



**Facciamo statistica! Conoscere la realtà che ci circonda  
attraverso i numeri**

**Materiale didattico**

*prof.ssa Nunziata Ribecco* [nunziata.ribecco@uniba.it](mailto:nunziata.ribecco@uniba.it)  
*prof.ssa Angela Maria D'Uggento* [angelamaria.duggento@uniba.it](mailto:angelamaria.duggento@uniba.it)

# Cos'è la Statistica (1/2)

La Statistica è uno strumento; serve a *tradurre l'informazione in conoscenza*.

Studia i **fenomeni collettivi** (osservazione di un **insieme** di manifestazioni individuali), allo scopo di:

- ✓ ottenere informazioni;
- ✓ descrivere un fenomeno;
- ✓ individuare relazioni.

## **A cosa serve la Statistica?**

Nella nostra società dell'informazione e della comunicazione, la Statistica e il ragionamento statistico sono ovunque:

- nei telegiornali compaiono ogni giorno grafici, tabelle, risultati di sondaggi;
- attraverso le carte fedeltà si registrano i dati sui nostri acquisti, abitudini, gusti;
- in farmacia acquistiamo prodotti testati e sperimentati;

.....

# Cos'è la Statistica (2/2)

La Statistica quindi

- aiuta a valutare ciò che osserviamo e a prendere decisioni sulla base delle osservazioni e dei dati rilevati;
- affronta ed interpreta aspetti del mondo reale: la casualità, l'aleatorietà e l'incertezza;
- è uno strumento fondamentale per le scienze sperimentali, naturali ed empiriche;
- non dice mai che un'affermazione è vera o falsa, ma che è vera o falsa con una determinata **probabilità**.

E la teoria della probabilità è il principale strumento che consente di *modellare l'incertezza*.

# Cosa fa uno Statistico?

La Statistica ha «pacificamente invaso» ogni campo dello scibile umano e, quindi, non esiste UN SOLO lavoro tipico dello statistico, spaziando da quello che può svolgere negli Uffici statistici della Pubblica Amministrazione (ISTAT, Regioni, Comuni, ecc) a quello in aziende e società di consulenza.

Uno statistico è un **data manager**, che applica metodi quantitativi, a seconda del campo di applicazione in cui è inserito.

Già nel 2009 Hal Varian, capo economista di Google ha affermato che quello dello statistico «sarà uno dei mestieri "più sexy" del prossimo decennio».

Nel 2011 the Wall Street Journal ha classificato 200 lavori in rapporto a diversi parametri (stipendio, ambiente di lavoro, stress, impegno fisico, prospettive di carriera, etc): lo statistico si è classificato in quarta posizione.

## ***Perché tutto questo interesse?***

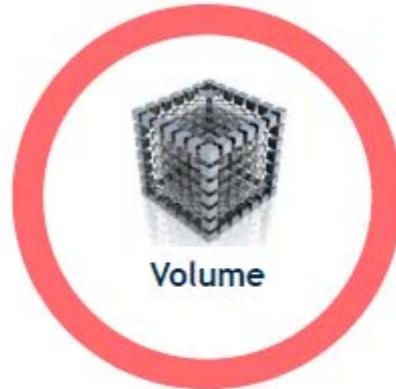
Perché i dati sono sempre più abbondanti e pervasivi nella nostra società e, di conseguenza, la capacità di trattarli e capirli, di dare loro un significato diventa sempre più rilevante.

# L'era dei BIG DATA

Il **90%** dei dati che ci circondano è stato creato nei soli ultimi due anni. I dati arrivano ormai da ogni parte e sono sempre più *digitali*:

- **sensori** che acquisiscono dati metereologici,
- contenuti condivisi sui **social media**,
- registrazioni di dati relativi agli acquisti o **dati bancari** che ogni giorno si accumulano,
- archivi di **immagini e video, dati telefonici, segnali Gps**

# Big data: what are they?



## Big Data

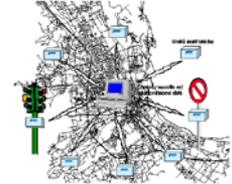


# Sommersi da un diluvio di dati

Oggi produciamo e raccogliamo dati a un ritmo esplosivo.



*Dispositivi*



*Reti di sensori*



*Social network*



*Cloud*



*Big Data*

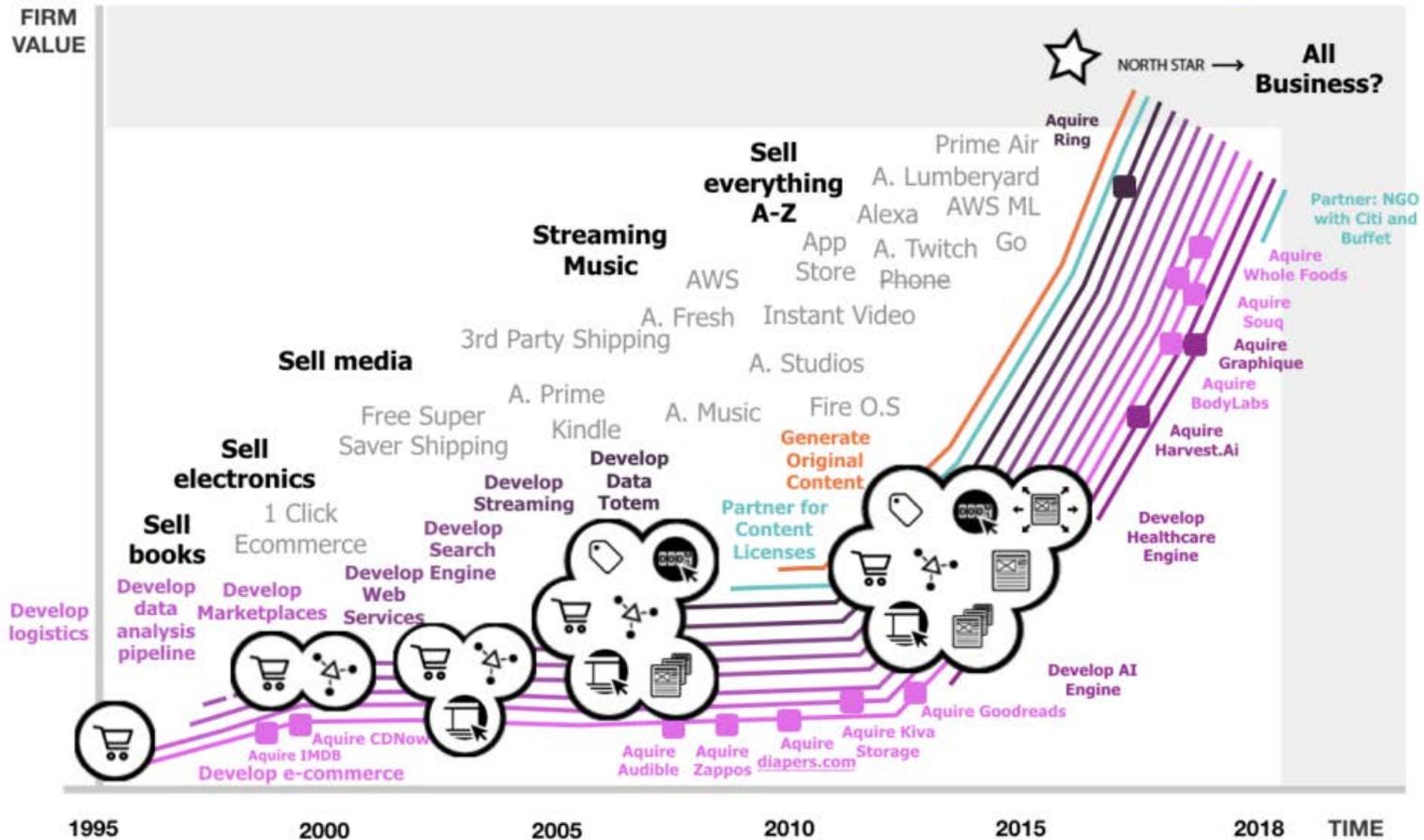


*Machine to machine*

# Le «nuove» fonti di dati medici: i wearables



# New business model



# Big data across sectors

Big data can generate significant financial value across sectors



## US health care

- \$300 billion value per year
- ~0.7 percent annual productivity growth



## Europe public sector administration

- €250 billion value per year
- ~0.5 percent annual productivity growth



## Global personal location data

- \$100 billion+ revenue for service providers
- Up to \$700 billion value to end users



## US retail

- 60+% increase in net margin possible
- 0.5–1.0 percent annual productivity growth

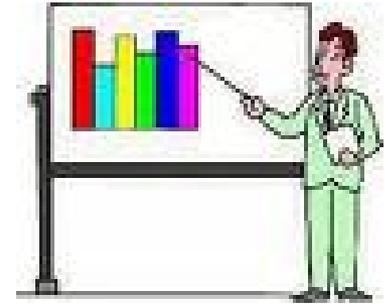


## Manufacturing

- Up to 50 percent decrease in product development, assembly costs
- Up to 7 percent reduction in working capital



# Le professioni dello statistico



- Data scientist
- Consulente per la progettazione e la realizzazione di indagini di mercato e sondaggi di opinione.
- Esperto di elaborazione ed analisi di dati medici, genetici, farmaceutici, assicurativi.
- Esperto in progettazione certificazione e controllo della qualità.
- Analista economico-finanziario.
- Risk Analyst.
- Gestore di sistemi informativi.
- Analista di politiche industriali e di politiche pubbliche.

## Dove può lavorare?

- ✓ Imprese del settore finanziario (Banche, Assicurazioni, Società di certificazione e revisione);
- ✓ Organizzazioni internazionali (World Bank, Ocse, ONU, Fao, Eurostat, ...);
- ✓ Enti di ricerca ed uffici studi (Istat, Banca d' Italia, CNR, Istituto Superiore di Sanità, ...);
- ✓ Società di informatica;
- ✓ Uffici di statistica degli enti locali;
- ✓ Pubblica Amministrazione (Ministeri, Questure, Unità Sanitarie Locali, Ospedali, ...);
- ✓ Università e Scuole medie superiori.

