

Università degli Studi di Napoli «Federico II»  
Facoltà di Architettura

**UPTA**

Corso di laurea in Urbanistica e Scienze della Pianificazione Territoriale e Ambientale  
AA 2017-2018

**Corso integrato di Matematica e statistica**

Docente: Livia D'Apuzzo.

**STATISTICA**

**Terminologia, classificazione caratteri, tabelle di distribuzione unitaria**



**STATISTICA: Scienza che indaga fenomeni collettivi, sia sociali sia naturali, che presentano variabilità** (possono cioè presentarsi con diverse modalità).

**Il fine della statistica è di fornire, attraverso l'elaborazione e l'interpretazione dei dati, i modelli matematici atti a descrivere il fenomeno e a formulare previsioni.**

### 1. CENNI STORICI

Il termine “statistica” deriva da “statista” e quindi da “Stato”. E, infatti, la statistica nasce dall'esigenza proprio della statista di conoscere e quantificare attraverso numeri fenomeni che interessano la popolazione di uno stato o una sua parte: distribuzioni di beni, incidenza di determinate malattie, mortalità, etc. . . .

Già nel 3800 A.C. i Sumeri compivano indagini sulla popolazione al fine di contare uomini e i beni di cui una comunità poteva disporre. Indagini per rilevare l'ammontare della popolazione si hanno in Egitto già ai tempi della prima dinastia (intorno al 3000 A.C.) e durante la seconda si rilevavano i beni posseduti dalle famiglie a fini fiscali. Durante le dinastie successive si tenevano elenchi delle famiglie dei soldati, dei dipendenti statali, delle merci. Durante la ventesima dinastia (1186 A.C.- 1078 A. C) le indagini portano alla stesura di liste delle abitazioni e dei loro abitanti. Il Libro dei Numeri (quarto libro della Torah ebraica e della Bibbia cristiana) racconta la storia degli ebrei durante il periodo che trascorsero nel deserto del Sinai e riporta il conteggio di tutte di tutti i figli di Israele secondo l'appartenenza alle singole famiglie, la discendenza paterna, e l'elenco dei maschi dai venti anni in su al fine di scegliere quelli adatti a combattere.

Quelle ricordate sono le prime indagini di tipo statistico che prenderanno successivamente nell'antica Roma il nome di **censimenti** (dal verbo latino **censere**, che significa **dichiarare, accertare, valutare**): a partire dalla fine del sesto secolo avanti Cristo i censimenti furono utilizzati per contare gli abitanti dello stato, determinarne l'appartenenza ad una classe sociale, al fine anche di determinare la quantità di tasse dovute. I censimenti furono poi estesi anche alle province conquistate da Roma; Gesù Cristo nacque durante uno dei censimenti voluti dall'imperatore Augusto.

I censimenti e le prime indagini statistiche hanno essenzialmente il carattere di semplice raccolta di dati e conteggio. La statistica si è in seguito sviluppata come *scienza dei fenomeni collettivi*, capace di analizzare i risultati e descriverli in termini matematici.

**Le origini della statistica moderna.** Nelle scienze sperimentali (economia fisica, scienze del territorio etc..) il metodo scientifico, che ha come fine la conoscenza di un dato fenomeno, si articola in tre fasi:

- *raccolta dei dati;*
- *correlazione dei dati attraverso l'elaborazione di un modello matematico (o teoria) che leghi i dati in modo preciso e coerente;*
- *uso del modello matematico come modello predittivo.*

L'uso del modello matematico come strumento per predire i risultati di altri esperimenti, ha anche il fine di verificare l'affidabilità del modello stesso.

La statistica moderna si articola secondo le tre fasi proprie del metodo scientifico ma è rivolta allo studio dei fenomeni collettivi in condizioni di incertezza e non determinismo (cioè nel caso che il fenomeno possa presentarsi secondo modalità diverse); essa determina i modi in cui, attraverso strumenti matematici, il fenomeno, nelle sue diverse modalità, può essere sintetizzato e quindi compreso. Essa è di supporto per molte discipline applicative come la fisica, l'economia etc.

Le origini della statistica moderna quale “**scienza del collettivo**” in cui i dati sono analizzati e descritti in termini matematici, si fanno risalire al XVII° o XVIII° secolo con lo sviluppo della demografia (dal greco **demos**, cioè popolo, popolazione, e **graphia**, scrittura, descrizione), analisi quantitativa dei fenomeni di stato e di movimento della popolazione.

### **Alcune date i riferimento per la nascita e lo sviluppo della statistica moderna**

**-1662: J. Graunt** (mercante londinese (1620 - 1674)) pubblica in Inghilterra : **“Osservazioni naturali e politiche fatte sui bollettini di mortalità”** .

Partendo dalla registrazione delle nascite e delle morti Graunt:

- a) *suddivide le informazioni per classi omogenee di individui;*
- b) *rileva caratteristiche di regolarità nei fenomeni indagati (ad esempio la prevalenza di nascite maschili su quelle femminili in rapporto di 14 a 13 nella città di Londra e di 15 a 14 in campagna);*
- c) *usa i dati e le regolarità osservate per risolvere problemi di previsione.*

John Graunt fu il primo a calcolare, sulla base di registrazione settimanale delle morti, la probabilità di morte in funzione dell'età. Alcuni anni più tardi Jan Hudde (1628- 1704) e Jan de Witt (1625-1672) in Olanda pervennero a conclusioni simili e utilizzarono il calcolo delle probabilità di morte in funzione dell'età alla determinazione delle assicurazioni sulla vita.

**-1671 W. Petty** (economista e matematico inglese) pubblica i **“Saggi di aritmetica politica”** “ nei quali pone in rilievo l'importanza delle cifre e del ragionare su di esse per il governo dei fenomeni sociali cui le cifre si riferiscono:

*“l'aritmetica politica è l'arte di ragionare con cifre su cose che riguardano il governo”*

La demografia e l'indagine sulla ricchezza della popolazione (elemento di forza per competere con altri stati) furono i principali ambiti di indagine per gli aritmeti politici. Tra l'altro **Petty**, che riprende il metodo statistico elaborato da Graunt per il settore demografico giunge alla conclusione che causa fondamentale delle disuguaglianze nella ricchezza è la distribuzione della popolazione sul territorio.

**-1812 P.S. Laplace** (astronomo e matematico francese (1749-1827)), pubblica *“Theory analytique des probabilités”* rivolto ai matematici e seguito nel **1814** da *“Essay philosophique sur les probabilités”*, in cui spiega, rivolto a un pubblico non specialistico, gli scopi e le applicazioni della teoria della probabilità.

**-1854 G. Boole**, pubblica *“Un'indagine delle leggi del pensiero sulle quali sono fondate le teorie matematiche della logica e della probabilità”*, opera nella quale applica l'algebra alle proposizioni logiche e alla probabilità e butta le basi per la classificazione multipla, di grande importanza nella rappresentazione dei dati rilevati nell'indagine statistica.

**-1869 A. L. Quételet**, astronomo belga, pubblica l'opera in quattro volumi *“Fisica sociale, saggio sullo sviluppo delle facoltà dell'uomo”* in cui tra l'altro sviluppa la teoria dell'“*uomo medio*”:  
*“le caratteristiche fisiche, morali, intellettuali di una popolazione vanno riferite ad un individuo astratto nel quale le caratteristiche assumono il valore medio dei valori posseduti da tutti gli individui della popolazione”*.

**- 1889 F. Galton** (1822-1911) pubblica  
*“Natural Inheritance”*

in cui, tra l'altro indaga sull'ereditarietà della statura, scoprendo la regressione (verso la media) della statura dei figli rispetto a quella dei padri e dando così origine allo studio della correlazione tra caratteri.

**- 1893 - 1912 K. Pearson**, (1857-1936) pubblica 18 lavori su  
*“Mathematical Contribution to the Theory of Evolution”*

che alla correlazione ha dato un contributo fondamentale, con l'introduzione dell'indice di correlazione  $\chi^2$ .

Ricordiamo infine che nel 1861, in Italia si costituì un Ufficio Statistico Nazionale, che poi diventò ISTAT nel 1926.

## 2. TERMINOLOGIA

**POPOLAZIONE O UNIVERSO O COLLETTIVO STATISTICO:** è l'insieme o collettivo

$$\Omega = \{\dots\dots\dots\}$$

di oggetti, individui, osservazioni o altro su cui vengono indagati i fenomeni di interesse .

L'uso della parola **Popolazione** in statistica deriva storicamente dal fatto che le prime statistiche venivano fatte per descrivere fenomeni demografici o economici riguardanti la popolazione di un paese o stato.

**UNITÀ STATISTICA O INDIVIDUO O CASO O OSSERVAZIONE:** è ognuno degli elementi costituenti la popolazione

$$\Omega = \{\dots u_{\dots}\} \text{ unità statistica.}$$

**CAMPIONE:** è una parte rappresentativa della popolazione su cui è effettivamente svolta l'indagine relativa al fenomeno di interesse, quando, per motivi vari, l'indagine non può essere svolta su tutta popolazione.

**CARATTERE:** è il fenomeno, caratteristica o proprietà osservabile sugli individui di una popolazione, il cui diverso modo di manifestarsi è oggetto di indagine.

**MODALITÀ DI UN CARATTERE:** è ognuno dei modi in cui si può manifestare il carattere negli individui della popolazione.

**DETERMINAZIONE DEL CARATTERE NELL'UNITÀ  $u$ :** è la modalità con cui si presenta il carattere nell'unità statistica  $u$ .

**STATISTICA UNIVARIATA:** studio della popolazione rispetto ad un carattere;

**STATISTICA BIVARIATA:** studio della popolazione rispetto a due caratteri al fine anche di stabilire la relazione che intercorre tra i due caratteri;

**STATISTICA MULTIVARIATA:** studio della popolazione rispetto a più caratteri e della relazione che intercorre tra essi.

### Esempi

#### 1. Indagine sul reddito delle famiglie residenti in Italia nell'anno 2010.

Il fenomeno indagato è il reddito e l'indagine è fatta sull'insieme delle famiglie residenti in Italia nell'anno 2010, quindi:

$C = \text{reddito}$

$\Omega = \{u : u \text{ è una famiglia residente in Italia nel 2010}\}$

$u = \text{famiglia res. in Italia nel 2010}$

*carattere*

*popolazione*

*unità statistica*

*I diversi valori del reddito costituiscono le **modalità** del carattere reddito.*

Se la famiglia Russo (unità statistica) ha avuto nel 2010 un reddito 30.000 euro allora **30.000** è la **determinazione del carattere reddito nell'unità statistica famiglia Russo**

Se la popolazione su cui si svolge l'indagine "distribuzione del reddito" è troppo ampia, si può indagare su campione  $X$  della popolazione ad esempio su

$$X = \{u: u \text{ è una famiglia residente nel 2010 in un capoluogo di provincia italiana}\} \\ = \{u \in \Omega: x \text{ residente in un capoluogo di provincia}\} \quad \text{campione}$$

Dall'indagine su  $X$  si inferiscono con opportuni strumenti risultati relativi a tutta la popolazione.

