UNIVERSITA' DEGLI STUDI DI PERUGIA
DIPARTIMENTO DI FILOSOFIA SCIENZE SOCIALI UMANE E DELLA FORMAZIONE
Corso di Laurea in Scienze per l'Investigazione e la Sicurezza

1. Metodologia Classificazione dei dati e Distribuzioni statistiche

Prof. Maurizio Pertichetti

Statistica sociale

1. Metodologia e Classificazione dei dati

La **Statistica**, sebbene darne una definizione non sia semplice, è propriamente l'applicazione dei metodi scientifici alla programmazione della raccolta dei dati, alla loro classificazione, elaborazione, analisi, sintesi e presentazione ed alla inferenza di conclusioni attendibili da essi. In pratica tende alla individuazione dei metodi logico-matematici più idonei per la scoperta di leggi e relazioni tra fenomeni e ciò a fini descrittivi, esplicativi, previsionali, ecc. Ha come fine lo studio quantitativo e qualitativo di fenomeni collettivi (demografici, economici, sociali, sanitari.....), cioè di situazioni singole ripetibili con caratteristiche comuni, riscontrando le regolarità che sono alla base di tale fenomeno collettivo.

Nell'ambito della metodologia statistica si distinguono due filoni fondamentali:

la **Statistica descritiva** che attiene alla descrizione, attraverso strumenti matematici, di un fenomeno reale conducendo lo studio sulla **popolazione (o universo)** sulla quale si manifesta il fenomeno stesso oggetto di esame. Essa studia i criteri di rilevazione, di classificazione e di sintesi delle informazioni relative agli aggregati di riferimento. Raccoglie tali informazioni in distribuzioni, semplici o multiple, le rappresenta in forma grafica e realizza una sintesi, attraverso indici dei risultati ottenuti;

la **Statistica inferenziale** che attiene all'induzione probabilistica della struttura sconosciuta di una popolazione. Si propone di risolvere il cosiddetto problema inverso, ossia, sulla base di osservazioni condotte su un **campione** di unità selezionate con date procedure dalla popolazione, perviene a conclusioni valide, entro dati livelli di probabilità di errore, per l'intera stessa popolazione. La teoria della probabilità e il campionamento statistico, che utilizza la teoria dei campioni, sono alla base della statistica inferenziale.

Si premette che questo corso limita il proprio interesse di studio al filone della statistica descrittiva.

Definizioni fondamentali

La statistica come già accennato acquisisce le informazioni dalla **popolazione** o **collettivo** o da un **campione**.

- una **popolazione** finita si definisce **empirica** quando l'insieme delle unità che vi appartengono costituisce concretamente l'oggetto della ricerca, mentre è detta **teorica** se alcune delle unità non hanno possibilità di essere effettivamente osservate. E' in sostanza l'**aggregato** che forma l'oggetto dello studio statistico. Costituiscono popolazioni statistiche: l'insieme dei residenti di una regione, l'insieme degli studenti di un Ateneo, le imprese di un dato settore localizzate in una provincia, il numero di auto circolanti in un dato territorio, il numero di immigrati nel nostro Paese, ecc. Deve sottolinearsi che le popolazioni studiate nella ricerca sociale devono sempre essere definibili nel tempo e nello spazio.

- il **campione** è invece un sottoinsieme di una popolazione ottenuto per estrazione tramite tecniche definite dalla teoria dei campioni. L'impossibilità di osservare tutte le unità statistiche, il costo elevato della rilevazione, i limiti di tempo fanno si che spesso lo studio si limiti ad un campione rappresentativo dell'intera popolazione, ossia ad un limitato numero di unità che riproduca le caratteristiche dell'intera popolazione.