## Computer Science

Matematica, Informatica,  
Ingegneria

## Economia

Business, Finanza, Mercato  
del lavoro, Marketing

## Salute e Medicina

Genetica, Epidemiologia,  
Farmacologia

# DATA

Statistica Metodologica

Statistica Computazionale

Statistica Matematica

Statistica Psicometrica

Statistica Economica

Statistica Aziendale

Statistica Sociale

Statistica Attuariale

Statistica Demografica

Statistica Medica

Statistica Giudiziarla

Statistica Ambientale

Statistica Territoriale

# SCIENCE

## Governo Pubblico

Leggi, Censimenti,  
Sviluppo locale

## Scienze Sociali

Educazione, Sociologia,  
Psicologia, Sport

## Scienze Fisiche

Astronomia,  
Fisica, Chimica

## Ambiente

Ecologia, Geologia,  
Agricoltura, Veterinaria

# Rassegna stampa

## Lavoro, chi sono gli «introvabili»: i 6 profili che tutte le aziende vorrebbero assumere

di Diana Cavalcoli | 23 gennaio 2021

1 Il gap

2 Broad band architect

3 Cloud architect

4 Data scientist

5 Energy manager

6 Growth hacker

7 Ux Designer

### Data scientist

I dati sono il nuovo petrolio. E chi lavora con i dati sarà avvantaggiato nel mercato del lavoro. Ecco perché le aziende cercano data scientist in grado di leggere i numeri e interpretarli. **I professionisti in grado di usare i dati con l'obiettivo primario di aiutare a raggiungere precisi goal aziendali sono sempre più irrinunciabili.** A un data scientist è richiesta capacità di programmazione, di analisi quantitativa, di comprensione del prodotto, di comunicazione e ovviamente deve saper fare lavoro di squadra. Lo stipendio medio di un data scientist, secondo dati della Technology and IT Salary Guide per il 2018, varia in base all'esperienza **da 87 mila a 140 mila euro all'anno.**

Fonte: <https://www.corriere.it/economia/lavoro/cards/lavoro-chi-sono-introvabili-6-profilo-che-tutte-aziende-vorrebbero-assumere/data-scientist.shtml>

Come si diventa Statistico?



# Il Corso di Laurea in Scienze Statistiche

Dipartimento di Economia e Finanza (ex Facoltà di Economia)

E' ad Accesso Libero.

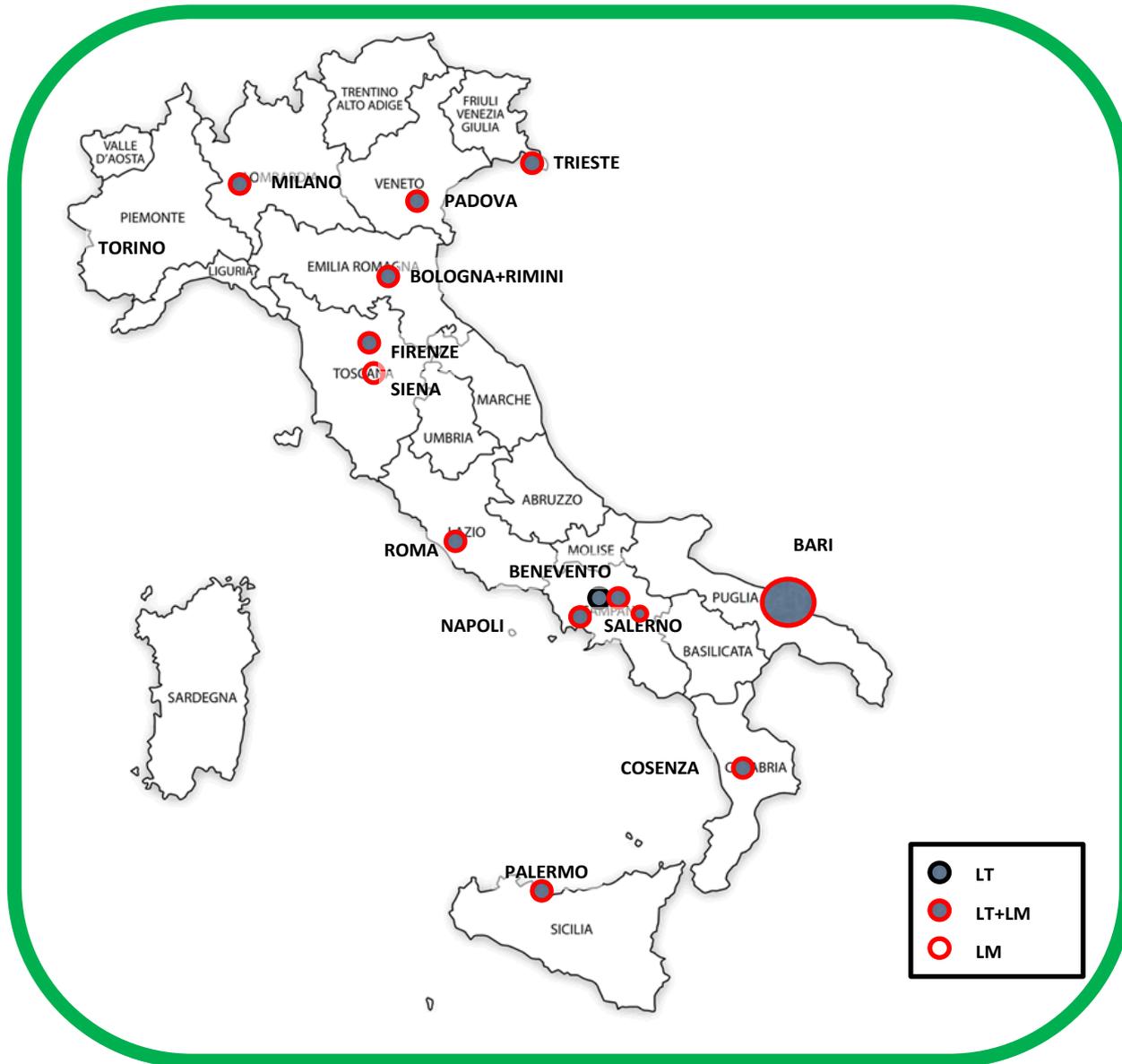
Test per gli OFA dopo l'immatricolazione.

É l'unico corso di **Statistica** in Puglia e Basilicata

É un corso di laurea della durata di **tre anni** (180 CFU).

**Dopo la laurea Triennale si può proseguire con la laurea  
Magistrale (120 CFU)**

# Sedi di Lauree Triennali e Magistrali in Statistica in Italia





# Cosa si studia: la Statistica

## **Strumenti e metodi**

per indagare/analizzare e interpretare fenomeni che riguardano un collettivo, in presenza di grandi moli di dati o di informazioni parziali.

## **Ambiti di applicazione**

Medico/sanitario, sociale, economico, industriale, servizi, etc.

## **Esempi**

Economia: modelli comportamentali, previsione azioni, marketing

Medicina: Studi clinici per nuove terapie/farmaci

Scienze politiche e sociali: sondaggi, decisioni di governo

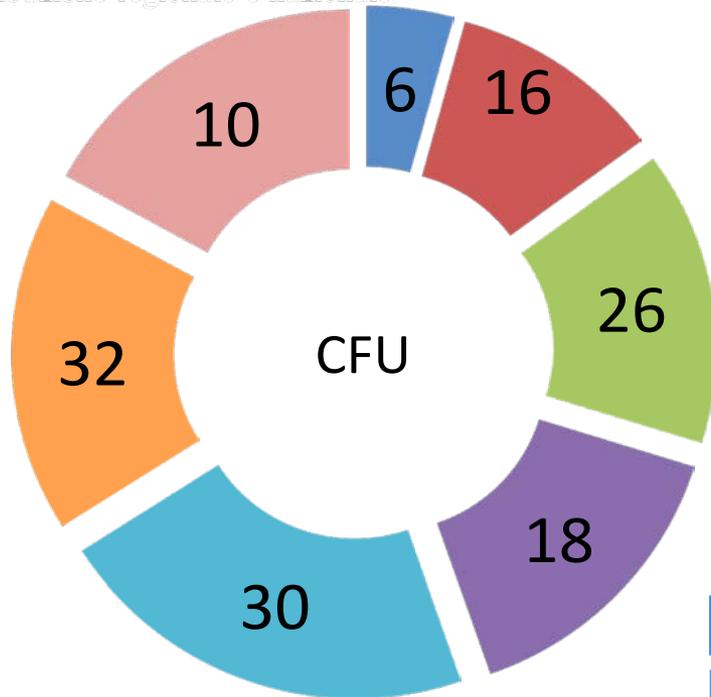


# Cosa si studia a Scienze statistiche?

distribuzione regionale e nazionale

## Ambiti disciplinari

- Inglese
- Economico
- Statistico Socio-Sanitario
- Statistico-Economico
- Statistico
- Matematico-Probabilistico
- Informatico



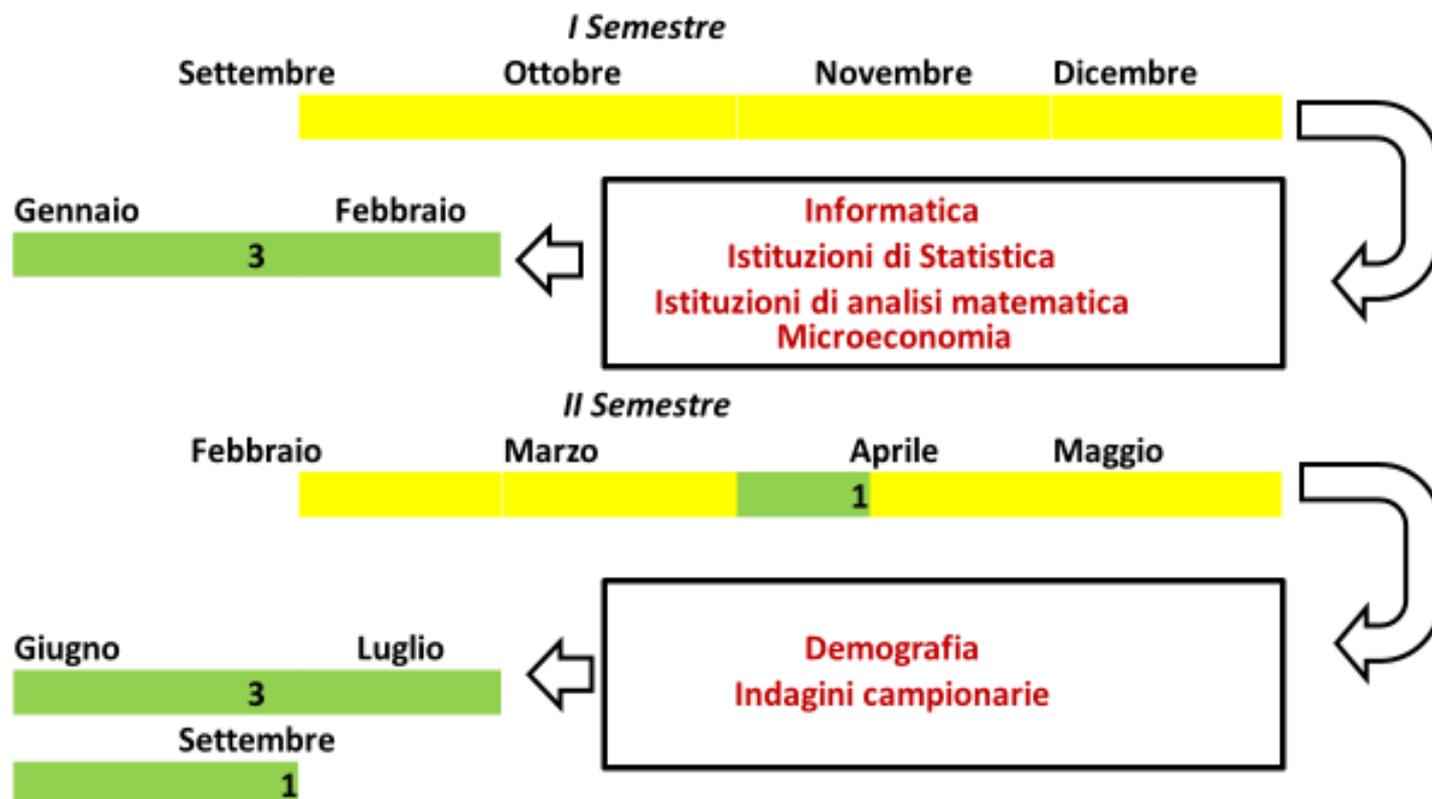
- 24 CFU a scelta tra due liste di esami
- 12 CFU a scelta dello studente (es. tirocinio opzionale)
- 3 CFU per Abilità richieste dal Mondo del lavoro  
3 CFU per la prova finale
- Nessun limite di accesso: test solo per OFA

n.b. 1 CFU = 25 h di studio complessivo, di cui 7 di lezione frontale es: corso da 8 CFU = 56 ore di lezioni frontali

