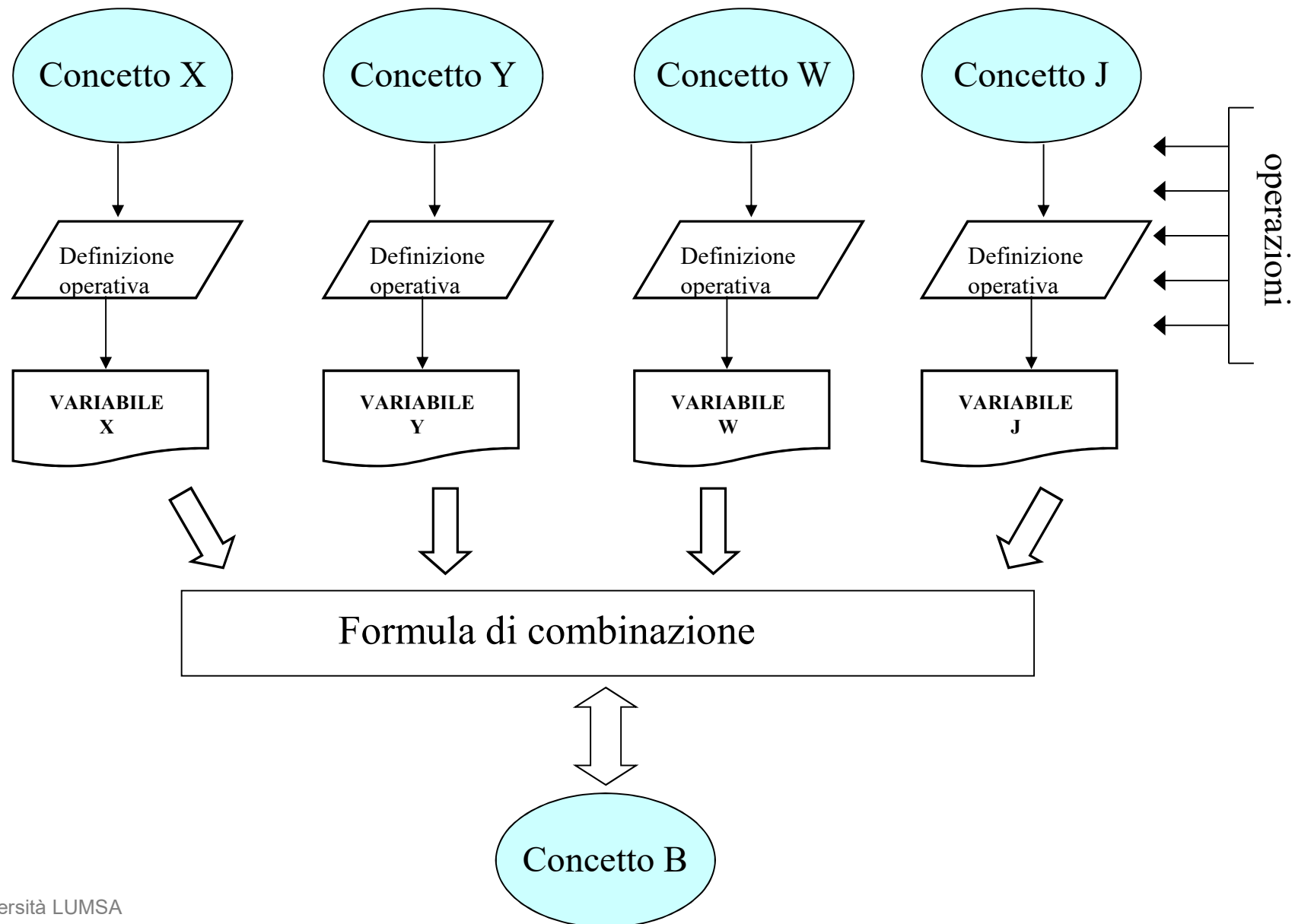
Le fasi dell'indagine statistica – quarta fase