Le **unità statistiche** sono le componenti elementari dell'**aggregato**, **popolazione** o **campione**, sulle quali vengono compiute le osservazioni e rilevate le variabili oggetto delle analisi statistiche. Si distinguono in:

- **unità semplici**, come una singola persona, una singola abitazione, una singola impresa, un singolo autoveicolo, ecc;
- **unità composte**, se sono insiemi di unità semplici simili considerate anche a prescindere dall'unità composta: per esempio una famiglia intesa come insieme di persone, un edificio inteso come insieme di abitazioni, ecc.;

Con il termine **carattere** si indica invece una entità logica, adatta a descrivere una popolazione o un campione. Tale entità può essere di **natura qualitativa** (colore occhi, sesso, titolo di studio di una persona, marca delle autovetture circolanti, ecc), in tal caso si definisce anche **mutabile**, o di **natura quantitativa** (peso, altezza, età di una persona, cilindrata delle autovetture circolanti, ecc), e si definisce **variabile**. La statistica studia come si distribuisce una certa popolazione in relazione ad uno o più caratteri osservati sulle unità statistiche.
Infine per **modalità** si intende l'insieme dei diversi modi di presentarsi di un carattere statistico. In sostanza una popolazione statistica si può classificare in vari modi, in relazione alle diverse qualità o valori posseduti da ciascuno dei singoli elementi.

I caratteri che formano l'oggetto di una rilevazione statistica, che si distinguono, come detto, in:

- **qualitativi** e si esprimono mediante **attributi**, ossia attraverso l'insieme **delle diverse qualità del carattere**: titolo di studio, colore degli occhi, tipo di impiego, ecc; si distinguono ulteriormente in:

  Sconnessi o non ordinati quando le modalità mancano di qualsiasi ordine di tipo logico o naturale: un elenco di professioni, la nazionalità, il sesso ...

  Ordinati o ordinali quando le modalità possiedono un ordine naturale: il titolo di studio, le attitudini ad una disciplina (ottime, buone, sufficienti, insufficienti, scarse), il tempo in giorni, mesi e anni.

- **quantitativi** e si esprimono mediante **numeri**, ossia attraverso l'insieme **dei valori del carattere**: altezza, peso, abitanti; si distinguono ulteriormente in:

  Discreti (quando si conta) quando le modalità sono espresse da valori appartenenti all'insieme dei numeri naturali 1,2,3,... , ad esempio abitanti di un comune, numero di figli, ecc.

  Continui (quando si misura) quando le modalità possono assumere tutti i numeri reali appartenenti ad un dato intervallo, ad esempio l'altezza, il peso, le temperature, ecc.

Distribuzioni statistiche

Il punto di partenza di ogni indagine statistica è quello della rilevazione dei dati, ovvero la raccolta dei dati statistici individuali che compongono il fenomeno collettivo oggetto dell'indagine, e della loro classificazione. La raccolta dei dati ha come presupposto la definizione di uno o più caratteri e di come essi si articolano nelle corrispondenti modalità e altresì dell'insieme delle unità statistiche oggetto dell'indagine. Definiti questi elementi si procede **al conteggio o alla misurazione del carattere su ciascuna unità statistica**.

Il conteggio o la misurazione, che altro non sono che l'insieme delle osservazioni, possono essere visti come l'assegnazione di una modalità a ciascuna unità statistica .
A tale operazione segue l'associazione a ciascuna modalità del numero delle ricorrenze, ossia far corrispondere a ogni modalità del carattere, il numero di unità statistiche (o osservazioni) che rientrano in tale modalità. Questo numero è definito, in statistica, **frequenza assoluta** o valore assoluto della modalità.

L'insieme delle modalità e delle rispettive frequenze assolute prende il nome di **distribuzione statistica** o **distribuzione di frequenze assolute**, che è un modo di organizzare i dati rilevati in una tabella.

Una distribuzione statistica può essere rappresentata in tre modi distinti:

- elencando **in forma tabellare** tutte le modalità e le corrispondenti frequenze (tabella statistica);
- **in forma grafica**;
- **in forma analitica** (ossia tramite espressione matematica);

nella statistica descrittiva si utilizzano solamente le prime due tipologie di rappresentazione.