# Primo Anno – Scienze Statistiche

Lezioni

Esami (n. Appelli)





## II ANNO

1. Analisi matematica e Calcolo delle probabilità
2. Algebra lineare
3. Inferenza statistica e modelli lineari
4. Matematica finanziaria
5. Macroeconomia
6. Lingua inglese
7. 2 esami a scelta

## III ANNO

1. Statistica economica
2. Statistica multivariata
3. Statistica per la ricerca sociale
4. Analisi delle serie storiche
5. 2 esami a scelta

# AGENDA

## Saper leggere ed usare i dati

- ✓ La statistica nella vita quotidiana
- ✓ I metadati: le informazioni per capire i dati
- ✓ Esempi di uso “improprio” della statistica
- ✓ Tabelle e grafici
- ✓ Media e variabilità
- ✓ Indagine statistica

# Saper leggere i dati

Le statistiche fanno parte della nostra vita quotidiana:

Tabelle, grafici, indicatori, medie, ... ci aiutano a rappresentare in maniera sintetica il mondo in cui viviamo e a prendere decisioni in condizioni di incertezza

Concorrono a determinare la nostra visione del mondo, a formare il senso comune

Senso comune: maniera ordinaria e semplice, propria della maggior parte della gente, di intendere e giudicare (Il nuovo Zingarelli)

# La Statistica nel quotidiano

Pensiamo alla **sveglia del mattino**, ognuno la regola sulla base della propria esperienza, per arrivare in orario, per non arrivare in ritardo ma senza rinunciare al sonno “più del necessario”....

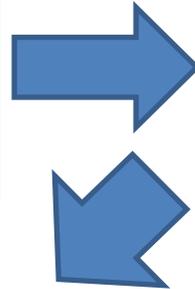
ognuno di noi ha “fatto una statistica” per puntare la sveglia ... anche senza rendersene conto ...

I **biscotti che mangiamo a colazione** sono il frutto di studi statistici sui gusti dei consumatori. Anche la loro disposizione sugli scaffali del supermercato nasce da analisi statistiche sui comportamenti in fase di acquisto.

La statistica è alla base della **programmazione televisiva**: le fasce orarie dei programmi e la loro eventuale riproposizione sono determinati sui dati di audience. Anche la pubblicità ha costi diversi in base all'audience e quindi alle fasce orarie.

# La «gestione» dei dati

Gestire i dati e saperli interpretare genera un «potere»



I dati sono la base per formulare corrette strategie e prendere decisioni

**Lo statistico partecipa ai processi decisionali**

Ma attenzione **perché talvolta i dati statistici vengono usati impropriamente**

Ad esempio per (di)mostrare una certa verità, quando magari non è supportata da alcun dato e, anzi, sembra essere smentita

## Informazioni non precise....a causa di banali errori

Un giorno, in occasione della pubblicazione dei dati INAIL sugli infortuni nei luoghi di lavoro, in un telegiornale nazionale italiano fu spiegato che il fenomeno aveva subito una riduzione apprezzabile in termini di numero di decessi.

E' corretto confrontare il dato assoluto tra un anno e il precedente?

Siamo certi che le situazioni dei due anni siano «a parità di condizioni» ???

E se i decessi fossero di meno perché ci sono meno aziende e meno occupati?

E' più corretto **relativizzare**.....

## Esempi di «uso improprio»

Quando la John Hopkins University (USA) accettò anche le donne come studenti, qualcuno pensò di riportare la notizia secondo la quale il 33,3% delle studentesse aveva sposato un insegnante. Ma a quell'epoca le donne iscritte erano solo tre ed **una** aveva sposato un professore

### «Uso corretto» per “completare” l'informazione

Quando le statistiche “condensano” le informazioni e sintetizzano i dati di un collettivo, è fondamentale conoscere:

- sia il **valore assoluto** (l'aspetto dimensionale) del fenomeno;
- sia le **percentuali**, ovvero il “contributo” che ciascuna parte apporta al collettivo nel suo complesso. Esempio:

Studentesse	Frequenze assolute
Hanno sposato un insegnante	1
<b>NON</b> hanno sposato un insegnante	2
TOTALE	3



$$\frac{1}{3} \cdot 100 = 33,3\%$$

In un certo anno le studentesse erano 3 ed una di loro ha sposato un insegnante

**Valore ASSOLUTO e PERCENTUALE: l'informazione è corretta quando sono forniti entrambi**

## «Leggere» tra i dati

Di **16 donne** di successo americane selezionate dalla Camera di Commercio di Boston intorno alla metà del XX secolo, fu detto che esse **possedevano complessivamente 60 lauree ed avevano 18 figli**

... Ma nel gruppo erano presenti due donne “particolari”, Virginia Gildersleeve, presidente del Barnard College, e Lillian M. Gilbreth, nota, insieme al marito, nello sviluppo delle tecnologie industriali

**Le due, insieme, presentavano un terzo di tutti i titoli accademici e 12 dei “18 figli” erano della signora Gilbreth** (quindi le altre 15 donne avevano 6 figli in tutto)

# Qual è il numero medio di figli per donna?

Riprendendo l'esempio delle 16 donne di successo e dei loro 18 figli:

- Se manteniamo nel collettivo tutte le donne con la rispettiva prole:

$$\frac{18}{16} = 1,12 \text{ figli per donna}$$

Traduzione in «statistichese»

*media aritmetica semplice*

$$\mu = \frac{\sum_{i=1}^S x_i}{N}$$

*media aritmetica ponderata*

$$\mu = \frac{\sum_{i=1}^N x_i * n_i}{\sum n_i}$$

- Se dal gruppo escludiamo la signora Gilbreth con i suoi 12 figli:

$$\frac{6}{15} = 0,4 \text{ figli per donna}$$



Nemmeno "mezzo" figlio per donna!!!

**Attenzione ai valori ANOMALI !!!!!!!**

**Chiamare le cose con il loro  
nome....**

**Un po' di «statistichese»**

- **fasi dell'indagine statistica**
- **tabelle statistiche**
- **rappresentazioni grafiche**
- **media aritmetica**
- **variabilità**

# L'indagine statistica

Lo studio di un fenomeno collettivo si può articolare nelle seguenti fasi:

1. progettazione dell'indagine
2. rilevazione dei dati
3. elaborazione dei dati
4. presentazione dei dati
5. interpretazione dei dati

## 1. Progettazione

E' una fase molto importante in cui si pianifica la ricerca/indagine e si definiscono:

- contesto, obiettivi, destinatari, tempi e costi..
- modalità di somministrazione, di raccolta ed elaborazione dei dati,
- attori interni ed esterni coinvolti nelle diverse fasi dell'indagine
- tutto ciò che riguarda la redazione del documento di presentazione dei risultati.

## 2. Rilevazione dei dati

La rilevazione è quel complesso di operazioni necessarie per conoscere i dati o le "manifestazioni" di uno o più caratteri di una popolazione/campione.

### 3. Elaborazione dei dati

L'elaborazione è quell'insieme di operazioni con cui i dati rilevati vengono classificati e sintetizzati per ottenere informazioni più espressive.

### 4. Presentazione dei dati

La *presentazione* è l'esposizione dei dati statistici in forma chiara e compatta, attraverso tabelle, grafici, medie, indici, ecc. con la finalità di far conoscere i risultati dell'indagine anche a chi non vi ha partecipato ma ne ha interesse.

### 5. Interpretazione dei dati

L'*interpretazione* consiste nella spiegazione dei risultati dell'indagine statistica; attraverso il commento delle tabelle, dei grafici e delle elaborazioni fatte si producono report utili alla diffusione dei risultati.

# Rilevazione dei dati

- 1. Diretta:** Questionario, intervista, sondaggio postale o telefonico, metodo Delphi.....
- 2. Indiretta:** Istat, Eurostat, OCSE, Osservatori, Ministeri (banche dati varie), Caritas, .....

## Atteggiamenti psicosociali sugli usi di mezzi tecnologici

Ti invitiamo a completare il questionario in ogni sua parte, esprimendo le tue personali opinioni. Leggi bene ciascuna domanda e dai la tua valutazione, barrando la casella che contrassegna la risposta prescelta. Il questionario è anonimo e destinato solo ad elaborazioni statistiche che verranno svolte in maniera aggregata ed in conformità alla vigente legge sulla privacy.