**2. Indagine sulle cause di malfunzionamento di un tipo di macchine utensili.** Il fenomeno o carattere indagato è "**causa di malfunzionamento**" e le diverse cause di malfunzionamento costituiscono le **modalità del carattere**: ad esempio, "fluttuazione di tensione" è una possibile causa di malfunzionamento e quindi modalità del carattere; l'indagine viene fatta sulle macchine utensili, quindi il **collettivo statistico  $\Omega$**  è **l'insieme delle macchine utensili del tipo considerato**, ed ogni macchina utensile è una unità statistica, un individuo o caso.

**3. Indagine sulle cause di mortalità infantile.** Il carattere indagato è "**causa di mortalità infantile**" e una sua possibile modalità è ad esempio "soffocamento", un'altra è "morbillo" etc.. In tal caso l'indagine può essere fatta solo guardando al passato e utilizzando i registri presenti presso gli ospedali o l'anagrafe che indicano anche le cause della morte. Allora il **collettivo statistico** è **l'insieme dei casi riportati nei registri di mortalità infantile**, e ogni suo elemento è un caso o osservazione (termine che sembra più appropriato di unità statistica).

**4. Indagine sul titolo di studio dei telespettatori che seguono il programma  $X$  in TV.** Il carattere indagato è il "**titolo di studio**" e sua modalità sono: nessun titolo, licenza elementare, licenza media etc..

La popolazione è **l'insieme degli spettatori che seguono il programma** e l'unità statistica o individuo è il singolo spettatore che appartiene alla popolazione considerata;

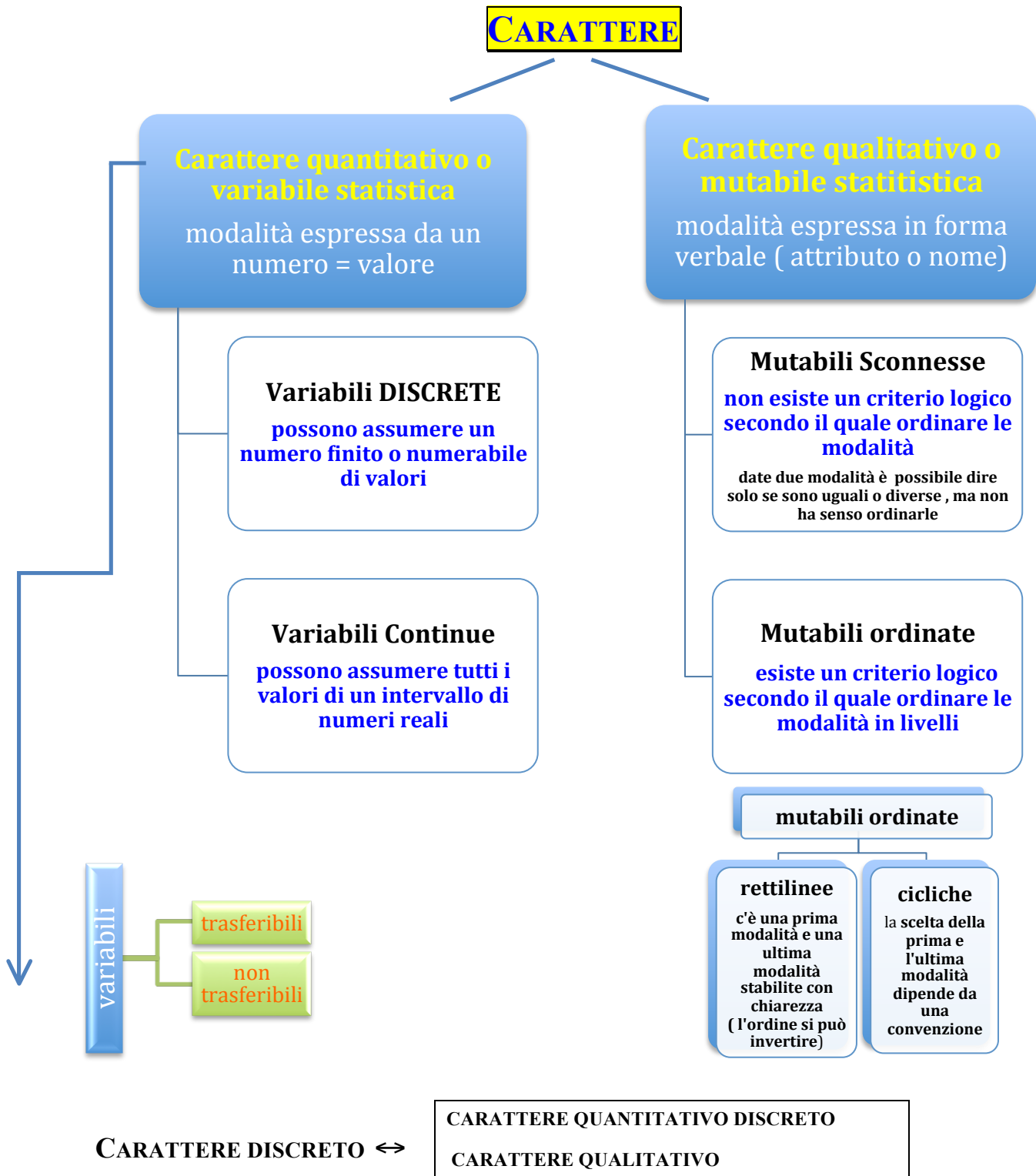
**5. Esempio di statistica bivariata.** Indagine sulla popolazione

$$\Omega = \{u : u \text{ vive in Italia e ha 60 anni o più}\},$$

con riferimento ai caratteri peso e diabete al fine di stabilire se e quanto il peso incida sulla presenza del diabete.

### 3. CLASSIFICAZIONE DEI CARATTERI

Nella scelta della metodologia statistica più adeguata all'analisi dei dati, gioca un ruolo importante la distinzione tra caratteri qualitativi e quantitativi, oltre alla natura della rilevazione, (completa o parziale cfr. sulla rilevazione statistica in appendice) e l'obiettivo di fondo dell'analisi (descrittiva o inferenziale).

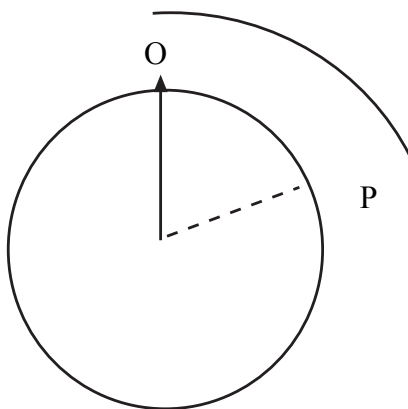


**CARATTERE QUANTITATIVO o VARIABILE STATISTICA:** **carattere le cui modalità sono espresse da un numero** (peso, altezza di un individuo, reddito di una famiglia, numero delle stanze di un appartamento, ...); quindi il carattere è una variabile numerica e le modalità sono i valori che la variabile può assumere, pertanto ogni modalità è detta **valore**.

Un carattere quantitativo è detto *carattere discreto o variabile discreta*, se può assumere una quantità *finita* di valori o una quantità *numerabile* di valori (cioè i suoi valori possono essere messi in corrispondenza biunivoca con i numeri naturali); allora i valori possono essere *elencati*, cioè possono essere indicati con una sequenza  $x_1, x_2, \dots, x_n, \dots$ ; essi sono sovente espresse da un numero intero, risultante ad esempio da un conteggio (numero delle stanze di un appartamento, numero dei figli di una coppia, numero di petali di un fiore,...). Un siffatto carattere è anche carattere *enumerabile*.

Un carattere quantitativo è detto *carattere continuo o variabile continua* se valori che può assumere sono, almeno teoricamente, tutti i valori di un intervallo dei numeri reali (durata di una corsa, peso di un individuo, durata di vita di una lampadina etc ....)

**Esempio. LA RUOTA DELLA FORTUNA**



Indagine sui risultati dei giri dell'ago della ruota della fortuna:

- il carattere è il risultato del giro
- il collettivo è l'insieme dei giri

Il risultato di un giro è visualizzato dal “ punto in cui ferma l’ago sulla circonferenza che fa da frontiera”; ogni punto è individuato dalla misura dell’arco di estremi O e P e allora il carattere può essere pensato come la variabile statistica  $X$  = “ misura dell’arco OP” e i suoi valori cadono nell’intervallo  $[0, 2\pi]$ . Oppure l’esperimento è “ arco descritto dalla freccia sulla circonferenza ” e il carattere può essere pensato come la variabile statistica  $X$  = “ misura dell’arco descritto” , la cui misura varia in  $[0, +\infty[$  .

Un carattere quantitativo è detto *trasferibile* se un’unità statistica può trasferire in parte o totalmente la quantità di carattere che possiede a una altra unità (reddito di una famiglia, numero abitanti di una regione etc...)

**CARATTERE QUALITATIVO o MUTABILE STATISTICA:** carattere le cui modalità non sono misurabili numericamente e sono espresse attraverso attributi o nomi o gradi (sesso, disponibilità verso gli altri, colore dei capelli, fiducia, carriera militare, etc...). Un carattere qualitativo è considerato un carattere discreto, che assume cioè un numero finito di modalità (ad esempio il carattere sesso assume solo due modalità: maschile, femminile; il carattere stagione assume quattro modalità: primavera, estate, autunno, inverno).

I caratteri qualitativi si distinguono in caratteri *ordinabili* e caratteri *sconnessi*.

Un carattere qualitativo è detto *ordinabile o ordinato* se tra le modalità è possibile stabilire una relazione d'ordine secondo un criterio logico (es. il grado nella carriera militare, il titolo di studio, il livello di gravità di un terremoto, i mesi dell'anno, ).