La costruzione del dato



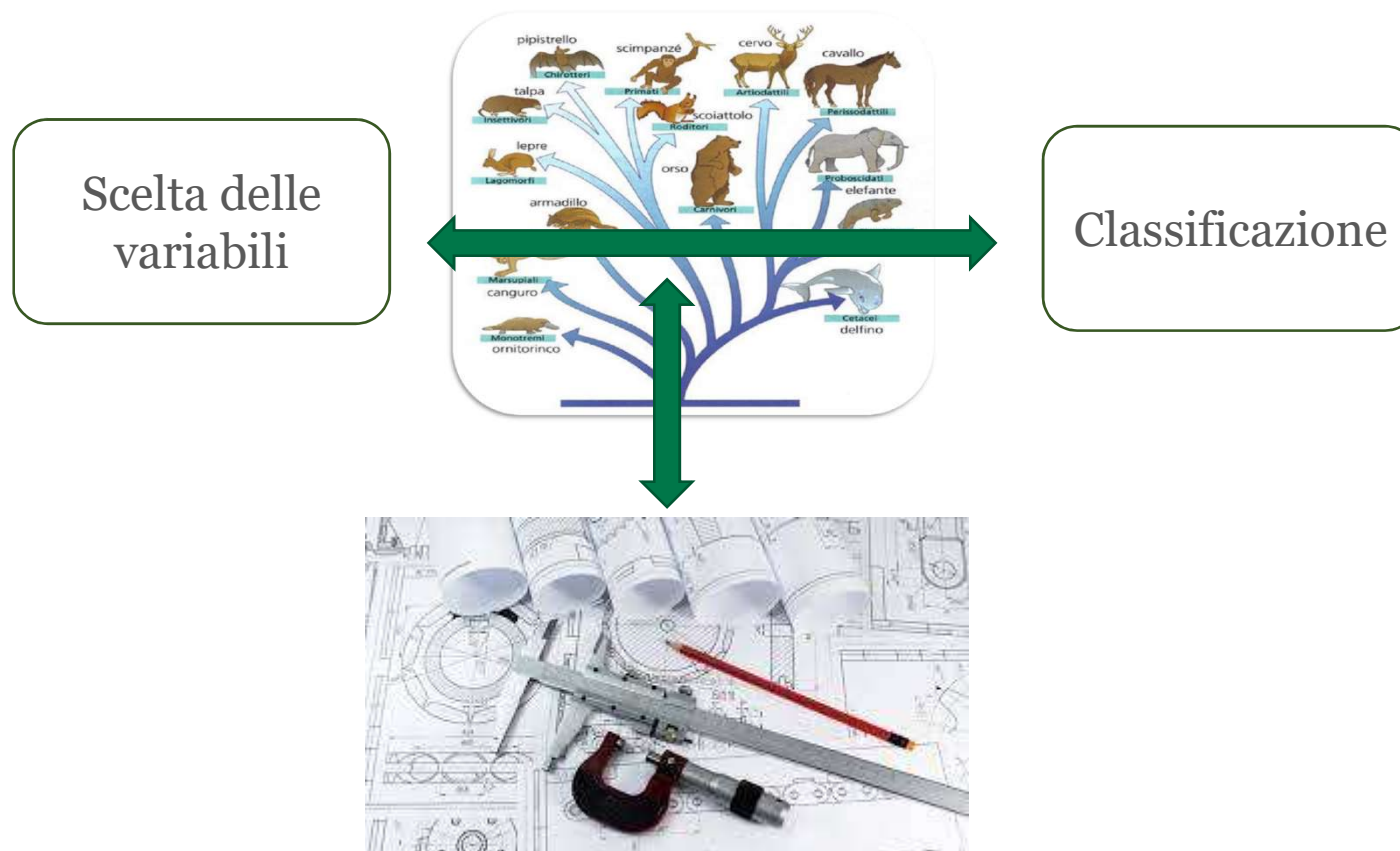
La costruzione del dato



La costruzione del dato

Le fasi dell'indagine statistica – quarta fase

QUARTA FASE



La misurazione

La costruzione del dato

Le fasi dell'indagine statistica – quarta fase

CLASSIFICAZIONE

Per la classificazione bisogna considerare che:

1. Ogni caso rilevato deve essere attribuito ad una classe
2. Nessun caso deve poter essere attribuito a più di una classe
3. Il criterio deve essere univoco (*fundamentum divisionis*)

La costruzione del dato

Le fasi dell'indagine statistica – quarta fase

“Ai membri della British Association ci permettiamo di suggerire, in conclusione, che la definizione meno restrittiva e più utile di misurazione è, come ebbe a dire uno dei suoi membri, "l'assegnazione di numeri alle cose in modo tale da rappresentare i fatti e le relazioni tra essi". Il problema su cosa sia e cosa non sia misurazione si riduce allora alla semplice domanda: quali sono le regole, se ci sono, con le quali i numeri vengono assegnati?”

Stevens S. S.