### 1. Genere:

- .1  Maschio
- .2  Femmina

### 2. Età (in anni compiuti): \_\_\_\_\_

### 3. Nazionalità:

- .1  Italiana
- .2  Altra (specificare) \_\_\_\_\_

### 4. Da chi è composta, oltre te, la tua famiglia?

- .1  Madre
- .2  Padre
- .3  Fratelli/sorelle (specificare quanti): \_\_\_\_\_
- .4  Altri famigliari

### 5. Titolo di studio dei tuoi genitori:

- |                            | Madre                    | Padre                    |
|----------------------------|--------------------------|--------------------------|
| .1 Licenza elementare      | <input type="checkbox"/> | <input type="checkbox"/> |
| .2 Licenza media inferiore | <input type="checkbox"/> | <input type="checkbox"/> |
| .3 Licenza media superiore | <input type="checkbox"/> | <input type="checkbox"/> |
| .4 Laurea                  | <input type="checkbox"/> | <input type="checkbox"/> |

### 6. Condizione lavorativa:

- |  | Madre                    | Padre                    |
|--|--------------------------|--------------------------|
| .1 Occupato                                    | <input type="checkbox"/> | <input type="checkbox"/> |
| .2 Disoccupato o in cerca di prima occupazione | <input type="checkbox"/> | <input type="checkbox"/> |
| .3 Casalinga                                   | <input type="checkbox"/> | <input type="checkbox"/> |
| .4 Altra condizione                            | <input type="checkbox"/> | <input type="checkbox"/> |

### 7. Quanto sei soddisfatto/a del tuo aspetto fisico?

- .1  Per nulla
- .2  Poco
- .3  Abbastanza
- .4  Molto

### 8. Quando hai qualche problema, con chi ti confidi

### 12. Hai accesso ad internet (tramite smartphone, tablet o pc):

- |               | Sì                       | No                       |
|---------------|--------------------------|--------------------------|
| .1 In casa    | <input type="checkbox"/> | <input type="checkbox"/> |
| .2 Fuori casa | <input type="checkbox"/> | <input type="checkbox"/> |

### 13. Quando sei su internet, hai un adulto che controlla (anche se non sempre) su quali siti navighi?

- .1  Sì
- .2  No

### 14. Quali dei seguenti social network o mezzi di messaggistica istantanea usi ?

- .1  Facebook
- .2  Twitter
- .3  Whatsapp/Viber/WeChat/Line ecc
- .4  Ask.fm
- .5  Instagram
- .6  Youtube
- .7  Altro (specificare): \_\_\_\_\_
- .8  Nessuno dei precedenti (vai alla definizione di cyberbullismo prima della domanda 17)

### 15. I tuoi genitori possiedono le password per accedere ai tuoi account dei social network?

- .1  Sì
- .2  No

### 16. Riesci a dire quanto tempo al giorno dedichi all'incirca ai social network o ai mezzi di messaggistica istantanea?

- .1  Meno di 1 ora
- .2  Da 1 a 3 ore

# Campione o collettivo???

L'insieme di tutte le unità statistiche che sono uguali rispetto ad alcuni loro caratteri è definito **collettivo statistico o popolazione**

*Esempio: gli Italiani sono un collettivo statistico costituito da tutti gli esseri umani nei quali il carattere "nazionalità" è presente con la modalità italiani*

Si parla di **rilevazione totale** quando vengono rilevate tutte le unità di un collettivo o popolazione (che si denota con  $N$ ). Ne è un esempio il censimento.

Si parla di **rilevazione parziale o campionaria** quando vengono rilevate solo alcune unità della popolazione (la cui numerosità si denota con  $n$ ).

Il **campione statistico** è una parte di tutte le unità statistiche costituenti il collettivo/popolazione, con la caratteristica di dare un'immagine **ridotta** ma **fedele** delle caratteristiche del collettivo.

Si definisce **rappresentativo** se riproduce **in scala tutte** le caratteristiche della popolazione da cui è estratto.

## Tabella statistica

Un carattere o variabile osservato assume, nella realtà, diverse manifestazioni o modalità.

L'insieme delle unità statistiche rilevate secondo le modalità vengono schematizzate attraverso le **tabelle statistiche** semplici o multiple.

Nel caso **qualitativo**, le **modalità** sono **attributi o categorie**, in quello **quantitativo** sono **numeri**.

Le **frequenze** rappresentano l'esito di un conteggio oppure l'intensità del fenomeno.

Modalità del carattere ( $X_i$ )	Frequenze assolute ( $n_i$ )
$x_1$	$n_1$
$x_2$	$n_2$
$x_3$	$n_3$
.	.
$x_i$	$n_i$
.	.
$x_s$	$n_s$
<b>Totale</b>	<b>N</b>

N. B. in una tabella le modalità di un carattere quantitativo sono ordinate in senso crescente

# La sommatoria

Il simbolo  $\Sigma$

$$\sum_{i=1}^s n_i = N$$

indica la sommatoria delle frequenze  
ossia:

$$n_1 + n_2 + n_3 + n_i + \dots + n_s = N$$

# Vari tipi di frequenze

Frequenze  
Assolute



$$\sum_{i=1}^s \mathbf{n}_i = N$$

Frequenze  
Relative



$$y_i = \frac{\mathbf{n}_i}{N}$$

Frequenze  
Percentuali



$$P_i = (\mathbf{n}_i / N) * 100$$

## Mutabili rettilinee

**Mutabile rettilinea** si ha quando le modalità ammettono un ordine naturale di successione, con una modalità iniziale ed una finale.

<b>Grado di istruzione</b>	<b>Popolazione residente</b>
Analfabeti	10
Scuola elementare	30
Scuola media inferiore	40
Diploma	50
Laurea	70
Totale	200

## Mutabili cicliche

La **Mutabile ciclica** si ha quando le modalità della distribuzione presentano un ordine naturale di successione, ma non si può dire, a meno di una convenzione, quale sia prima e quale sia l'ultima modalità.

Giorni della settimana	Ore lavorate
Lunedì	6
Martedì	9
Mercoledì	6,5
Giovedì	8
Venerdì	7,5
Sabato	0
Domenica	0
Totale	37

## Mutabili sconnesse

La **mutabile sconnessa** si ha quando le modalità non presentano un ordine naturale di successione

Gruppi di corsi di studio	Studenti Intervistati
Scienze Formazione	30
Economia	15
Giurisprudenza	15
Medicina e Chirurgia	20
Totale	80

# Tabelle statistiche doppie

Rappresentano contemporaneamente due caratteri quantitativi  $X_1$  e  $X_2$ , per cui la tabella statistica doppia si configura come l'insieme delle coppie di valori  $(x_{1i}, x_{2i})$  corrispondenti alla stessa unità  $i$ .ma del collettivo  $(i=1, 2, 3, \dots, n)$ .

La tabella statistica doppia è rappresentata da una tabella a doppia entrata.

# Mutabili statistiche doppie

Rappresentano contemporaneamente due caratteri qualitativi  $A_1$  e  $A_2$ , per cui la mutabile statistica doppia si configura come l'insieme delle coppie di valori  $(a_{1i}, a_{2i})$  corrispondenti alla stessa unità  $i$ .ma del collettivo  $(i=1,2,3,\dots,n)$ .

La mutabile statistica doppia è rappresentata da una tabella detta di **contingenza**.

# Le rappresentazioni grafiche

Sono **strumenti statistici** che permettono di:

- interpretare più velocemente le informazioni raccolte sul fenomeno osservato,
- di coglierne **immediatamente** alcune **caratteristiche**
- di **evidenziare/far capire meglio** anche ai «non addetti» ai lavori.

Talvolta i grafici **sostituiscono** le tabelle, altre volte li integrano e sono un utile supporto per l'analisi statistica.

Per ogni distribuzione statistica semplice o doppia o multipla esiste il tipo di **rappresentazione grafica adatta**; talvolta una stessa distribuzione può essere rappresentata con più tipologie di grafico.

**Affinché una rappresentazione grafica sia utile ed efficace dovrebbe contenere con immediatezza e chiarezza tutte le informazioni necessarie alla comprensione dei dati in essa rappresentati.**

# LA RAPPRESENTAZIONE GRAFICA DEI DATI

(tratto da ISTAT, Pacchetti didattici

<http://www.istat.it/it/istituto-nazionale-di-statistica/attivita/ex-scuola-superiore-statistica/under-21/pacchetti-didattici>).

The screenshot shows the ISTAT website interface. At the top, there is a red navigation bar with links for 'Istituto nazionale di statistica', 'Bandi di gara', 'Concorsi', and 'Amministrazione trasparente'. Below this is the ISTAT logo and a search bar. The main content area features a large image of the ISTAT building facade with the inscription 'NUMERVS RERV OMNIVM NODVS' and 'ANNO-IX'. To the right of the image, there is a text block stating: 'Dal 1926 l'Istat è il principale produttore di statistica ufficiale a supporto dei cittadini, degli operatori economici e dei decisori pubblici'. Below the image, there is a breadcrumb trail: 'Home / Istituto nazionale di statistica / Attività / Ex-Scuola superiore statistica / Under 21 / pacchetti-didattici'. The main section is titled 'Pacchetti didattici a cura della rete territoriale' and includes a sub-section for 'ASCOLTA'. Underneath, there are three categories of educational packages: 'Scuola primaria', 'Scuola secondaria di primo grado', and 'Scuola secondaria di secondo grado'. To the right of the main content, there is a sidebar with sections for 'Organizzazione', 'Attività', 'Qualità', and 'Presidente'. The 'Organizzazione' section lists 'Organigramma', 'Sedi territoriali', and 'Normativa'. The 'Attività' section lists 'Rilevazioni ed elaborazioni', 'Ex-Scuola superiore statistica', 'Open data', 'Attività internazionali', and 'Sistema statistico europeo'. The 'Qualità' section lists 'Il percorso fatto', 'Codici della statistica', and 'Strumenti'. The 'Presidente' section lists 'Articoli e interviste'. At the bottom of the sidebar, there are social media icons (Twitter, Facebook, YouTube, LinkedIn, RSS) and a section titled 'In evidenza' which lists 'Calendario delle diffusioni e degli eventi', 'SEC 2010: il nuovo sistema europeo dei conti', 'Previsioni economiche e microsimulazioni', 'Informazioni territoriali e cartografiche', 'Immigrati e nuovi cittadini', and 'Attività della ex-Scuola superiore di statistica'. At the very bottom of the page, there is a banner for 'Indici per aggiornare AFFITTI e ASSEgni FAMILIARI' and another for 'naitalia | edizione 2015'.

Le **rappresentazioni grafiche** hanno l'obiettivo di illustrare, mediante:

- figure,
- linee o segmenti,
- superfici o aree,
- solidi,
- simboli convenzionali,

una distribuzione di frequenze o di intensità in funzione delle modalità di uno o più caratteri per

-una **visualizzazione immediata dell'andamento del fenomeno e della distribuzione**, che consente una efficace e globale descrizione dei dati;

-una **sintesi** con la possibilità, in poco spazio, di confrontare più distribuzioni (curve, spezzate ecc.);

- per finalità **investigative**:

- mette in risalto i casi "anomali" (particolari *picchi* grafici), imputabili a errori nei dati o a effettivi casi anomali, che vanno ulteriormente approfonditi,
- evidenzia le relazioni tra caratteri,
- permette di individuare andamenti di fondo (*trend*) interpolabili con funzioni matematiche (ad es. curva normale, retta ecc.).