In particolare:

- il carattere è detto *ordinato rettilineo o mutabile rettilinea* se è logicamente individuabile una gerarchia di livelli, con primo ed ultimo livello ben determinati, lungo i quali disporre le modalità del carattere (titolo di studio, gradi di una gerarchia militare....);

- il carattere è detto *ordinato ciclico o mutabile ciclica*, se le modalità sono ordinabili logicamente in una gerarchia, ma il primo e ultimo livello della gerarchia non sono determinati in modo univoco (giorni della settimana, mesi dell'anno<sup>(\*)</sup>,...).

Un carattere qualitativo è detto *sconnesso* quando non esiste un criterio logico che individui un ordinamento nell'insieme delle modalità (es. il colore degli occhi, il sesso etc...): le modalità possono essere allora elencate in modo arbitrario.

**Esercizio-** *Classificare i caratteri: sesso, età (in anni compiuti), stato civile, numero di clienti in coda in banca, titolo di studio, peso, numero di fratelli, nazionalità, grado di soddisfazione di un certo servizio....*

NOTA- Nella terminologia anglosassone:

**Carattere qualitativo =fattore**

**Modalità= livello del fattore**

---

<sup>(\*)</sup> ad esempio gli estremi dell'ordinamento per i mesi dell'anno cambiano se si pensa all'anno civile o a quello scolastico

- **Esempi**

1. Indagine sul *numero dei componenti* delle famiglie residenti in Italia nell'anno 2010:  
*c = numero componenti*                      *carattere quantitativo discreto o variabile discreta*
  
2. Indagine sul *durata di funzionamento di una lampadina a led di una data marca*  
*c = durata funzionamento di una lampadina a led*    *carattere quantitativo continuo o*  
*variabile continua*
  
3. Indagine sul *reddito* delle famiglie residenti in Italia nell'anno 2010:  
*c = reddito*                                      *carattere quantitativo* che può  
essere trattato come variabile continua
  
3. Indagine sulle incidenze di una malattia, la cataratta, negli individui che hanno 60 anni o più in Italia:  
*c = cataratta*                                      *carattere qualitativo sconnesso*  
*o mutabile sconnessa*  
con due modalità: **presente, non presente**
  
4. Indagine sul *livello di istruzione* degli abitanti della città di Napoli  
*c= livello di istruzione*                                      *mutabile ordinata rettilinea*  
*le cui modalità sono date dai titoli di studio*  
*e dalla modalità "nessun titolo" se tra gli abitanti di Napoli*  
*c'è chi non ha preso nessun titolo di studio*

Sia per le variabili quantitative che e che per le mutabili qualitative un'ulteriore distinzione viene fatta suddividendo tali variabili in variabili di stato e in variabili di flusso

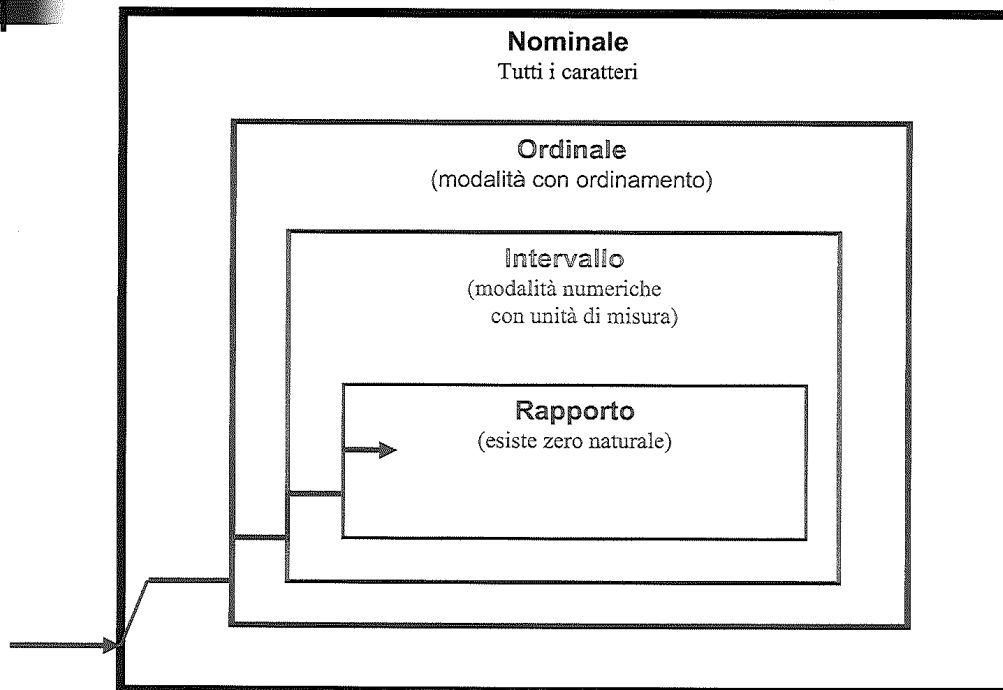
**Variabili di stato:** sono i caratteri che vanno rilevati per avere la fotografia precisa di un dato collettivo in un determinato istante (ad esempio il numero degli iscritti all'università nelle varie sedi universitarie italiane nell'anno 2003 – 2004; tale carattere si riferisce alla popolazione degli iscritti alle università per l'anno accademico considerato e viene rilevato in una data successiva al termine fissato per le iscrizioni per fotografare la situazione per quell'anno accademico). Il collettivo per il quale è rilevato un siffatto carattere è a sua volta un **collettivo di stato**

**Variabili di flusso:** sono i caratteri che vengono rilevati in almeno due momenti diversi per dar conto dell'evolversi di un fenomeno in un dato periodo, e dei cambiamenti che intervengono nel collettivo, che quindi non è un collettivo statico, potendo le unità di esso cambiare, aumentare o diminuire, ed è perciò detto **collettivo di flusso** (ad esempio il numero dei nuovi iscritti alle università italiane in un dato anno accademico richiede due rilevazioni: il numero degli iscritti nel precedente anno accademico e il numero degli iscritti nel nuovo anno: il numero dei nuovi iscritti è una variabile di flusso che dà idea del cambiamento che avviene nel periodo di un anno nella popolazione universitaria).

## 4. SCALE DI MISURA

Un carattere quantitativo o qualitativo presenta variabilità, perché può manifestarsi secondo diverse modalità; l'operazione di "misurare" le modalità con cui un carattere si presenta può consistere in una semplice operazione di classificazione o riconoscimento delle modalità attraverso nomi, attributi e anche numeri (così da distinguere una modalità dall'altra), può essere più sofisticata e permettere, oltre la classificazione, l'ordinamento delle modalità, può infine attribuire a ogni modalità un numero, che non solo serve per individuarla, ma ha il significato di "peso o valore numerico" della modalità. Si hanno quindi diverse scale di misurazione e l'adozione dell'una o dell'altra dipende dal carattere considerato; tale scale si possono disporre in una gerarchia che va dalla scala più semplice a quella più complessa ed è tale che le operazioni permesse da una scala di misura sono incluse in quelle permesse da una scala di misura più complessa e quindi disposta ad un livello superiore della gerarchia

### Classificazione gerarchica delle scale di misurazione



## 4.1 ILLUSTRAZIONE DELLE DIVERSE SCALE DI MISURA

### A. SCALE RIFERITE PRINCIPALMENTE ALLE MUTABILI STATISTICHE

**Scala Nominale e Scala ordinale:** sono riferite principalmente ai caratteri qualitativi le cui modalità sono espresse da nomi o aggettivi.

**A<sub>1</sub>. Scala Nominale o Classificatoria:** è la scala più semplice che *distingue, classifica* le modalità di un carattere attraverso il nome o il numero che li identifica; quindi essa è particolarmente riferita a caratteri qualitativi le cui modalità sono identificate attraverso nomi o aggettivi e sono tali che non ha (necessariamente) senso stabilire tra esse un ordinamento di preferenza. Tuttavia in tale scala possono essere considerati anche i caratteri quantitativi.

Alla scala classificatoria è collegata anche l'operazione di classificazione delle unità statistiche o individui attraverso le modalità, cioè la costruzione di classi di unità statistiche, ognuna delle quali è costituita da tutte e sole le unità che presentano la stessa modalità

In tale scala nessuna relazione esiste tra le modalità o tra le classi di unità e gli unici simboli matematici utilizzabili per confrontare due classi sono i segni = e ≠.

**Esempio:**  $\Omega = \{x : x \text{ vive in Italia}\}$

#### 1. Carattere: sesso                      modalità: F , M

Gli individui sono suddivisi in due classi ognuna individuata da una modalità del carattere  *sesso*:

<b>F</b>	<b>M</b>
----------	----------

due individui appartengono alla stessa classe se e solo se sono dello stesso sesso

#### 2. Carattere: zona di residenza

**modalità: Nord, Centro, Sud, Isole**

Gli individui sono suddivisi in quattro classi, ognuna individuata da una modalità del carattere  *zona di residenza*

<b>Nord</b>	<b>Centro</b>	<b>Sud</b>	<b>Isole</b>
-------------	---------------	------------	--------------

due individui appartengono alla stessa classe se e solo se risiedono nella stessa zona .

La **mutabile sconnessa** è l'unico carattere misurabile solo in una scala nominale

**A<sub>2</sub>. Scala ordinale:** è la scala che *classifica le modalità e le ordina*, pertanto è *particolarmente riferita alle mutabili ordinate le cui modalità, identificate attraverso nomi o attributi, sono ordinabili logicamente, tuttavia essa è utilizzabile anche per i caratteri quantitativi*; mentre non sono ad essa riferibili le mutabili sconnesse.

A una scala ordinale è collegata sia l'operazione di suddivisione delle unità statistiche in classi (ognuna costituita dalle unità statistiche che presentano la stessa modalità) sia l'operazione di ordinamento di tali classi, corrispondente all'ordinamento delle modalità secondo cui sono determinate tali classi.