Individui secondo il colore degli occhi

Modalità / Mutabile	Frequenze assolute
Verdi	12
Azzurri	9
Castani	53
Neri	26
Totale	**100**

Famiglie secondo il numero dei componenti

Modalità / Variabile	Frequenze assolute
1	39
2	31
3	23
4 e oltre	7
Totale	**100**

Il **numero associato** a ciascuna delle modalità con le quali si presenta un certo carattere non sempre tuttavia esprime quante volte si manifesta quella modalità nel collettivo di riferimento, ovvero non sempre si identifica come la **frequenza della modalità** così come si è visto, ma a volte esprime l'ammontare globale del carattere per ogni modalità ed allora in tal caso si chiama **intensità** o **quantità della modalità**. Pertanto oltre alle tabelle che prendono il nome di distribuzione di frequenze, vi sono anche quelle dette **distribuzione di quantità** o **di intensità** che indicano la suddivisione del totale tra le diverse voci.

Ad esempio le principali importazioni dell'Italia nel 2016 secondo alcuni settori, costituiscono una distribuzione di intensità:

Principali importazioni Italia - anno 2016	**mld $ cif**
Macchinari e mezzi di trasporto	107,7
Combustibili minerali e lubrif.	93,9
Prodotti chimici e derivati	73,4
Alimenti, bevande e tabacco	44
Totale incluso altro	**479,3**

Il carattere qualitativo del fenomeno è importazioni, le sue modalità sono Macchinari e mezzi di trasporto, Combustibili minerali e lubrif., ecc. I valori numerici invece sono costituiti dalla misura del valore in miliardi di dollari delle importazioni.

In presenza di un numero elevato di modalità (+ di 20) è conveniente definire **classi di valori**. Generalmente la necessità di segmentare i dati in classi si presenta per i caratteri continui e il procedimento, in questo caso, prende il nome di **aggregazione in classi** o di **discretizzazione**.

Come esempio, supponiamo di analizzare i dati sull'altezza espressa in centimetri di 1.000 individui: è evidente che è poco interessante sapere quanti sono coloro alti esattamente 172,5 o 172,6 o 172,7 e così via, mentre è certamente più significativo conoscere quanti individui hanno un'altezza compresa fra 170 e 174 cm, quanti fra 175 e 179 e via dicendo.

Per la costruzione di una scala numerica per classi, si devono rispettare alcune regole fondamentali:

- il numero delle classi dovrebbe essere compreso tra 10 e 20 e comunque non superiore a 1/10 del numero totale delle osservazioni;
- soprattutto, le classi devono essere stabilite in modo da evitare ambiguità nella assegnazione delle osservazioni, ovvero va evitata la sovrapposizione delle classi. Chiaramente il dubbio si pone per i valori limite, per cui occorre specificare se essi vadano inclusi nella classe inferiore o in quella superiore.

Per stabilire in modo univoco dove collocare gli estremi delle classi, viene utilizzata la seguente simbologia:

- ⊢— chiusa a sinistra e aperta a destra, il limite sup è escluso dalla classe
- —⊣ aperta a sinistra e chiusa a destra, il limite inf è escluso dalla classe
- ⊢⊣ chiusa sia a sinistra che a destra, entrambi i limiti inf e sup sono inclusi nella classe

Si definiscono inoltre: **Modulo =** $x_{i+1} - x_i$ **Valore centrale =** $\frac{x_i + x_{i+1}}{2}$

L'ampiezza (**α**) del modulo può essere uguale o diversa per tutte le classi.

Altezza in cm di 1000 studenti

Modalità/ Classi di altezza	Valore centrale	Frequen assolute studenti
165 ⊢⊣ 169	167,0	x_1
170 ⊢⊣ 174	172,0	x_2
175 ⊢⊣ 179	177,0	x_3
180 ⊢⊣ 184	182,0	x_4
185 ⊢⊣ 189	187,0	x_5
Totale		**1.000**

Altezza in cm di 1000 studenti

Modalità/ Classi di altezza	Valore centrale	Frequen assolute studenti
165 ⊢— 170	167,5	x_1
170 ⊢— 175	172,5	x_2
175 ⊢— 180	177,5	x_3
180 ⊢— 185	182,5	x_4
185 ⊢— 190	187,5	x_5
Totale		**1.000**

Prima di entrare nel dettaglio della rappresentazione di una distribuzione statistica è però opportuno specificare alcune convenzioni sulle notazioni che in genere si utilizzano.