## 2. Rappresentazioni grafiche di caratteri qualitativi/mutabili

- a) Grafici a barre
- b) Diagrammi circolari (torta)
- c) Diagrammi in coordinate polari (radar)
- d) Cartogrammi, mappe tematiche

### RICORDA:

mutabili sconnesse: serie in cui non è possibile riscontrare alcun ordine di successione tra le modalità. (professione, nazionalità, religione e partito politico degli intervistati ecc)

mutabili rettilinee: serie in cui è possibile riscontrare un ordine logico naturale di successione dei termini. Esiste cioè un termine che rappresenta una modalità iniziale del fenomeno, ed un altro che rappresenta una modalità finale di esso.

## a) Grafici a barre

(1/3)

I grafici a barre (o ortogrammi) sono impiegati per rappresentare graficamente **serie sconnesse** o **rettilinee** e possono essere di due tipi:

-**A colonne**, se sono costituiti da una successione di colonne, segmenti verticali o rettangoli (a base uguale o arbitraria) equidistanti, in numero pari alle modalità del carattere, e hanno *altezza* uguale o proporzionale alla frequenza (assoluta o relativa) o all'intensità della modalità da rappresentare;

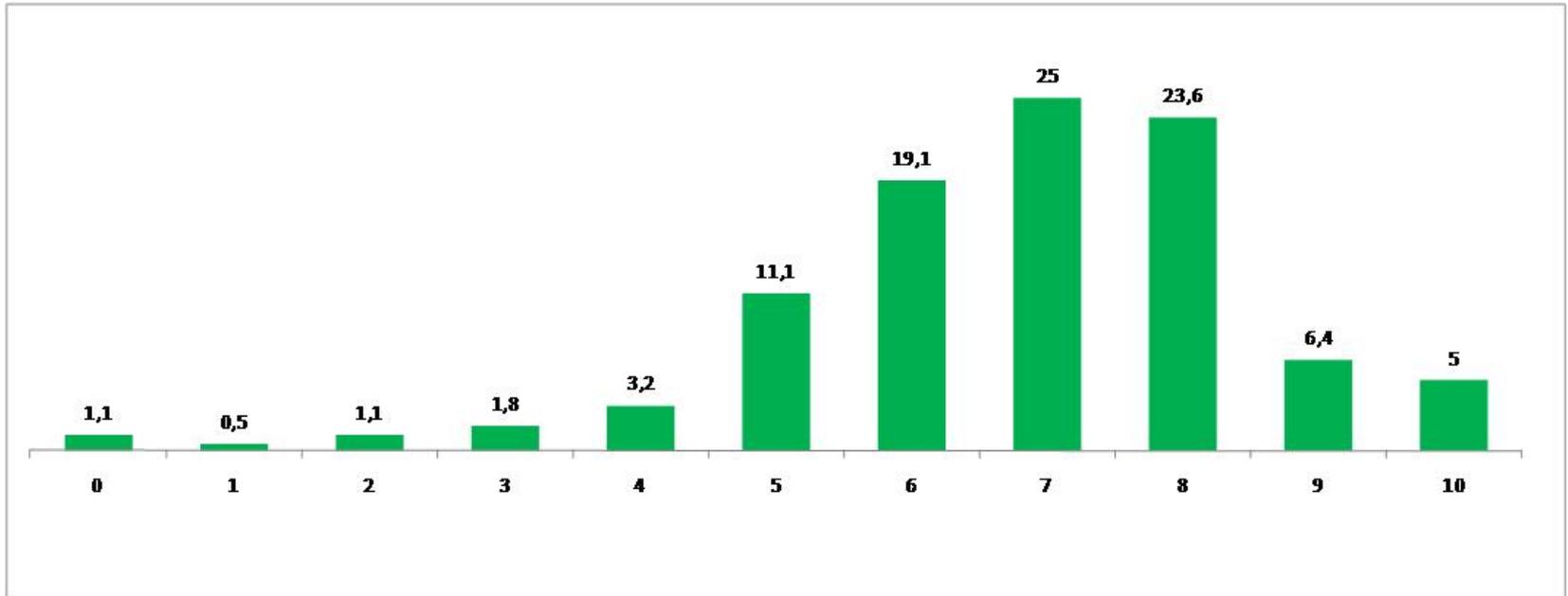
-**A nastri**, se sono costituiti da tanti nastri (segmenti orizzontali, rettangoli) sovrapposti ed equidistanti, in numero pari alle modalità del carattere, e hanno *lunghezza* uguale o proporzionale alla frequenza (assoluta o relativa) o all'intensità della modalità da rappresentare.

I grafici a barre sono pertanto caratterizzati dall'averne un solo asse (verticale nel caso a colonne, orizzontale nel caso a nastri) in scala graduata, secondo l'unità di misura che si è scelta per rappresentare le frequenze o intensità. Sull'altro asse, invece, figureranno le modalità (qualitative), per convenzione equidistanti.

## a) Ortogramma a colonne

(2/3)

**Persone di 14 anni e più per livello di soddisfazione per la vita nel complesso (a) in Italia  
Anno 2013** (per 100 persone di 14 anni e più)



Fonte: Istat, Indagine Multiscopo "Aspetti della vita quotidiana"

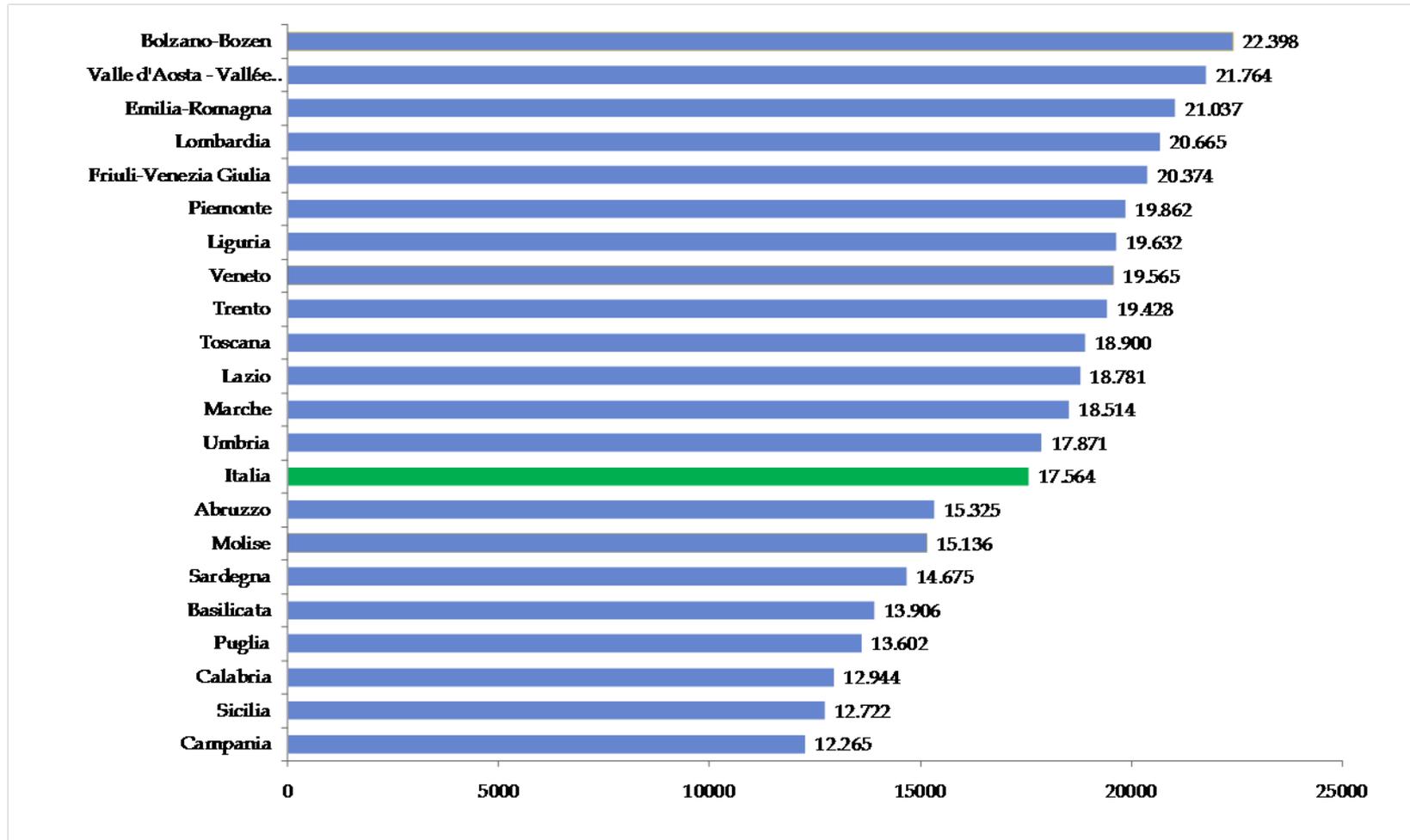
(a) Espresso con un punteggio da 0 (per niente soddisfatto) a 10 (molto soddisfatto)

Se la rappresentazione grafica riguarda una serie sconnessa, l'ordine in cui saranno poste le modalità è arbitrario; se si tratta invece di una serie rettilinea (come in questo caso), le modalità saranno poste nell'ordine naturale che esse presentano nella serie.

## a) Ortogramma a nastri

(3/3)

Reddito disponibile delle famiglie per abitante per regione – Anno 2012 (*valori in euro*)



## b) Diagrammi circolari

(1/2)

I diagrammi circolari (o aereogrammi) per la loro forma circolare, sono comunemente noti come “diagrammi a torta”.

Sono particolarmente adatti alle serie **sconnesse o rettilinee**.

Sono efficaci per mettere in evidenza l'importanza relativa delle singole modalità rispetto al totale.