Nella gerarchia delle scale, *la scala ordinale si pone a un livello superiore a quella nominale perché realizza la stessa classificazione della scala nominale e in più le classi vengono allineate in un ordinamento*. Tra due classi (oltre ai simboli = e ≠) si può porre il simbolo < ( risp, >) per dire che la prima classe precede ( risp. segue) la seconda nell'ordinamento,

**Esempio:**  $\Omega = \{x : x \text{ vive in Italia}\}$

**Carattere:** titolo di studio

<b>Modalità:</b> nessun titolo	<b>liv. 0</b>
Lic. Elementare	<b>liv. 1</b>
Lic. media inferiore	<b>liv. 2</b>
Lic. media superiore	<b>liv. 3</b>
Laurea triennale	<b>liv. 4</b>
Laurea specialistica	<b>liv.5</b>

n.t.	Lic. elementare	Lic. media infer.	Lic . media super.	Laurea triennale	Laurea specialistica
------	--------------------	-------------------------	--------------------------	---------------------	-------------------------

0	1	2	3	4	5
---	---	---	---	---	---

**Codifica delle modalità di un carattere qualitativo**

*Allo scopo di inserire i dati in un computer o rendere più semplice rispondere a un questionario i nomi o aggettivi che indicano le modalità di una mutabile sono sostituiti da numeri o codici numerici; ad esempio, per le zone di residenza, si può porre Nord= 1, Centro =2, Sud= 3, Isole=4. L'attribuzione di numeri per un carattere sconnesso può avvenire in un ordine qualsiasi: si può porre ad esempio seguire la seguente attribuzione di codici alternativa alla precedente: Sud= 1, Centro =2, Nord= 3, Isole=4.*

*Per un carattere ordinato invece l'attribuzione di codici numerici è fatto in modo che tali codici riflettano una scala di valori o di importanza (vedi esempio sul titolo di studio le cui modalità si traducono in livelli o gradi di istruzione).*

*La codifica è in ogni caso un fatto strumentale che non trasporta alle modalità le proprietà dei numeri, se non le proprietà dell'ordine limitatamente al caso di un carattere ordinato: non ha senso perciò sommare o moltiplicare i codici che rappresentano le modalità di un carattere qualitativo o mutabile. La codifica delle modalità può essere fatta attribuendo alle modalità lettere invece di numeri.*

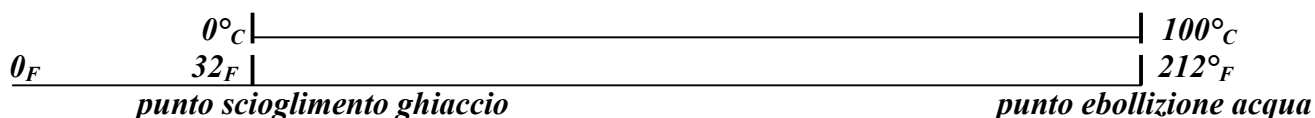
## B. SCALE RIFERITE SOLO ALLE VARIABILI STATISTICHE

**Scala a intervalli e Scala di Rapporti:** sono scale riferite a caratteri quantitativi le cui modalità possono essere misurate una volta che sia stata fissata una unità di misura ed una origine.

**B<sub>1</sub>. Scala a intervalli:** una *scala ad intervalli* fa riferimento a un intervallo numerico diviso in sotto intervalli uguali che individuano l'unità di misura, mentre l'origine è convenzionale e non determinata in modo assoluto: si possono perciò determinare più scale di misura con origini diverse. Una grandezza misurabile in una scala a intervalli è la temperatura: essa è misurata sia nella scala Celsius che in quella Fahrenheit, che hanno origini e unità di misura diverse. In conseguenza della scelta opzionale dell'origine del sistema di misura, si possono confrontare e rapportare tra loro le distanze tra valori (ottenute attraverso una differenza) ma non ha senso fare il rapporto di valori, come illustra il seguente esempio.

**Esempio.** Nella scala Celsius l'intervallo di riferimento è quello compreso tra il punto di scioglimento del ghiaccio e il punto di ebollizione dell'acqua: l'intervallo viene suddiviso in cento parti uguali (gradi centigradi: °C) ognuna delle quali vale 1°C e rappresenta l'unità di misura: lo zero nella scala Celsius corrisponde al punto di scioglimento del ghiaccio e quindi viene posto uguale a 100° la temperatura di ebollizione dell'acqua.

Nella scala Fahrenheit punti di riferimento sono ancora riferimento il punto di scioglimento del ghiaccio e il punto di ebollizione dell'acqua: l'intervallo tra i due punti è diviso in 180 parti uguali (gradi F), ma lo zero della scala è fissato 32 unità sotto il punto di scioglimento del ghiaccio per cui a 0° C corrispondono +32° F e a 100° C corrispondono 212° F.



Le formule di passaggio tra le due scale sono le seguenti

$$F = \frac{9}{5}C + 32 \qquad C = \frac{5}{9}(F - 32)$$

I° quesito: Se le temperature di due corpi A e B sono rispettivamente di 5°C e 10°C nella scala Celsius ha senso dire in assoluto che la temperatura di B è doppia di quella di A?

NO: nella scala Fahrenheit le temperature di A e B sono rispettivamente 41°F e 50°F, ma 50 non è il doppio di 41.

Le temperature dei corpi A, B, X e Y sono sono rispettivamente di 5°C , 10°C, 20°C, 30°C nella scala Celsius. Risulta  $30 - 20 = 2(10-5)$

II° quesito: Ha senso dire che la differenza di temperatura tra i corpi Y e X è doppia della differenza di temperatura tra i corpi B ed A?

SI: nella scala Fahrenheit le temperature di A, B, X e Y sono rispettivamente 41°F, 50°F, 68°F e 86°F e risulta  $86 - 68 = 18 = 2(50 - 41)$

SCHEMA esplicativo per i due quesiti

Le temperature di due corpi A e B sono rispettivamente di 5°C e 10°C nella scala Celsius. Risulta  $10 = 2 \cdot 5$ .

Ha senso dire in assoluto che la temperatura di B è doppia di quella di A?

**NO**

	°C	F
A	5°	41
B	10°	50

$$10/5 = 2$$

$$50/41 = 1,219519512\dots$$

Le temperature dei corpi A, B, X e Y sono rispettivamente di 5°C, 10°C, 20°C, 30°C nella scala Celsius. Risulta

$$30 - 20 = 2(10 - 5).$$

Ha senso dire che la differenza di temperatura tra i corpi Y e X è doppia della differenza di temperatura tra i corpi B ed A?

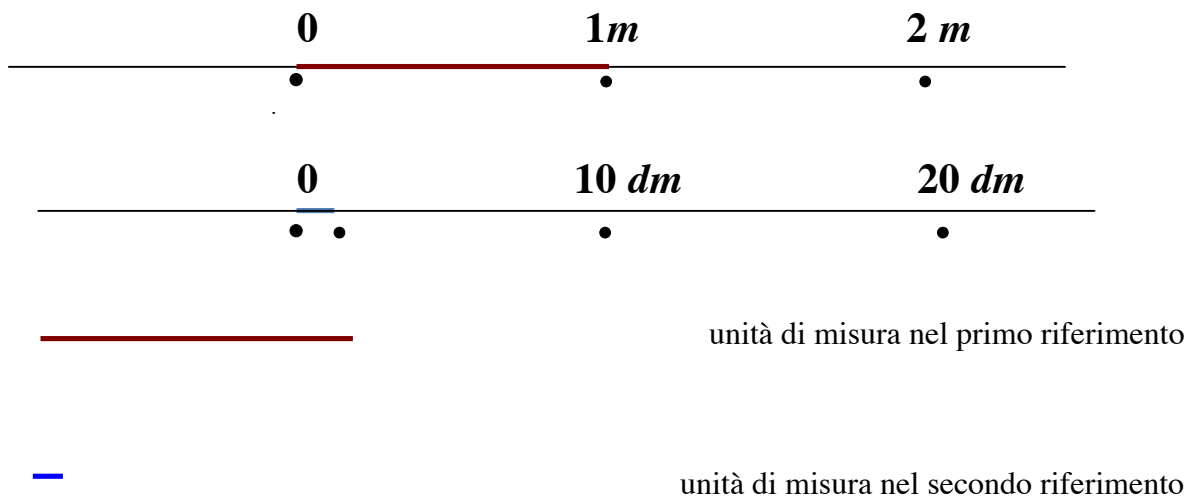
**SI**

	C	F
A	5°	41
B	10°	50
X	20°	68
Y	30°	86

$$30 - 20 = 10 = 2(10 - 5)$$

$$86 - 68 = 18 = 2(50 - 41)$$

**B<sub>2</sub>. Scala di Rapporti:** una *scala di rapporti* è una scala per caratteri quantitativi in cui è **univocamente determinata l'origine per le misure**: infatti lo **zero assoluto** viene posto nel punto in cui la caratteristica da misurare “scompare” non è più visibile. E' il caso di caratteri quali peso, altezza, lunghezza, velocità etc.. Per tali caratteri ha senso parlare di rapporto di valori perché tali valori rappresentano distanze dallo zero assoluto la cui posizione non cambia se cambia l'unità di misura (per esempio da metri si passa a decimetri). Allora il rapporto delle misure (di due modalità) non cambia se si passa alle misure effettuate in un'altra scala con la stessa origine ma unità diversa.



## 5. LE BRANCHE DELLA STATISTICA

**Rilevazione statistica.** Non da sempre considerata una vera branca della statistica, essa si interessa dei metodologie per la raccolta dei dati e informazioni relativi al fenomeno oggetto di studio.

**Statistica descrittiva.** Ha il compito di rappresentare i dati raccolti attraverso tabelle e grafici, di operarne la sintesi attraverso indici analitici (medie), e di studiarne la variabilità; ciò al fine di evidenziare le particolarità e gli elementi d'interesse del fenomeno o dei fenomeni indagati e fare confronti.

Nota: la statistica descrittiva si sviluppa abbastanza tardi, nel 1800, per la difficoltà di trovare strumenti per la classificazione rispetto a più criteri e per la mancanza di uno strumento logico per rappresentare il flusso di beni e servizi da un settore verso altri settori.

**Statistica induttiva o inferenziale.** Scopo di tale della statistica inferenziale, collegata al calcolo della probabilità, è inferire il comportamento del fenomeno sulla popolazione di riferimento, partendo da una conoscenza limitata, parziale del fenomeno, ad esempio dai dati raccolti su un campione particolare. In particolare essa fornisce le metodologie e gli strumenti a individuare i modi per:

- risalire correttamente dai dati raccolti su un campione alle caratteristiche del fenomeno sul collettivo da cui il campione è tratto, costruire quindi un **modello teorico** che spieghi i dati raccolti e sia applicabile alla popolazione (**teoria della stima del campionamento**) e sia utilizzabile come modello predittivo, capace cioè di predire risultati futuri con un certo margine di certezza;
- decidere l'appartenenza di un campione ad un collettivo, sulla base anche delle informazioni sui possibili collettivi di provenienza (**teoria delle decisioni statistiche**); ad esempio dai risultati di analisi compiute su un gruppo (campione) di pazienti risalire al collettivo di appartenenza, quello costituito dai pazienti che presentano una data patologia o quello costituito dai pazienti che non soffrono di tale patologia.