La costruzione del dato


Le fasi dell'indagine statistica – quarta fase




Le variabili possono essere distinte in:

1. Quantitative / qualitative
2. Discrete / continue

La costruzione del dato

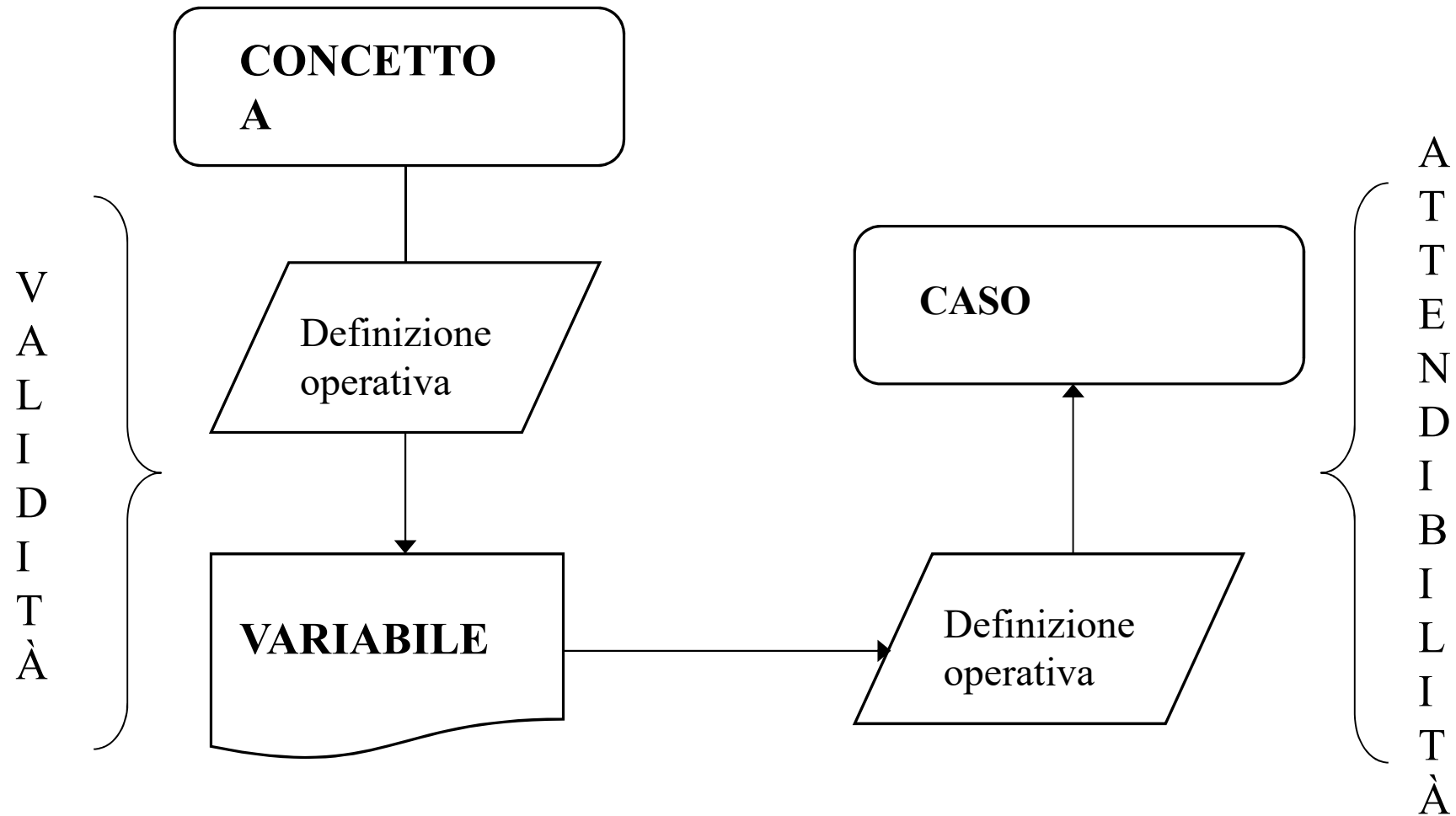
Le fasi dell'indagine statistica – quarta fase

 Livelli di misura	Proprietà	
	NOMINALE	Classificazioni
	ORDINALE	Ordinamento gerarchico
	A INTERVALLI	Misure aritmetiche
DI RAPPORTI	Misure proporzionali alle distanze	