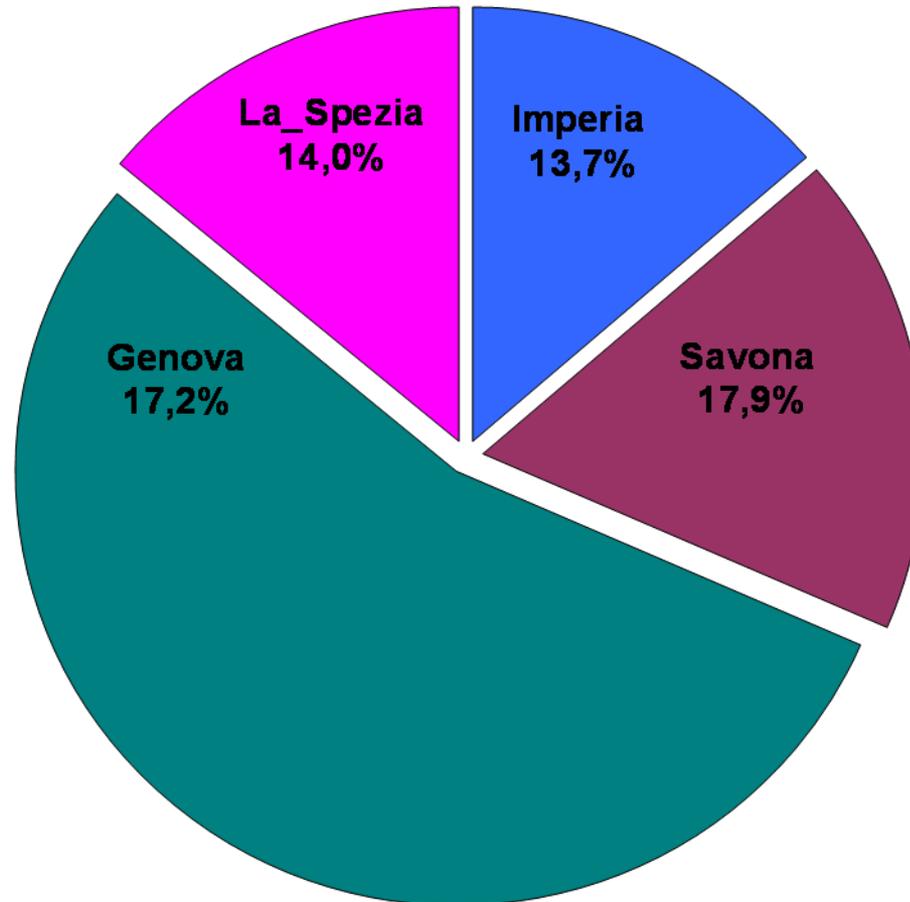
**Gli spicchi o settori variabili**, presentano angoli al centro corrispondenti alle frequenze assolute (o relative) delle singole modalità e raggio fisso.

## b) Diagrammi circolari

(2/2)

Composizione percentuale della popolazione residente in Liguria per provincia

Censimento 2011



Fonte: Istat

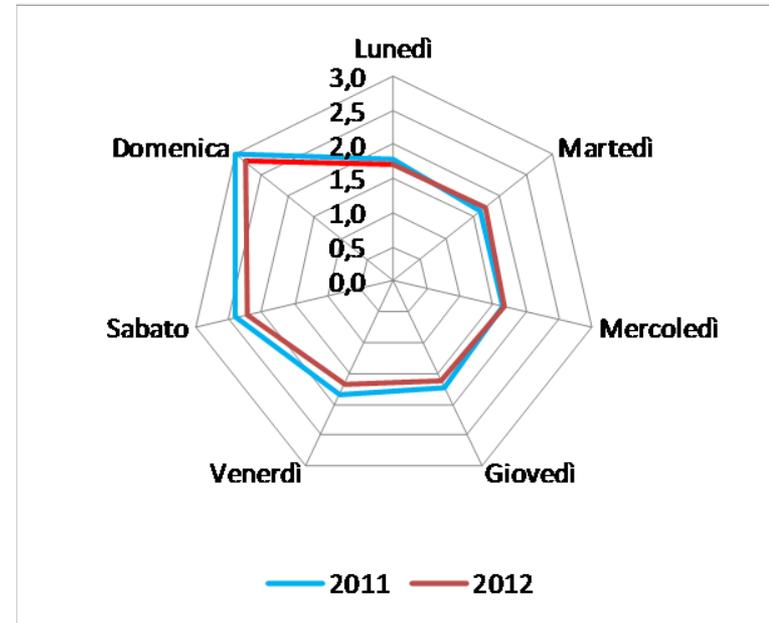
## c) Diagrammi in coordinate polari

Si utilizzano per rappresentare serie cicliche.

Le **serie cicliche** sono distribuzioni di mutabili che dispongono di un ordinamento naturale ma non di una prima e un'ultima modalità.

**Esempio.** I nati, i matrimoni, le vendite di determinati prodotti, fenomeni astronomici o atmosferici e così via secondo i mesi dell'anno, i giorni della settimana ecc.

Indice di mortalità in incidenti stradali per giorno della settimana  
Anni 2011 e 2012



Fonte: Istat

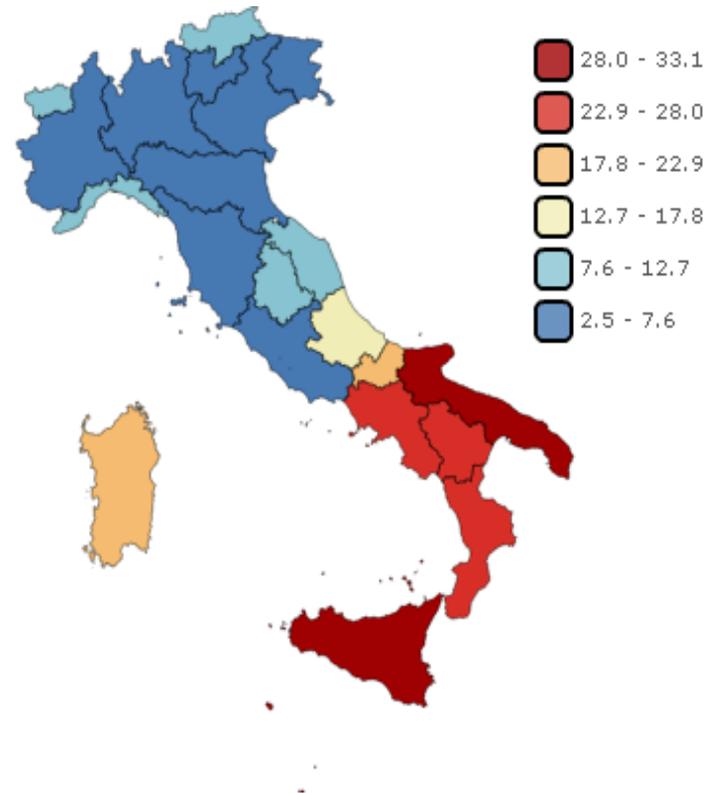
## d) Cartogrammi, mappe tematiche

(1/2)

I cartogrammi sono grafici utili per rappresentare **serie territoriali**.

Per costruire un cartogramma occorre disporre di una carta geografica o topografica in cui siano chiaramente delimitate le diverse zone, regioni, circoscrizioni (geografiche, politiche, amministrative) rispetto alle quali viene analizzata l'intensità o la frequenza di uno o più caratteri (es. nati, morti, reddito pro capite, secondo le Regioni, Province, Comuni).

Incidenza della povertà relativa per regione  
Anno 2012 (valori percentuali)



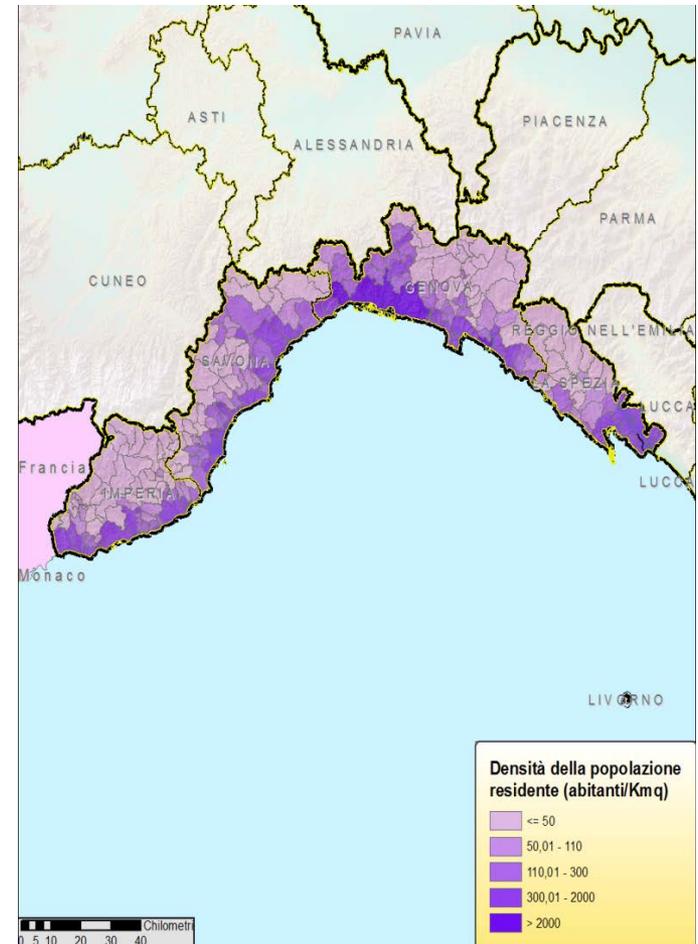
Fonte: Istat

## d) Cartogrammi, mappe tematiche

(2/2)

Le mappe tematiche consentono, anche se molto empiricamente e in prima approssimazione, di tener conto dell'autocorrelazione spaziale, intesa come possibilità che un fenomeno che interessa un certo territorio sia influenzato nelle sue manifestazioni dalla contiguità spaziale esistente tra i luoghi in cui il fenomeno è osservato.

Per esempio nella mappa della Liguria qui a destra, si nota come la densità della popolazione residente sia maggiore nei comuni costieri e in quelli circostanti i comuni capoluogo rispetto a quelli dell'entroterra.



Fonte: Istat

### 3. Rappresentazioni grafiche di caratteri quantitativi

- a) Istogrammi
- b) Diagrammi cartesiani a segmenti
- c) Poligono e curve di frequenza per variabili continue
- d) Rappresentazioni di tipo informatico:
  - a ramo e foglia (stem and leaf plot)
  - a scatola (boxplot)
  - le facce di Chernoff
  - ideogrammi o pictogrammi

## a) Istogrammi

(1/3)

Gli istogrammi si impiegano per rappresentare graficamente distribuzioni di frequenza di **caratteri quantitativi** le cui modalità sono costituite da **classi di valori**.

L'ipotesi di base su cui si fonda la rappresentazione per istogrammi è che **all'interno di ciascuna classe le unità siano equidistribuite**.