Connessa alla statistica inferenziale è la

**Statistica matematica.** Fornisce i modelli teorici che permettono di inquadrare e spiegare i dati raccolti.

In questo corso ci occuperemo di statistica descrittiva, di cui uno degli scopi è l'elaborazione e la rappresentazione dei dati raccolti attraverso una rilevazione statistica. Per tal fine utile è la classificazione delle unità statistiche a partire dalle modalità del carattere, alla quale abbiamo fatto cenno parlando di scale di misura.

## 6. LA CLASSIFICAZIONE DELLE UNITÀ STATISTICHE SECONDO LE MODALITÀ DEL CARATTERE

Per la classificazione di cui al titolo di questo paragrafo è necessario che le *modalità* di un carattere rilevate nell'indagine siano:

- a) **non sovrapponibili**: in una stessa unità non possono essere presenti due modalità distinte del carattere (una modalità esclude le altre).
- b) **esaustive**: devono esaurire tutti i modi in cui un carattere effettivamente si presenta, in altre parole *per ogni modalità  $m$  deve esistere almeno una unità statistica  $u$  in cui il carattere si presenta con quella modalità*;

Prima di iniziare un'indagine si devono stabilire le *possibili* modalità, che il carattere può presentare e tali modalità devono essere pensate in modo da non essere sovrapponibili; è solo dopo la rilevazione delle modalità sui singoli individui, si ha un insieme di modalità (quelle effettivamente rilevate), che fornisce un insieme *esaustivo*.

Supponiamo che le modalità siano in numero finito:  $m_1, m_2, \dots, m_k$ ; allora, a partire da ogni modalità  $m_j$ , costruiamo la parte del collettivo costituita da tutte e sole le unità statistiche che presentano  $m_j$  come modalità:

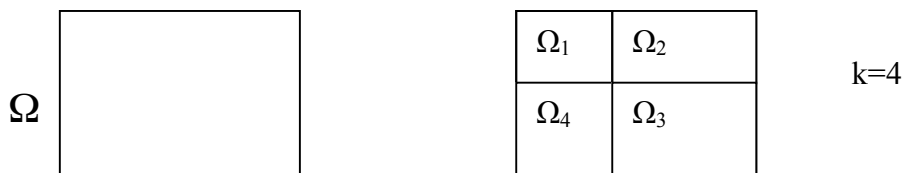
$$\Omega_j = \{u \in \Omega : \text{in } u \text{ il carattere presenta la modalità } m_j\} \quad j=1, 2, \dots, n.$$

Otteniamo così una *partizione del collettivo*  $\Omega$ :

$$\Pi = \{ \Omega_1, \Omega_2, \dots, \Omega_k \};$$

infatti le classi  $\Omega_h$ , per la proprietà **b)** di esaustività, sono *non vuote e la loro unione è uguale*  $\Omega$ , e, per la proprietà **a)** di non sovrapponibilità, sono *a due a due disgiunte*:

$$\Omega_j \neq \emptyset \quad \text{per ogni } j \quad \Omega_i \cap \Omega_j = \emptyset \quad \text{se } i \neq j \quad \cup \Omega_j = \Omega$$



La partizione  $\Pi$  è detta **classificazione** delle unità statistiche ( o di  $\Omega$  ) secondo le modalità del carattere e le parti  $\Omega_1, \Omega_2, \dots, \Omega_k$  sono dette **classi di unità statistiche**.

Nel caso di un carattere quantitativo continuo o trattato come tale perché presenta un numero rilevante di valori (= modalità) distinti, invece dei singoli valori sono presi in considerazione intervalli di valori  $I_1, I_2, \dots, I_K$  a due a due disgiunti (\*) e la classificazione delle unità statistiche si ottiene raggruppando in una stessa classe le unità che presentano un valore del carattere appartenente allo stesso intervallo:

$$\Omega_1 = \{u \in \Omega : \text{in } u \text{ il carattere prende un valore appartenente all'intervallo } I_1\}$$

.....

$$\Omega_k = \{u \in \Omega : \text{in } u \text{ il carattere prende un valore appartenente all'intervallo } I_k\}$$

(\*)---- sulle regole da seguire per la costruzione di tali intervalli vedi fascicolo 2, seconda parte

## Esempi di classificazione delle unità statistiche secondo le modalità del carattere

1. Riprendiamo l'esempio 2 del paragrafo 2. relativo *all'indagine sulle cause di malfunzionamento di un tipo di macchine utensili*.

*Le diverse cause di malfunzionamento costituiscono le modalità del carattere e suddividono la popolazione  $\Omega$  di macchine utensili in classi ognuna delle quali è costituita da tutte le macchine utensili il cui malfunzionamento è dovuto alla stessa causa.*

2. Riprendiamo l'esempio 3 del paragrafo 2. relativo *all'indagine sulle cause di mortalità infantile*.

*Le diverse cause di morte (caduta, soffocamento, morbillo etc..) sono le modalità del carattere; la classificazione dei casi riscontrati di mortalità infantile avviene secondo le cause di morte: in una stessa classe sono raggruppati i casi di morte avvenuta per la stessa causa, ad esempio morbillo.*

3. Riprendiamo l'esempio 4 del paragrafo 2. relativo *all'indagine sul titolo di studio dei telespettatori che seguono il programma X in TV*.

*Le modalità sono i diversi titoli di studio rilevati e gli spettatori sono classificati secondo il loro titolo di studio: in una stessa classe sono raggruppati gli spettatori con lo stesso titolo di studio*

4. Riprendiamo l'esempio 1 del paragrafo 2. relativo *all'indagine sul reddito delle le famiglie residenti in Italia nell'anno 2010*:

*I diversi valori del reddito costituiscono le modalità del carattere reddito e suddividono  $\Omega$  in classi ognuna delle quali è costituita da tutte le famiglie che hanno lo stesso reddito: le unità statistiche vengono quindi raggruppate e classificate secondo il reddito.*

*Se i valori distinti del reddito rilevati sono in quantità elevata e quindi non facilmente gestibile conviene considerare non i singoli valori ma intervalli di valori, quali, ad esempio,*

$I_1 = [0, 20.000[$  ,  $I_2 = [20.000, 40.000[$ ,  $I_3 = [40.000, 60.000[$ ,  $I_4 = [60.000, 100.000[$

$I_5 = [100.000, 200.000[$ ,  $I_6 = [200.000, 400.000[$ ,  $I_7 = [400.000, 600.000[$ ,  $I_8 = [600.000, 1.000.000[$

$I_9 = [1.000.000, 10.000.000[$  ,  $I_{10} = \{x \in R : x \geq 10.000.000\} = [10.000.000, +\infty [$  .

Allora la classificazione di  $\Omega$  è fatta raggruppando in una stessa classe  $\Omega_h$  tutte e sole le famiglie che hanno un reddito appartenente ad uno stesso intervallo di valori  $I_h$ .

**ESERCIZIO.** Nella tabella della pagina seguente sono riportati i dati di un'indagine fatta sul collettivo degli iscritti a un corso di perfezionamento relativamente alla loro facoltà di provenienza (carattere): accanto al nome e cognome di ogni iscritto è riportata la facoltà di provenienza

*Le modalità sono le diverse facoltà di provenienza e gli iscritti al corso vanno classificati secondo tali facoltà, cioè in una stessa classe vanno raggruppati gli iscritti provenienti dalla stessa facoltà.*

<b>COGNOME E NOME</b>	<b>PROVENIENZA</b>
<b>RUSSO GIUSEPPINA</b>	<b>INGEGNERIA CIVILE</b>
<b>BIANCHI ANNA</b>	<b>MATEMATICA</b>
<b>RUSSO MICHELE</b>	<b>ARCHITETTURA</b>
<b>PESCE SALVATORE</b>	<b>ARCHITETTURA</b>
<b>DE ROSSI MARIA</b>	<b>ARCHITETTURA</b>
<b>SEPE ASSUNTA</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO UMBERTO</b>	<b>ARCHITETTURA</b>
<b>CIMMINO ROBERTO</b>	<b>ECONOMIA E COMM.</b>
<b>RUSSO MARIA</b>	<b>SCIENZE GEOLOGICHE</b>
<b>IACCARINO PASQUALE</b>	<b>INGEGNERIA EDILE</b>
<b>ESPOSITO FABIO</b>	<b>ARCHITETTURA</b>
<b>BRUNO ROSA</b>	<b>ARCHITETTURA</b>
<b>MORRA GABRIELLA</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO RAFFAELLA</b>	<b>ARCHITETTURA</b>
<b>SASSO SANDRA</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO MARIANNA</b>	<b>ARCHITETTURA</b>
<b>CARDONE VINCENZO</b>	<b>ARCHITETTURA</b>
<b>DE MARCO FIORELL</b>	<b>INGEGNERIA AMBIENTALE</b>
<b>DE ROSSI ROCCO</b>	<b>ARCHITETTURA</b>
<b>BIANCHI MAURIZIO</b>	<b>ARCHITETTURA</b>
<b>IACCARINO DIANA</b>	<b>SCIENZE POLITICHE</b>
<b>SQUILLO GABRIELLA</b>	<b>SCIENZE STATISTICHE</b>
<b>CARDONE ROSA</b>	<b>ARCHITETTURA</b>
<b>MORRA MANUELA</b>	<b>ECONOMIA E COMM.</b>
<b>SQUILLO MARIA</b>	<b>ARCHITETTURA</b>
<b>BRUNO CIRO</b>	<b>SCIENZE GEOLOGICHE</b>
<b>DI LELLO MARIO</b>	<b>ARCHITETTURA</b>
<b>GATTI ILARIA</b>	<b>ARCHITETTURA</b>
<b>LEONE DANILA</b>	<b>ARCHITETTURA</b>
<b>RICCIO CARMINE</b>	<b>ARCHITETTURA</b>

**TABELLA A**

Per giungere alla classificazione delle unità statistiche secondo le modalità del carattere “provenienza” *evidenziamo con lo stesso colore le unità che presentano la stessa modalità del carattere*