Il **carattere** viene normalmente indicato con una lettera maiuscola tipicamente X o Y.

Le **modalità** si indicano con la stesa lettera del carattere cui afferiscono, scritta in minuscolo e per caratteri qualitativi o numerici discreti dotata di pedice:

$$x_1 \quad x_2 \quad \dots \quad x_i \quad x_{i+1} \quad \dots \quad x_k$$

in questo caso **k**, o altra lettera, indica il numero totale delle modalità.

Con **n** si indica il totale delle osservazioni e con $\mathbf{n_i}$ la **frequenza assoluta**, ossia il numero di ricorrenze associate ad $\mathbf{x_i}$, modalità i-esima di **X**. Di conseguenza la somma delle frequenze sarà:

$\sum_{i=1}^{k} n_i = n$ ovvero uguale al totale delle osservazioni.

La **frequenza relativa** o **proporzione** è $f_i = \frac{n_i}{n}$ ed è data dal rapporto tra la frequenza assoluta associata ad una modalità e il numero totale di casi della distribuzione. Come si vedrà più avanti si tratta di un rapporto di composizione detto anche di una parte al tutto.

Modalità di X	Frequen assolute	Frequen relative
x_1	n_1	f_1
x_2	n_2	f_2
…	…	…
x_i	n_i	f_i
…	…	…
x_k	n_k	f_k
Totale	**n**	**1**

Modalità di X	Frequen assolute	Frequenze relative	
1	26	26/500 =	0,052
2	92	92/500 =	0,184
3	186	186/500 =	0,372
4	135	135/500 =	0,270
5	45	45/500 =	0,090
6 o +	16	16/500 =	0,032
Totale	**500**		**1,000**

La somma delle frequenze relative è sempre $\sum_{i=1}^{k} f_i = 1$

Ulteriormente è necessario introdurre il concetto di **frequenza cumulata**, che indica quante sono le unità che possiedono le prime **i** modalità del carattere ed è indicata con $\mathbf{n'_i}$

Freq ass	Frequenze cumulate assolute
n_1	$n_1' = n_1$
n_2	$n_2' = n_1 + n_2$
...	...
n_i	$ni' = n_1 + n_2 + \dots + n_i$
...	...
n_k	$nk' = n_1 + n_2 + \dots + n_i + \dots + n_k = n$

Analogamente, le **frequenze cumulate relative**, indicate con $\mathbf{f'_i}$ esprimono la frazione di soggetti che possiedono le prime **i** modalità del carattere.

Freq rel	Frequenze cumulate relative
f_1	$f_1' = f_1$
f_2	$f_2' = f_1 + f_2$
...	...
f_i	$fi' = f_1 + f_2 + \dots + f_i$
...	...
f_k	$fk' = f_1 + f_2 + \dots + f_i + \dots + f_k = 1$

Modalità	Freq ass	Freq rel	Freq cumulate assolute		Freq cumulate relative	
1	26	0,052	26 =	26	0,052 =	0,052
2	92	0,184	26+92 =	118	0,052+0,184 =	0,236
3	186	0,372	26+92+186 =	304	0,052+0,184+0,372 =	0,608
4	135	0,270	26+92+186+135 =	439	0,052+0,184+0,372+0,270 =	0,878
5	45	0,090	26+92+186+135+45 =	484	0,052+0,184+0,372+0,270+0,090 =	0,968
6 o +	16	0,032	26+92+186+135+45+16 =	**500**	0,052+0,184+0,372+0,270+0,090+0,032 =	**1,000**
Totale	**500**	**1,000**	----		----	

Il **valore massimo** che una distribuzione di frequenze assolute cumulate può assumere **è sempre uguale al totale delle frequenze assolute**, mentre Il **valore massimo** che una distribuzione di frequenze relative cumulate può assumere **è sempre uguale ad 1**.
Il calcolo delle frequenze cumulate ha significato solo per i caratteri quantitativi e qualitativi ordinabili.