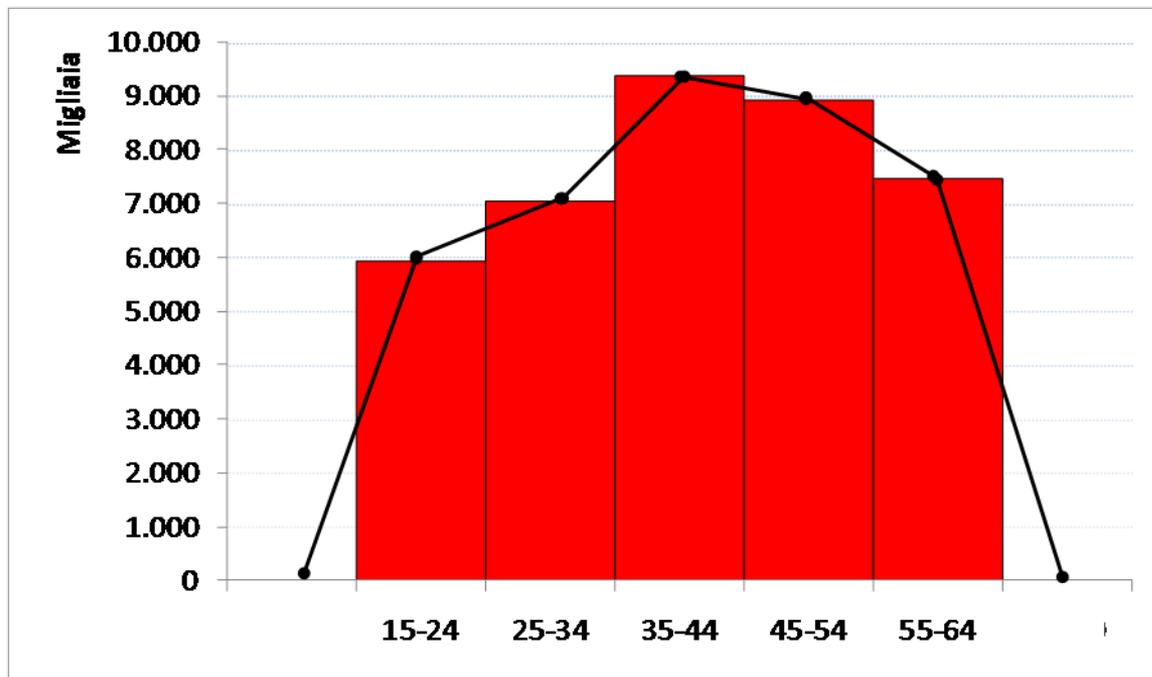
Per rendere più evidente l'andamento e la forma della distribuzione, si possono congiungere con segmenti di retta i punti centrali dei lati superiori dei rettangoli ottenendo una linea spezzata detta **poligono di frequenza**, la quale deve essere chiusa toccando l'asse delle ascisse all'esterno delle classi estreme, in modo che l'area all'interno del poligono di frequenza equivalga a quella dell'istogramma.

## a) Istogrammi

(2/3)

-Le **classi di valori** hanno **uguale ampiezza**. In questo caso avremo tanti rettangoli contigui, ciascuno avente base uguale all'ampiezza della classe e altezza uguale o proporzionale alla frequenza (assoluta o relativa) assunta nell'insieme delle unità della classe;

Istogramma e poligono di frequenza della popolazione residente in Italia in età lavorativa per classi decennali di età – Censimento 2011 (popolazione in migliaia)

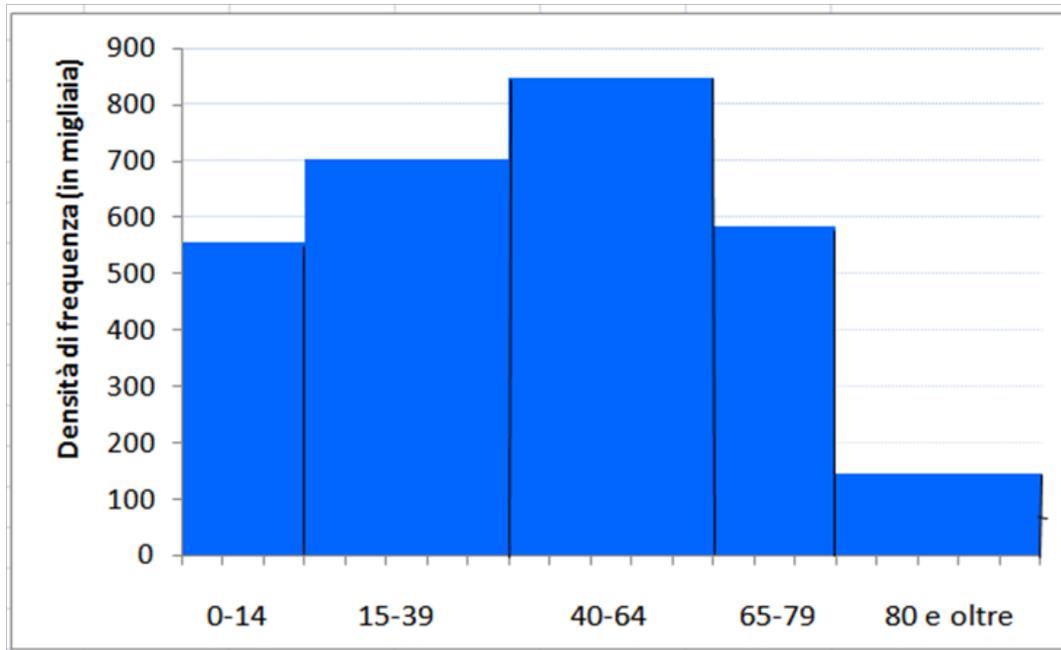


## a) Istogrammi

(3/3)

- Le **classi di valori** hanno **diversa ampiezza**. In quest'altro caso avremo una serie di rettangoli aventi basi diverse uguali all'ampiezza delle classi e altezze da calcolarsi, in modo che le frequenze siano proporzionali alle aree dei rispettivi rettangoli.

Istogramma della popolazione residente in Italia per classi di età – Censimento 2011 (popolazione in migliaia)



In ordinata abbiamo le **densità di frequenza** date dal rapporto tra la frequenza (assoluta o relativa) di ciascuna classe e la relativa ampiezza.

Fonte: Istat

## b) Diagrammi cartesiani a segmenti (1/2)

Sono impiegati per rappresentare graficamente caratteri **quantitativi discreti**, non divisi in classi, e possono configurarsi a segmenti verticali.

**Esempio.** Numero dei componenti per famiglia, numero delle stanze delle abitazioni, numero di unità locali delle aziende e così via.

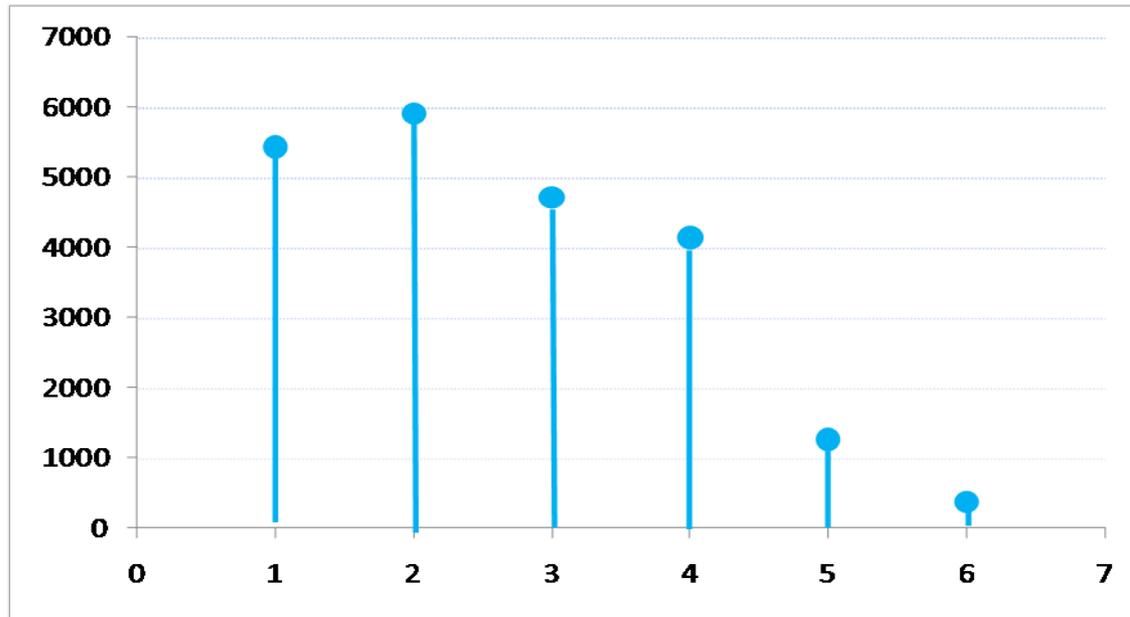
Essi si costruiscono come gli usuali diagrammi cartesiani aventi due assi perpendicolari: l'asse delle ascisse ( $x$ ) e l'asse delle ordinate ( $y$ ), aventi origine comune in zero. Ogni coppia ordinata di valori  $(x_i, y_i)$  determinerà un punto nel piano e l'insieme di tutte le coppie ( $x_i =$  modalità quantitativa  $i$ -esima,  $y_i =$  frequenza della modalità  $i$ -esima) determinerà l'insieme dei punti nel piano che costituiscono la rappresentazione grafica della distribuzione considerata.

Per rendere maggiormente visibili tali punti, si tracciano dei segmenti verticali congiungenti l'ascissa ( $x_i$ ) con il punto del piano corrispondente all'ordinata ( $y_i$ ).

## b) Diagrammi cartesiani a segmenti

(2/2)

Distribuzione delle famiglie residenti in Italia per numero di componenti – Censimento 2001 (valori in migliaia)



Fonte: Istat

E' da notare che in questo caso è scorretto costruire il poligono o spezzata di frequenza congiungendo tra loro i punti poiché il **carattere** considerato è **discreto** e quindi non possiede i valori intermedi a quelli indicati dalle modalità quantitative.

Una spezzata di frequenza che unisse tra loro le modalità, infatti, attribuirebbe anche valori intermedi alle modalità stesse.

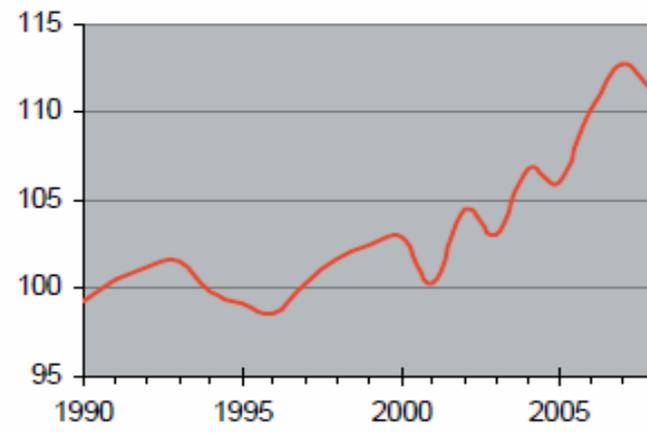
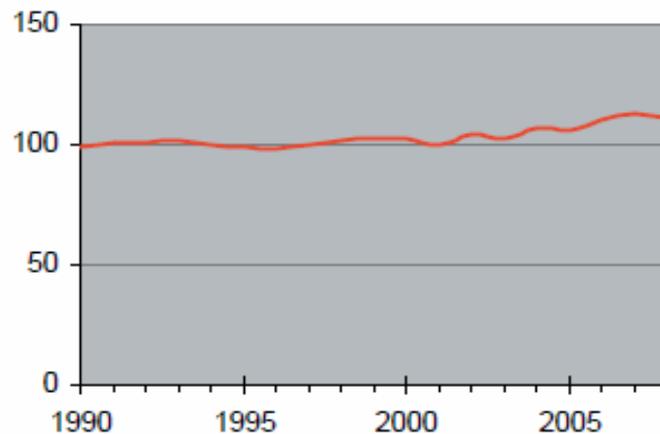
## c) Diagrammi cartesiani per serie storiche (1/4)

I diagrammi cartesiani ortogonali sono impiegati anche per rappresentare graficamente le serie storiche, per i caratteri che si suppone si modificano con continuità nel tempo.