<b>COGNOME E NOME</b>	<b>PROVENIENZA</b>
<b>RUSSO GIUSEPPINA</b>	<b>INGEGNERIA CIVILE</b>
<b>BIANCHI ANNA</b>	<b>MATEMATICA</b>
<b>RUSSO MICHELE</b>	<b>ARCHITETTURA</b>
<b>PESCE SALVATORE</b>	<b>ARCHITETTURA</b>
<b>DE ROSSI MARIA</b>	<b>ARCHITETTURA</b>
<b>SEPE ASSUNTA</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO UMBERTO</b>	<b>ARCHITETTURA</b>
<b>CIMMINO ROBERTO</b>	<b>ECONOMIA E COMM.</b>
<b>RUSSO MARIA</b>	<b>SCIENZE GEOLOGICHE</b>
<b>IACCARINO PASQUALE</b>	<b>INGEGNERIA EDILE</b>
<b>ESPOSITO FABIO</b>	<b>ARCHITETTURA</b>
<b>BRUNO ROSA</b>	<b>ARCHITETTURA</b>
<b>MORRA GABRIELLA</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO RAFFAELLA</b>	<b>ARCHITETTURA</b>
<b>SASSO SANDRA</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO MARIANNA</b>	<b>ARCHITETTURA</b>
<b>CARDONE VINCENZO</b>	<b>ARCHITETTURA</b>
<b>DE MARCO FIORELLA</b>	<b>INGEGNERIA AMBIENTALE</b>
<b>DE ROSSI ROCCO</b>	<b>ARCHITETTURA</b>
<b>BIANCHI MAURIZIO</b>	<b>ARCHITETTURA</b>
<b>IACCARINO DIANA</b>	<b>SCIENZE POLITICHE</b>
<b>SQUILLO GABRIELLA</b>	<b>SCIENZE STATISTICHE</b>
<b>CARDONE ROSA</b>	<b>ARCHITETTURA</b>
<b>MORRA MANUELA</b>	<b>ECONOMIA E COMM.</b>
<b>SQUILLO MARIA</b>	<b>ARCHITETTURA</b>
<b>BRUNO CIRO</b>	<b>SCIENZE GEOLOGICHE</b>
<b>DI LELLO MARIO</b>	<b>ARCHITETTURA</b>
<b>GATTI ILARIA</b>	<b>ARCHITETTURA</b>
<b>LEONE DANILA</b>	<b>ARCHITETTURA</b>
<b>RICCIO CARMINE</b>	<b>ARCHITETTURA</b>
<b>TOTALE 30</b>	

Le modalità distinte rilevate sono in numero di nove e determinano una partizione del collettivo in nove classi di unità statistiche

TABELLA B

$\Omega_1$ classe di Architettura =	<b>RUSSO MICHELE</b> <b>PESCE SALVATORE</b> <b>DE ROSSI MARIA</b> <b>SEPE ASSUNTA</b> <b>ESPOSITO UMBERTO</b> <b>ESPOSITO FABIO</b> <b>BRUNO ROSA</b> <b>MORRA GABRIELLA</b> <b>ESPOSITO RAFFAELLA</b> <b>SASSO SANDRA</b> <b>ESPOSITO MARIANNA</b> <b>CARDONE VINCENZO</b> <b>DE ROSSI ROCCO</b> <b>BIANCHI MAURIZIO</b> <b>CARDONE ROSA</b> <b>SQUILLO MARIA</b> <b>DI LELLO MARIO</b> <b>GATTI ILARIA</b> <b>LEONE DANILA</b> <b>RICCIO CARMINE</b>
$\Omega_2$ classe di Ingegneria civile =	<b>RUSSO GIUSEPPINA</b>
$\Omega_3$ classe di scienze biologiche =	<b>RUSSO MARIA</b> <b>BRUNO CIRO</b>
$\Omega_4$ classe di Ingegneria ambientale =	<b>DE MARCO FIORELLA</b>
$\Omega_5$ classe di Economia e Commercio =	<b>CIMMINO ROBERTO</b> <b>MORRA MANUELA</b>
$\Omega_6$ classe di Matematica =	<b>BIANCHI ANNA</b>
$\Omega_7$ classe di Scienze statistiche =	<b>SQUILLO GABRIELLA</b>
$\Omega_8$ classe di Scienze politiche =	<b>IACCARINO DIANA</b>
$\Omega_9$ classe di Ingegneria edile =	<b>IACCARINO PASQUALE</b>
	<b>30</b>

La classificazione ci permette di risalire a informazioni quali: quante unità statistiche presentano una certa modalità del carattere

PROVENIENZA $m_i$	CLASSI DI UNITÀ STATISTICHE $C(m_i)$	NUMEROSITÀ della classe $ C(m_i) $
ARCHITETTURA	RUSSO MICHELE PESCE SALVATORE DE ROSSI MARIA SEPE ASSUNTA ESPOSITO UMBERTO ESPOSITO FABIO BRUNO ROSA MORRA GABRIELLA ESPOSITO RAFFAELLA SASSO SANDRA ESPOSITO MARIANNA CARDONE VINCENZO DE ROSSI ROCCO BIANCHI MAURIZIO CARDONE ROSA SQUILLO MARIA DI LELLO MARIO GATTI ILARIA LEONE DANILA RICCIO CARMINE	20
ING. CIV.	RUSSO GIUSEPPINA	1
SCIENZE GEOLOGICHE	RUSSO MARIA BRUNO CIRO	2
ING. AMB.	DE MARCO FIORELLA	1
ECON. E COMM.	CIMMINO ROBERTO MORRA MANUELA	2
MAT.	BIANCHI ANNA	1
SCIENZE STATISTICHE	SQUILLO GABRIELLA	1
SCIENZE POLITICHE	IACCARINO DIANA	1
ING. EDILE	IACCARINO PASQUALE	1
		30

## 7. SULLA RILEVAZIONE STATISTICA

**RILEVAZIONE TOTALE:** si effettua su tutte le unità della popolazione che interessa studiare

**RILEVAZIONE PARZIALE O RILEVAZIONE PER CAMPIONE:** si effettua solo su una parte degli elementi della popolazione, che è detta *campione*.

**Ampiezza del campione:** numero delle unità costituenti il campione.

**Rapporto Campionario:** rapporto tra ampiezza del campione e il numero delle unità della popolazione

Le rilevazioni **totali** sono dette

- **generalì** se vengono enumerate e misurate tutte le unità statistiche di qualsiasi genere che costituiscono le popolazioni;
- **speciali** se vengono in qualche modo considerate e misurate sole le unità con date peculiarità.

**Unità di rilevazione:** non coincide necessariamente con l'unità statistica. Nel censimento i rilevatori prendono per comodità di riferimento le famiglie (unità di rilevazione) ma in realtà rilevano le notizie su tutti i componenti della famiglia che sono le unità statistiche.

**Rilevazione diretta:** si rilevano direttamente le modalità del carattere che interessa sulle unità statistiche (ad esempio si rileva il reddito delle famiglie intervistando direttamente le famiglie).

**Rilevazione indiretta:** si indaga su un altro carattere connesso al carattere di interesse e dalle informazioni raccolte si risale alle informazioni sul carattere oggetto reale dell'indagine (per risalire alla quantità di cittadini andati in vacanza ad agosto nelle diverse città di Italia si misura la quantità di nettezza urbana che si suppone proporzionale ai cittadini presenti in città, così come per misurare la *temperatura* si misura la *dilatazione* di una colonnina di mercurio)

**I dati raccolti in una rilevazione statistica si distinguono in**

**Dati sperimentali:** sono i dati raccolti direttamente dopo aver predeterminato le condizioni sotto le quali si svolge la rilevazione e il un metodo di rilevazione: intervista, creazione di una situazione, esperimento di laboratorio etc..

**Dati osservazionali:** sono i dati raccolti attraverso osservazioni di fenomeni o fatti non predisposti all'uopo o dati rilevabili da una lista dati già registrati (in archivi comunali o di centro studi o di società etc.) precedentemente raccolti per altri indagini

## 8. RISULTATI DI UNA RILEVAZIONE STATISTICA: TABELLE DI DISTRIBUZIONE UNITARIA O DI DATI GREZZI

Il risultato finale di una rilevazione statistica su uno o più fenomeni indagati su una popolazione o campione  $\Omega = \{u_1, \dots, u_n\}$  di  $n$  unità è una tabella di dati *grezzi*.

- Se il carattere indagato è uno solo, la tabella si presenta nella forma

*“unità/ modalità del carattere”* ed è detta **tabella di distribuzione unitaria semplice**: essa distribuisce le modalità unità per unità.

**Tabella di distribuzione unitari semplice ( forma unità/ modalità carattere):**

<i>Unità</i>	<i>Modalità Carattere nell'unità</i>
$u_1$	$m(u_1)$
$u_2$	$m(u_2)$
$\vdots$	$\vdots$
$u_i$	$m(u_i)$
$\vdots$	$\dots$
$u_n$	$m_n$

<i>Unità</i>	<i>Modalità Carattere nell'unità</i>
<b>1</b>	$m(1)$
<b>2</b>	$m(2)$
$\vdots$	$\vdots$
<b>i</b>	$m(i)$
$\vdots$	$\vdots$
<b>n</b>	$m(n)$

Nella fascia verticale a sinistra della prima tabella sono elencate le unità della popolazione o del campione e accanto ad ognuna di esse la modalità presa; nella seconda tabella le unità statistiche sono indicate attraverso un codice numerico in quanto al fine dell'indagine statistica non interessa l'identità dell'unità statistica. Nella tabella che segue le unità statistiche sono invece indicate nella prima riga e sotto ogni unità statistica è indicata la modalità del carattere che essa presenta.

<i>Unità</i>	$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$\dots$	$u_n$
<i>Mod. carattere</i>	$m(u_1)$	$m(u_2)$	$m(u_3)$	$m(u_4)$	$m(u_5)$	$\dots$	$m(u_n)$

**n= ampiezza di  $\Omega$**

La lettera  $x$  è preferita alla lettera  $m$  quando si tratta di un carattere quantitativo o variabile statistica e si parla allora di tabella nella forma *unità/ valori carattere*

<i>Unità</i>	<i>Valori Carattere nell'unità</i>
$u_1$	$x(u_1)$
$u_2$	$x(u_2)$
$\vdots$	$\vdots$
$u_i$	$x(u_i)$
$\vdots$	$\vdots$
$u_n$	$x(u_n)$

Ogni tabella rappresenta la funzione che a ogni unità associa la modalità del carattere in essa presente, funzione chiamata **distribuzione unitaria semplice**:

$$d: u_i \in \Omega \rightarrow m(u_i) \quad \text{distribuzione unitaria semplice}$$

La *distribuzione unitaria semplice* può essere indicata anche solo elencando una dopo l'altra le modalità così come sono state rilevate

$$m_1 \quad m_2 \quad m_3 \quad \dots \quad m_i, \quad \dots \quad m_n; \quad (1)$$

allora:  $m_1$  è la modalità del carattere rilevata nell'unità statistica  $u_1$ ,  $m_2$  la modalità del carattere rilevata nell'unità statistica  $u_2$ , etc.. ). Poiché su unità statistiche diverse può essere stata rilevata la stessa modalità, nell'allineamento (1) una stessa modalità può presentarsi più volte (vedi esempi 1-3 che seguono).