Ai fini della misurazione vengono introdotte le seguenti scale di misurazione:

Scale di misura per **caratteri qualitativi:**

- **Scala non ordinale**
  E' il tipo di scala di misura più semplice (la relativa distribuzione prende spesso il nome di mutabile statistica sconnessa).

Mezzo di trasporto utilizzato da 100 studenti per recarsi alla sede dell'università

Modalità / Mezzo di trasporto	Frequenze / Nr di studenti
Auto privata	4
Moto/scooter	23
Bicicletta	6
Mezzi pubblici	64
Altro	3
Totale	100

- **Scala ordinale**
  Si tratta di attributi per i quali è possibile introdurre una relazione di ordine, ossia una forma di ordinamento.

Giudizi dati su 40 studenti

Modalità / Giudizio	**Frequenze / Nr di studenti**
Insufficiente	2
Sufficiente	6
Discreto	13
Buono	11
Ottimo	8
Totale	40

Scale di misura per **caratteri quantitativi:**

- **Scala numerica discreta**
  La misura è espressa da un numero intero che non derivi da un arrotondamento di un numero reale, ma dalla natura stessa del dato (per esempio, l'età in anni, gli abitanti di una città, le autovetture prodotte da un marchio automobilistico in un anno, ecc).

Studenti iscritti in un corso quinquennale

Modalità/Classe	1^a	2^a	3^a	4^a	5^a	**Totale**
Frequenze/Nr di studenti	27	25	24	22	21	**119**

- **Scala numerica continua**
  La misura in questo caso è espressa da un numero reale.

Peso di 10 individui

Modalità/Peso (in Kg)	66,3	72,2	73,5	77,4	78,8	**Totale**
Frequenze/Nr di soggetti	1	2	4	2	1	10

od anche

Modalità/Peso (in Kg)	65 ⊢ 69	70 ⊢ 74	75 ⊢ 79	**Totale**
Frequenze/Nr di soggetti	1	6	3	10

La **distribuzione di frequenza** è dunque una particolare tipologia di rappresentazione dei dati statistici per illustrare la quale è necessario costruire una **tabella statistica**. Le tabelle sinora costruite si riferiscono all'esame di singoli caratteri (**distribuzioni monodimensionali**), in realtà la statistica sempre più è utilizzata per studiare diversi caratteri contemporaneamente. Una distribuzione di frequenza, e quindi una tabella statistica, può essere dunque semplice o multipla a seconda che figurino le modalità riferite a un solo carattere o a più caratteri. In tale ultima circostanza si parla di analisi multivariata, e naturalmente il caso più semplice è rappresentato dall'analisi bivariata che prevede lo studio congiunto di due caratteri e da luogo a **distribuzioni bidimensionali**.

La distribuzione doppia è la distribuzione congiunta di due caratteri X e Y (per i quali si suppone l'esistenza di una relazione) esaminata rispetto al contemporaneo verificarsi su ciascuna unità statistica del collettivo di riferimento di una modalità x_i , del carattere X, e di una modalità y_i , del carattere Y.

La **tabella a doppia entrata** è quindi una tabella statistica in cui figurano le frequenze assolute o relative riferite alle diverse combinazioni di modalità o classi di modalità di due caratteri X e Y, desumibili da una distribuzione doppia.

Esempio di rappresentazione tabellare di due caratteri da studiare contemporaneamente, ovvero di tabella a doppia entrata:

Occupati secondo i settori e la posizione professionale

Settori	Posizione professionale		Totale
	Dipendenti	Autonomi	
Agricoltura	485	776	1.261
Industria	4.147	956	5.103
Altre attività	4.941	2.546	7.487
Totale	9.573	4.278	13.851

riga madre

riga o distribuzione marginale di Y

distribuzioni condizionate del carattere X o del carattere Y

colonna madre

colonna o distribuzione marginale di X

In termini concettuali si può allora così generalizzare: siano **X** e **Y** due caratteri cui corrispondono rispettivamente le seguenti modalità:

$\mathbf{x_1}$ $\mathbf{x_2}$ ... $\mathbf{x_i}$ $\mathbf{x_{i+1}}$ ... $\mathbf{x_r}$ $\mathbf{y_1}$ $\mathbf{y_2}$ ... $\mathbf{y_j}$ $\mathbf{y_{j+1}}$ ... $\mathbf{y_c}$

ed $\mathbf{n_{ij}}$ il numero delle osservazioni che identificano contemporaneamente la modalità $\mathbf{x_i}$ e la modalità $\mathbf{y_j}$ **.** La tabella a doppia entrata che ne consegue sarà:

	y_1	y_2	...	y_j	...	y_c	**Totale**
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
...	...	...	...	...	...	...	...
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r.}$
Totale	$n._1$	$n._2$	...	$n._j$	...	$n._c$	n

La prima riga della tabella è detta riga madre e vi figurano le modalità del carattere Y :

$\mathbf{y_1}$ $\mathbf{y_2}$ ... $\mathbf{y_j}$ $\mathbf{y_{j+1}}$ ... $\mathbf{y_c}$

La prima colonna della tabella è detta colonna madre e vi figurano le modalità del carattere X:

$\mathbf{x_1}$ $\mathbf{x_2}$ ... $\mathbf{x_i}$ $\mathbf{x_{i+1}}$ ... $\mathbf{x_r}$

Il corpo della tabella è una matrice **r** x **c** (con **r** righe e **c** colonne), che contiene le frequenze $\mathbf{n_{ij}}$ ovvero il numero di quegli elementi della popolazione **n** che possiedono le modalità xi di X e yj di Y simultaneamente. Dei pedici che caratterizzano $\mathbf{n_{ij}}$ (i=1,2,…,r; j=1,2,..,c) il primo rappresenta la riga e il secondo la colonna.

Nell'ultima riga, detta **riga marginale**, figurano le frequenze marginali, che rappresentano i totali di ciascuna delle **c** colonne, e precisamente la frequenza:

$n._j = \sum_{i=1}^{r} n_{ij}$ indica il numero di elementi totali che possiedono la modalità y_j del carattere Y.

Analogamente nell'ultima colonna, detta **colonna marginale**, figurano le frequenze marginali, che rappresentano i totali di ciascuna delle **r** righe, e precisamente la frequenza:

$n_{i.} = \sum_{J=1}^{c} n_{ij}$ indica il numero di elementi totali che possiedono la modalità x_i del carattere X.

Da ultimo, in una tabella a doppia entrata si individuano quelle che sono le **distribuzioni condizionate**, ovvero distribuzioni semplici che si ottengono associando la riga madre con **una qualsiasi delle r righe successive**, oppure associando la **colonna madre** con **una qualsiasi delle c colonne successive**.

La distribuzione $Y \mid (X=x_i)$ è la distribuzione condizionata del carattere Y dato il valore x_i del carattere X.

La distribuzione $X \mid (Y=y_j)$ è la distribuzione condizionata del carattere X dato il valore y_j del carattere Y.

Per cui da una tabella a doppia entrata si desumono:
- **r** distribuzioni condizionate del carattere Y alle corrispondenti modalità del carattere X;
- **c** distribuzioni condizionate del carattere x alle corrispondenti modalità del carattere Y.