La costruzione del dato

Le fasi dell'indagine statistica – quarta fase



La costruzione del dato

Le fasi dell'indagine statistica – quinta fase

QUINTA FASE

Piano di
rilevazione

Piano di
elaborazione



Redazione del
questionario

La costruzione del dato

Le fasi dell'indagine statistica – quinta fase

SCHEDA GENERALE

N. d'ordine dei componenti	Sesso	Età (in anni compiuti)	Relazione di parentela o di convivenza con la persona di riferimento (PR) del questionario		Posizione con riferimento all'indirizzo della famiglia anagrafica	Anno di nascita	6 ANNI E PIÙ		15 ANNI E PIÙ		Solo per le persone coniugate		
							Stato civile	Titolo di studio più alto conseguito	Condizione unica o prevalente	Fonte principale di reddito	Anno del matrimonio attuale	Stato civile prima del matrimonio attuale	
1	2	3	4		5	6	7	8	9	10	11		12
0 1			PR	0 1									

La costruzione del dato

Le fasi dell'indagine statistica – quinta fase

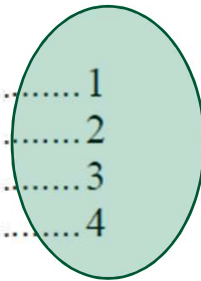
Le domande di un questionario possono avere diverse modalità di codifica delle variabili.

Per tutti

2.3 - Le capita di pensare di avere un carico eccessivo di lavoro domestico?

Leggere le risposte

Sì, spesso.....	1
Sì, di tanto in tanto	2
Sì, ma raramente	3
No, mai.....	4



La costruzione del dato

Le fasi dell'indagine statistica – quinta fase

Possono essere previste una sola o più risposte

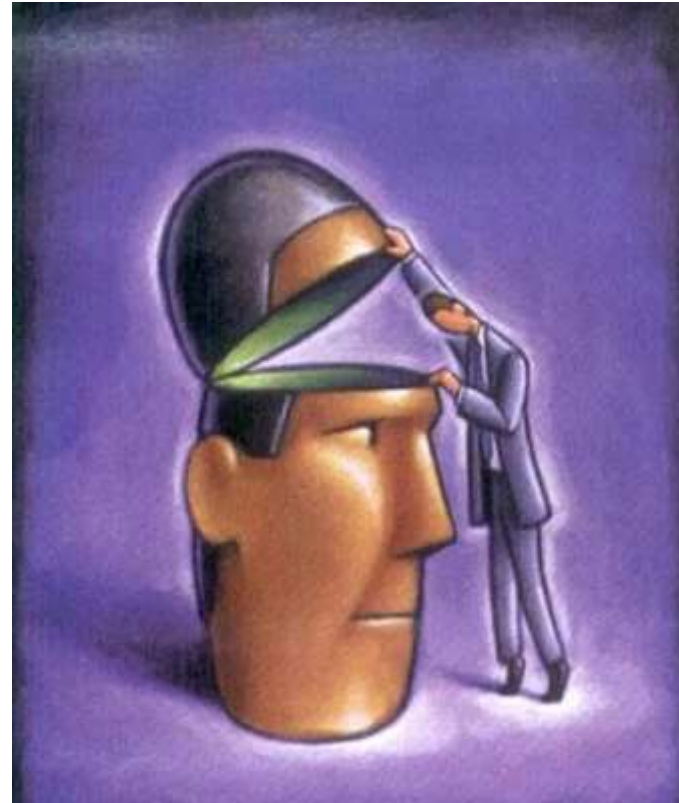
X_B8. Per quale motivo non ha sostenuto spese?

- Non dispone del servizio 01__
- Pagata da altri (parenti non conviventi, amici, datore di lavoro, enti di assistenza, ecc.) 02__
- Compresa nell'affitto 03__
- Non scorporabile da altre utenze 04__
- Nuova utenza, non sono ancora arrivate bollette 05__
- A credito con il fornitore 06__
- Mancato recapito della bolletta negli ultimi 12 mesi 07__
- Non utilizza il servizio 08__
- *Morosità* 09__
- **Altro (*Specificare*)** 10__
- Non sa/Non ricorda 88__

La costruzione del dato

I nostri più importanti risultati scientifici scaturiscono dalle riduzioni più perspicaci ed eleganti delle complessità superficiali della Natura, riduzioni che ne rivelano la semplicità di fondo; d'altra parte, spesso i nostri errori peggiori derivano dall'eccessiva semplificazione di aspetti della realtà che in seguito si dimostrano di gran lunga più complessi di quanto avessimo compreso.

Barrow J.



La costruzione del dato

Le fasi dell'indagine statistica – sesta fase

SESTA FASE

Rilevazione



Raccolta delle misure
statistiche
(totali o campionarie)



Elaborazione



Matrice dati
Indicatori
Modelli statistici

Diffusione

I.Stat



Archivi di dati
Basi di dati
Datawarehouse
Open data