- Nel caso che siano stati indagati più caratteri la tabella si presenta in uno dei modi seguenti ed è detta **tabella di distribuzione unitaria multipla**

**Tabella di distribuzione unitaria multipla**

Unità/ mod. carat.	Mod. carat. 1	Mod. carat. 2	Mod. carat. 3
$u_1$	$m_1(u_1)$	$m_2(u_1)$	$m_3(u_1)$
$u_2$	$m_1(u_2)$	$m_2(u_2)$	$m_3(u_2)$
$u_3$	$m_1(u_3)$	$m_2(u_3)$	$m_3(u_3)$
$u_4$	$m_1(u_4)$	$m_2(u_4)$	$m_3(u_4)$
$u_5$	$m_1(u_5)$	$m_2(u_5)$	$m_3(u_5)$
...	...	...	...
$u_n$	$m_1(u_n)$	$m_2(u_n)$	$m_3(u_n)$

Unità/ mod. carat.	Mod. carat. 1	Mod. carat. 2	Mod. carat. 3
<b>1</b>	$m_1(1)$	$m_2(1)$	$m_3(1)$
<b>2</b>	$m_1(2)$	$m_2(2)$	$m_3(2)$
<b>3</b>	$m_1(3)$	$m_2(3)$	$m_3(3)$
<b>4</b>	$m_1(4)$	$m_2(4)$	$m_3(4)$
<b>5</b>	$m_1(5)$	$m_2(5)$	$m_3(5)$
...	...	...	...
<b>n</b>	$m_1(n)$	$m_2(n)$	$m_3(n)$

Nella fascia verticale a sinistra sono elencate le unità della popolazione o del campione; nella fascia orizzontale in alto i caratteri. Nell'incrocio di una riga associata a un'unità e di una colonna corrispondente a un carattere è indicata la modalità con cui si presenta il carattere nell'unità considerata.

La tabella distribuisce unità per unità le modalità dei caratteri rilevati e rappresenta la funzione che associa ad ogni unità statistica l'ennupla delle modalità dei caratteri su di essa rilevate; tale funzione è detta **distribuzione unitaria multipla**:

$$d : u_i \in \Omega \rightarrow (m_1(u_i), m_2(u_i), \dots, m_n(u_i)) \quad \text{distribuzione unitaria multipla}$$

## Esempi di distribuzioni unitarie semplici (tabelle grezze di dati)

**1. Indagine sui chilometri percorsi in aereo in un anno.** L'indagine è svolta su un collettivo di 20 persone (unità statistiche), che, intervistate, hanno dato le seguenti risposte:

**0 1800 0 0 0 480 0 0 1250 0 0 0 0 0 6600 0 0 9300 0 0**

l'allineamento indica la distribuzione secondo il modello (1): poiché al secondo posto c'è 1800, vuol dire che la seconda unità statistica o individuo ha compiuto 1800 chilometri di viaggio aereo. Il valore 0 si presenta più volte. I dati sono rappresentati più compiutamente nella seguente Tabella **unità/valore**

### Distribuzione Unitaria Semplice:

Unità	valore $x_i = \text{chilometri in aereo}$
1	$x_1 = 0$
2	$x_2 = 1800$
3	$x_3 = 0$
4	$x_4 = 0$
5	$x_5 = 0$
6	$x_6 = 480$
7	$x_7 = 0$
8	$x_8 = 0$
9	$x_9 = 1250$
10	$x_{10} = 0$
11	$x_{11} = 0$
12	$x_{12} = 0$
13	$x_{13} = 0$
14	$x_{14} = 0$
15	$x_{15} = 6600$
16	$x_{16} = 0$
17	$x_{17} = 0$
18	$x_{18} = 9300$
19	$x_{19} = 0$
20	$x_{20} = 0$

Il **collettivo**  $\Omega$  è costituito dalle 20 persone intervistate e indicate nella prima colonna con i numeri interi da 1 a 20: ognuna **persona** è quindi una **unità statistica**;

il **carattere** **“Km percorsi in aereo in un anno”** è quantitativo e quindi è una **variabile statistica**;

i **valori distinti** rilevati sono: **0, 480, 1250 1880, 6600, 9300**;

lo **0** è stato rilevato più volte

Utilizzando i valori distinti **0, 480, 1250 1880, 6600, 9300**, il **collettivo**  $\Omega$  è ripartito nelle classi

$\Omega_1 = \{1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 16, 17, 19, 20\} = C(0)$  classe delle unità che hanno percorso 0 Km

$\Omega_2 = \{6\} = C(480)$  classe delle unità che hanno percorso 480 Km

$\Omega_3 = \{9\} = C(1250)$  classe delle unità che hanno percorso 1250 Km

$\Omega_4 = \{2\} = C(1880)$  classe delle unità che hanno percorso 1880 Km

$\Omega_5 = \{15\} = C(6600)$  classe delle unità che hanno percorso 6600 Km

$\Omega_6 = \{18\} = C(9300)$  classe delle unità che hanno percorso 9300 Km

$\Pi = \{\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_6\}$  è la classificazione del collettivo secondo i chilometri percorsi.

**2. Indagine sul costo in euro per i chilometri percorsi in aereo in un anno.** La seguente tabella riporta i dati dell'indagine svolta sullo stesso collettivo di venti persone di cui all'esempio precedente: ad ogni persona è chiesto di dire quanto ha speso per viaggi aerei nell'anno considerato

**Distribuzione Unitaria Semplice:** tabella unità/valori

UNITÀ	$q_i = \text{costo viaggio in aereo in un anno}$
1	0
2	140
3	0
4	0
5	0
6	60
7	0
8	0
9	100
10	0
11	0
12	0
13	0
14	0
15	300
16	0
17	0
18	500
19	0
20	0

Il **collettivo** è costituito dalle 20 persone intervistate e indicate nella prima colonna con i numeri interi da 1 a 20 : ogni **persona** è quindi una **unità statistica**;

il **carattere** "**costo per viaggi aereo in un anno**" è quantitativo e quindi è una **variabile statistica**;

i **costi** rilevati sono:

**0, 60, 100, 140, 300, 500,**

e lo **0** è stato rilevato più volte

**3. Indagine su gli esiti di dieci lanci di una moneta testa croce.** Supponiamo che in una moneta testa croce sia stata lanciata in aria 10 volte e che gli esiti dei lanci siano:

T, T, C, C, C, T, C, T, T, C

**Distribuzione Unitaria Semplice:** tabella unità/modalità

Lancio	Esito
1	T
2	T
3	C
4	C
5	C
6	T
7	C
8	T
9	T
10	C

Il **collettivo** è l'insieme dei lanci

Le **unità statistiche o casi** (osservati) o **prove** sono i **lanci**

Il **carattere** è "**esito del lancio**"

con **modalità**: **T** (testa) e **C** (croce);

quindi il carattere è

una **mutabile sconnessa**

## Esempi. di distribuzioni unitarie multiple

**4. Indagine sui chilometri percorsi in aereo in un anno e i relativi costi.** Nella seguente tabella sono riuniti i dati raccolti su un collettivo di 20 individui rispetto ai caratteri indicati negli esempi 1 e 2

### Distribuzione unitaria doppia

UNITÀ	$x_i$ =kilometri in aereo	$q_i$ = costo in euro
1	0	0
2	1800	140
3	0	0
4	0	0
5	0	0
6	480	60
7	0	0
8	0	0
9	1250	100
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	6600	300
16	0	0
17	0	0
18	9300	500
19	0	0
20	0	0

**20 = ampiezza del collettivo**

### 5. Indagine sull'anno di laurea e facoltà di provenienza degli iscritti a un corso di perfezionamento

Nella seguente tabella sono riportati i risultati di un indagine effettuata sugli iscritti ad un corso di perfezionamento, tenutosi presso la facoltà di Architettura qualche anno fa, relativamente ai caratteri *anno di laurea* e *provenienza*. I nomi dei partecipanti al corso sono stati modificati per la privacy; accanto ad ogni nome, nella seconda e terza colonna, le modalità dei due caratteri rilevati nel partecipante che ha quel nome.

### Distribuzione unitaria doppia

<b>UNITÀ STATISTICHE (INDIVIDUI INDICATI CON NOME E COGNOME)</b>	<b>ANNO DI LAUREA</b>	<b>PROVENIENZA</b>
<b>RUSSO GIUSEPPINA</b>	<b>1998</b>	<b>INGEGNERIA CIVILE</b>
<b>BIANCHI ANNA</b>	<b>2001</b>	<b>MATEMATICA</b>
<b>RUSSO MICHELE</b>	<b>2001</b>	<b>ARCHITETTURA</b>
<b>PESCE SALVATORE</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>DE ROSSI MARIA</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>SEPE ASSUNTA</b>	<b>2001</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO UMBERTO</b>	<b>1984</b>	<b>ARCHITETTURA</b>
<b>CIMMINO ROBERTO</b>	<b>1993</b>	<b>ECONOMIA E COMM.</b>
<b>RUSSO MARIA</b>	<b>1996</b>	<b>SCIENZE GEOLOGICHE</b>
<b>IACCARINO PASQUALE</b>	<b>2001</b>	<b>INGEGNERIA EDILE</b>
<b>ESPOSITO FABIO</b>	<b>1999</b>	<b>ARCHITETTURA</b>
<b>BRUNO ROSA</b>	<b>2002</b>	<b>ARCHITETTURA</b>
<b>MORRA GABRIELLA</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO RAFFAELLA</b>	<b>2001</b>	<b>ARCHITETTURA</b>
<b>SASSO SANDRA</b>	<b>2003</b>	<b>ARCHITETTURA</b>
<b>ESPOSITO MARIANNA</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>CARDONE VINCENZO</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>DE MARCO FIORELLA</b>	<b>2004</b>	<b>INGEGNERIA AMBIENTALE</b>
<b>DE ROSSI ROCCO</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>BIANCHI MAURIZIO</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>IACCARINO DIANA</b>	<b>1996</b>	<b>SCIENZE POLITICHE</b>
<b>SQUILLO GABRIELLA</b>	<b>2002</b>	<b>SCIENZE STATISTICHE</b>
<b>CARDONE ROSA</b>	<b>2003</b>	<b>ARCHITETTURA</b>
<b>MORRA MANUELA</b>	<b>2003</b>	<b>ECONOMIA E COMM.</b>
<b>SQUILLO MARIA</b>	<b>2004</b>	<b>ARCHITETTURA</b>
<b>BRUNO CIRO</b>	<b>2004</b>	<b>SCIENZE GEOLOGICHE</b>
<b>DI LELLO MARIO</b>	<b>1996</b>	<b>ARCHITETTURA</b>
<b>GATTI ILARIA</b>	<b>2003</b>	<b>ARCHITETTURA</b>
<b>LEONE DANILA</b>	<b>2001</b>	<b>ARCHITETTURA</b>
<b>RICCIO CARMINE</b>	<b>1999</b>	<b>ARCHITETTURA</b>

Poiché in un'indagine non ha importanza conoscere l'identità degli individui costituenti il collettivo, ma solo le modalità rilevate, le unità statistiche sono usualmente indicate con numeri ( 1 per prima unità, 2 per seconda unità etc.): l'ultimo numero ci dà l'ampiezza del collettivo.