Serie statistiche

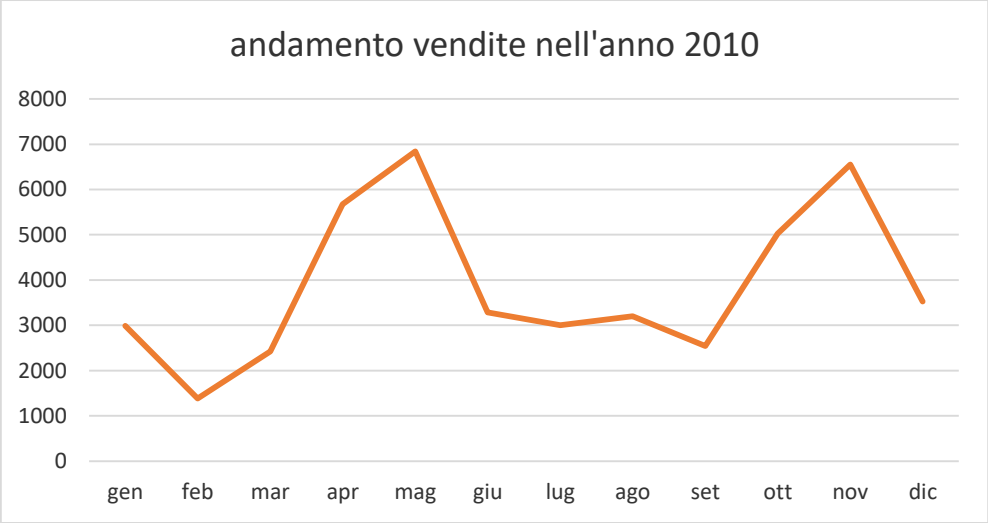
Molto spesso nell'attività di indagine statistica si pone l'esigenza di dover porre a confronto uno stesso fenomeno in contesti diversi, ovvero di porre in relazione tra loro due valori corrispondenti a modalità dello stesso carattere. Ad esempio, classico è il caso di dover valutare l'andamento del numero dei disoccupati, così come quello dei valori delle esportazioni o dei livelli di produzione, ecc. nel corso degli anni per avere informazioni sui relativi trend. In questo caso le distribuzioni statistiche possono essere trattate come **serie** statistiche.

La **serie statistica** si distingue tra:
- **serie temporale o serie storica**, quando la successione dei valori assunti dal carattere è ordinata logicamente secondo il tempo;
- **serie territoriale**, quando la successione dei valori assunti dal carattere nello stesso tempo è riferita a luoghi, territori, unità amministrative, ecc. diversi.

I valori che vengono messi a confronto possono essere di diversa natura: quantità varie, frequenze assolute, relative, percentuali, indici, coefficienti aventi riferimenti territoriali e temporali diversi.

Le tabelle che seguono sono un esempio di serie storiche, in quanto indicano valori distribuiti per mesi e anni. Ovvero il fenomeno è studiato in funzione del tempo.

anno 2010	vendite in €
gen	2.990
feb	1.382
mar	2.425
apr	5.673
mag	6.842
giu	3.285
lug	3.000
ago	3.200
set	2.540
ott	5.025
nov	6.552
dic	3.525

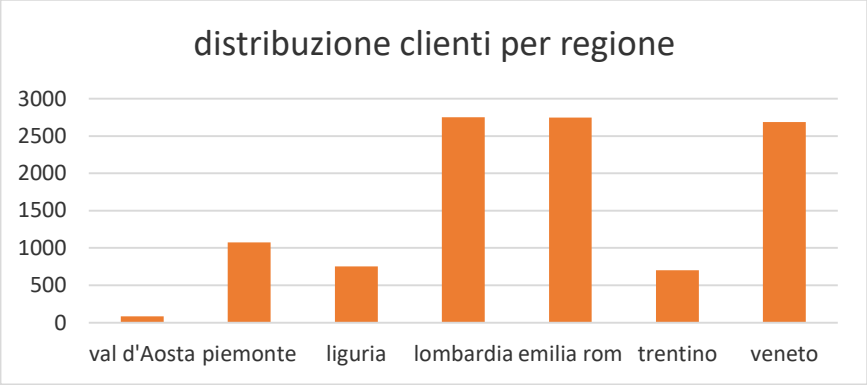


Retribuzione annua di una data categoria di lavoratori

anno	2004	2005	2006	2007	2008	2009
retribuz annua	17.166	17.853	18.818	19.552	19.884	20.242

Le tabelle sottostanti sono invece esempi di serie territoriali, in quanto i valori di un dato anno sono distribuiti in funzione dell'area geografica.

regione	clienti
val d'Aosta	83
piemonte	1.073
liguria	752
lombardia	2.752
emilia romagna	2.746
trentino	702
veneto	2.688



Distribuzione territoriale della natalità

area geografica	nord ovest	nord est	centro	sud	isole	Italia
natalità (*tasso di*)	9,36	9,69	9,36	10,20	9,75	9,67