### Distribuzione unitaria doppia

<b>UNITÀ STATISTICHE (INDIVIDUI INDICATI CON CODICE NUMERICO)</b>	<b>ANNO DI LAUREA</b>	<b>PROVENIENZA</b>
1	1998	INGEGNERIA CIVILE
2	2001	MATEMATICA
3	2001	ARCHITETTURA
4	2004	ARCHITETTURA
5	2004	ARCHITETTURA
6	2001	ARCHITETTURA
7	1984	ARCHITETTURA
8	1993	ECONOMIA E COMM.
9	1996	SCIENZE GEOLOGICHE
10	2001	INGEGNERIA EDILE
11	1999	ARCHITETTURA
12	2002	ARCHITETTURA
13	2004	ARCHITETTURA
14	2001	ARCHITETTURA
15	2003	ARCHITETTURA
16	2004	ARCHITETTURA
17	2004	ARCHITETTURA
18	2004	INGEGNERIA AMBIENTALE
19	2004	ARCHITETTURA
20	2004	ARCHITETTURA
21	1996	SCIENZE POLITICHE
22	2002	SCIENZE STATISTICHE
23	2003	ARCHITETTURA
24	2003	ECONOMIA E COMM.
25	2004	ARCHITETTURA
26	2004	SCIENZE GEOLOGICHE
27	1996	ARCHITETTURA
28	2003	ARCHITETTURA
29	2001	ARCHITETTURA
30	1999	ARCHITETTURA

30

Le unità statistiche sono 30, quindi l'indagine è stata condotta su un collettivo di 30 persone

## 6. Indagine su un collettivo di 20 famiglie relativamente a i caratteri

*numero componenti della famiglia, reddito annuo, titolo di studio del capofamiglia, zona di residenza*

### Distribuzione unitaria multipla

FAMIGLIA	NUMERO DI COMPONENTI	REDDITO ANNUO (in migliaia di)	TITOLO DI STUDIO DEL	ZONA DI RESIDENZA
1	2	28	ELEMENTARE	NORD
2	1	35	MEDIA	CENTRO
3	3	50	INFERIORE	NORD
4	1	45	MEDIA	NORD
5	1	40	INFERIORE	SUD (ISOLE)
6	2	30	MEDIA	SUD (ISOLE)
7	3	55	SUPERIORE	CENTRO
8	4	80	LAUREA	CENTRO SUD
9	5	60	MEDIA	(ISOLE)
10	6	85	INFERIORE	NORD
11	7	90	MEDIA	NORD
12	1	52	INFERIORE	CENTRO SUD
13	2	62	MEDIA	(ISOLE) SUD
14	3	75	SUPERIORE	(ISOLE)
15	5	60	LAUREA	NORD
16	4	45	LAUREA	NORD
17	3	42	LAUREA	CENTRO
18	2	28	MEDIA	NORD
19	8	70	SUPERIORE	SUD (ISOLE)
20	2	38	MEDIA	SUD (ISOLE)

Nella tabella le unità statistiche (famiglie) sono indicate attraverso numeri: i numeri sono 20 quindi 20 è l'ampiezza del campione indagato.

## 8.1 Codici numerici per indicare le modalità di un carattere qualitativo

Per facilitare l'inserimento dei dati in un elaboratore, le modalità di un carattere qualitativo possono essere rappresentate attraverso codici numerici; se il carattere è una mutabile ordinale i codici devono indicare i livelli della gerarchia lungo la quale si dispongono le modalità.

Con riferimento all'esempio 6 possiamo attribuire dei codici ai titoli di studio e alle zone di residenza, come di seguito indicato:

### DISTRIBUZIONE UNITARIA MULTIPLA

FAMIGLIA	NUMERO DI COMPONENTI	REDDITO ANNUO (in migliaia di euro)	TITOLO DI STUDIO DEL CAPOFAMIGLIA	ZONA DI RESIDENZA
1	2	28	1	1
2	1	35	2	2
3	3	50	2	1
4	1	45	3	1
5	1	40	4	3
6	2	30	2	3
7	3	55	2	2
8	4	80	3	2
9	5	60	4	3
10	6	85	4	1
11	7	90	4	1
12	1	52	3	2
13	2	62	3	3
14	3	75	3	3
15	5	60	1	1
16	4	45	2	1
17	3	42	2	2
18	2	28	1	1
19	8	70	3	3
20	2	38	4	3

#### Titolo di studio del capofamiglia (mutabile ordinata):

Nessun titolo = 0

Elementare = 1

Media infer. = 2

Media super. = 3

Laurea = 4

Specializzazione = 5

Poiché la mutabile è ordinata i codici devono riflettere la gerarchia dei titoli

**Residenza (mutabile sconnessa):** Nord = 1      Centro=2      Sud (isole) = 3

Poiché la mutabile residenza è sconnessa i codici possono essere anche attribuiti in modo diverso

Una tabella di dati grezzi è dispersiva e non leggibile; occorre allora costruire una tabella più sintetica (tabella elaborata) che ci dia in modo chiaro le informazioni *quantitative* che ci interessano. L'elaborazione dei dati e il passaggio dalle tabelle di distribuzione unitaria a tabelle elaborate si basa sulla classificazione delle unità statistiche secondo le modalità (cfr. paragrafo 6) e costituisce l'argomento del prossimo fascicolo.

## APPENDICE

### CAMPIONAMENTO

**Problema della scelta del campione.** E' sovente impossibile o non conveniente svolgere l'indagine su tutte le unità di una popolazione e si limita l'indagine alle unità di un campione. Si pone allora il problema di scegliere un campione che sia **rappresentativo** della popolazione, che rifletta cioè in qualche modo la diversità degli individui che compongono la popolazione, i rapporti numerici tra i vari gruppi omogenei etc., in modo da permettere di estendere alla popolazione i risultati dell'indagine parziale con piccolo margine di errore. Elementi che concorrono alla rappresentatività sono: **l'ampiezza** del campione e il **modo** in cui vengono scelte le unità campionarie. Fattori rilevanti per la determinazione dell'ampiezza del campione e la scelta delle unità campionarie sono: *il numero delle unità costituenti l'universo, il grado di omogeneità dell'universo, il ventaglio delle risposte possibili, il margine di errore* al di sotto del quale si vuole stare. Una volta che è stato determinato il numero delle unità che devono costituire il campione queste possono essere scelte seguendo più schemi.

#### *Schemi di campionamento*

**Campionamento casuale semplice:** le unità che devono costituire il campione vengono estratte a caso dall'universo in modo casuale. Si distinguono due tipi di estrazioni: estrazione con ripetizione o bernoulliana (il nome o numero dell'unità estratta viene rimesso tra i nomi o numeri delle unità dell'universo prima di procedere ad una nuova estrazione) ed estrazione senza ripetizione.



**Campionamento casuale sistematico:** posto  $k = \frac{\text{Numero elementi universo}}{\text{ampiezza del campione}}$  viene scelto un elemento ogni  $k$  elementi da una lista ordinata causale degli elementi dell'universo; la prima scelta è effettuata a partire da un numero casuale  $h \leq k$  ( $h = \text{numero di partenza}$ ).

Esempio

1	9
2	10
3	11
4	12
5	13
6	14
7	15
8.....	

**Campionamento a grappoli:** l'universo è diviso in sottinsiemi detti *grappoli*, assimilabili a campioni. Si estraggono poi alcuni grappoli e si prosegue o a formare il campione con tutte le unità dei grappoli estratti (campionamento a grappolo ad uno stadio), *oppure* si procede suddividendo l'insieme delle unità ottenute ancora in grappoli e tra questi ne vengono estratti alcuni le cui unità formeranno il campione (campionamento a grappolo a due stadi); *oppure* si procede ad una ulteriore divisione in grappoli delle unità prescelte e si procede come sopra (campionamento a grappolo a più stadi)

**Campionamento stratificato (o proporzionale o a quota):** si divide l'universo in parti dette *strati* costituite da unità abbastanza omogenee (mentre le unità di uno strato non sono omogenee a quelle di un altro) poi si estraggono campioni casuali semplici da ogni strato che si uniscono a formare il campione finale. Il campione così ottenuto potrà essere più piccolo di quello casuale scelto direttamente dall'universo perché le unità derivano da strati più omogenei.

Si utilizza il campionamento a strati quando sono individuabili parti della popolazione (strati) in cui si sa che il fenomeno indagato si manifesta in modo diverso e si vuole che tutte queste parti siano rappresentate .

Esempio da G. Arangio Ruiz “ Che cos'è la Statistica” , Editori Riuniti

**Indagine sulla spesa degli abitanti di una città per viaggi, vacanze, week-end**

<b>Strati</b> o settori (dal punto di vista economico)	<b>Universo</b> (famiglie che abitano nella città)	<b>Campione</b> (famiglie che verranno intervistate)
<b>Molto ricche</b>	<b>10.000</b>	<b>20</b>
<b>Ricche</b>	<b>50.000</b>	<b>100</b>
<b>Medie</b>	<b>250.000</b>	<b>500</b>
<b>Abbastanza povere</b>	<b>100.000</b>	<b>200</b>
<b>Povere</b>	<b>50.000</b>	<b>100</b>
<b>Totale famiglie</b>	<b>460.000</b>	<b>920</b>

Porzioni di popolazione scelte dal ogni strato in funzione della numerosità dello strato:

$$\frac{20}{10.000} = \frac{1}{500} \qquad \frac{1}{500} 10.000 = 20$$

$$\frac{100}{50.000} = \frac{1}{500}$$

$$\frac{500}{250.000} = \frac{1}{500}$$

$$\frac{200}{100.000} = \frac{1}{500}$$

$$\frac{100}{50.000} = \frac{1}{500}$$