La costruzione di tali diagrammi è del tutto analoga a quella vista per i diagrammi ad aste o segmenti salvo che, in questo caso, una volta ottenuti nel piano cartesiano i punti  $(x_i, y_i)$  rappresentanti la distribuzione considerata, essi vanno uniti mediante una spezzata.

## c) Attenzione ad adottare la giusta scala(3/4)

**Problemi di scala:** questi grafici rappresentano la stessa distribuzione ma l'andamento della serie viene percepito in maniera diversa.



Fonte: UNECE, Making Data Meaningful Part 2: A guide to presenting statistics, United Nations, Geneva 2009.

Il grafico a sinistra mostra un trend stabile per circa 10 anni, seguito da una moderata crescita.

Riducendo la scala sull'asse delle ordinate (grafico a destra), risultano più evidenti le variazioni intervenute nei primi dieci anni e sembra più marcata la crescita dell'ultimo periodo.

## c) Diagrammi a barre e cartodiagrammi

(1/4)

Nel caso di distribuzioni statistiche doppie in cui vi sono due mutabili o una variabile e una mutabile (**tabelle di contingenza**) esistono diverse possibili rappresentazioni grafiche, a seconda della natura dei due caratteri considerati e degli scopi della rappresentazione grafica.

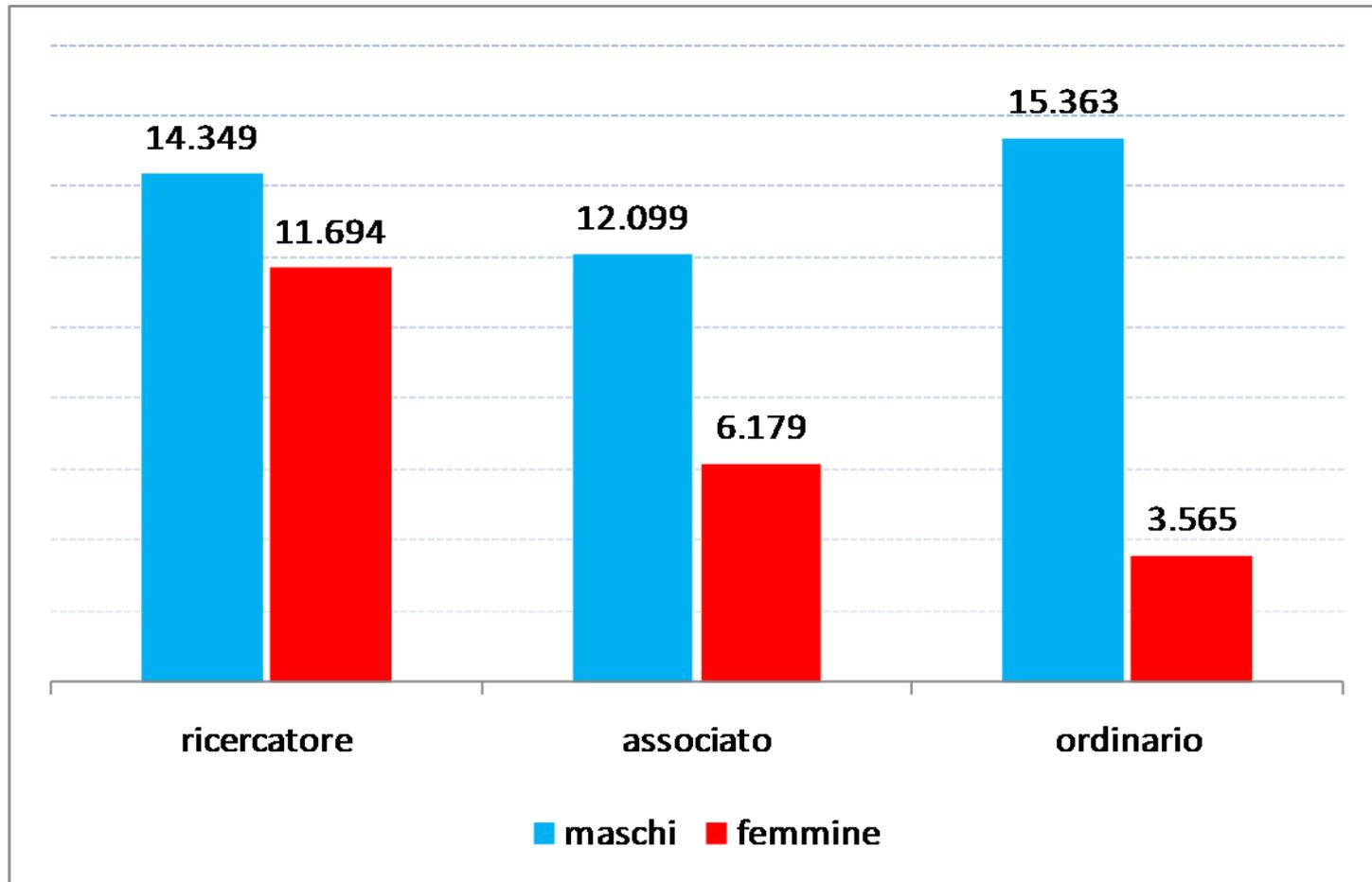
Tra le rappresentazioni più impiegate vi è quella già vista per le mutabili statistiche semplici, il diagramma a barre, anche se in questo caso sono rappresentati i due caratteri e le frequenze di associazione delle modalità dei medesimi: **Diagramma a barre verticali** e **Diagramma a barre per impilamento**.

## c) Diagrammi a barre e cartodiagrammi

(2/4)

### Diagramma a barre verticali

Docenti universitari per qualifica e genere – Italia – Anno 2008



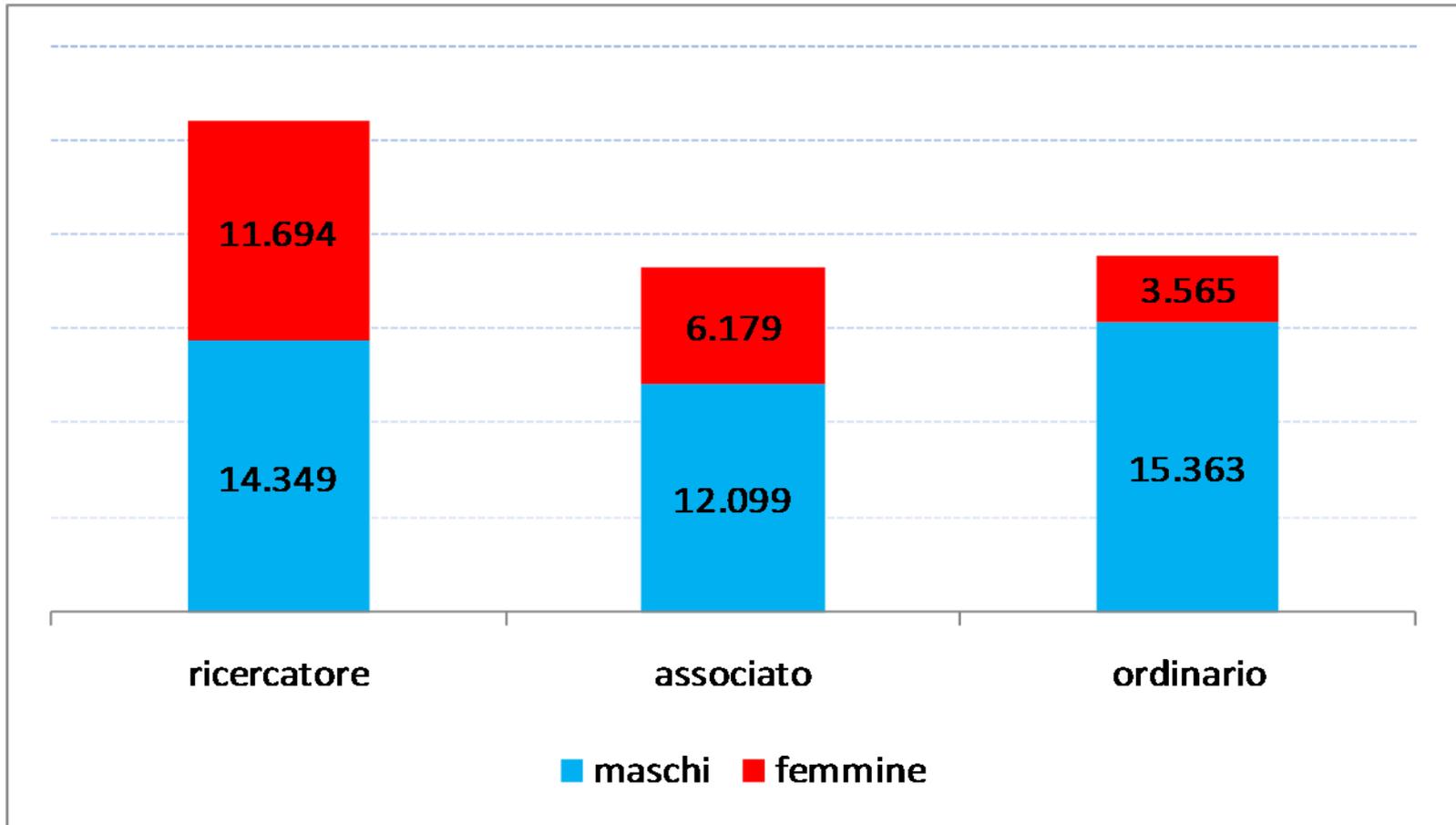
Fonte: Istat

## c) Diagrammi a barre e cartodiagrammi

(3/4)

### Diagramma a barre per impilamento

Docenti universitari per qualifica e genere – Italia – Anno 2008



## c) Diagrammi a barre e cartodiagrammi

(4/4)

I cartodiagrammi non sono altro che dei cartogrammi in cui, anziché delle serie territoriali semplici, vengono rappresentate delle serie territoriali di due o più caratteri.

Esempio: I nati vivi e i morti per 1.000 abitanti nelle 20 Regioni italiane nel 1986.



# Tabella 1 - Tipo di grafico secondo il livello di misurazione dei caratteri

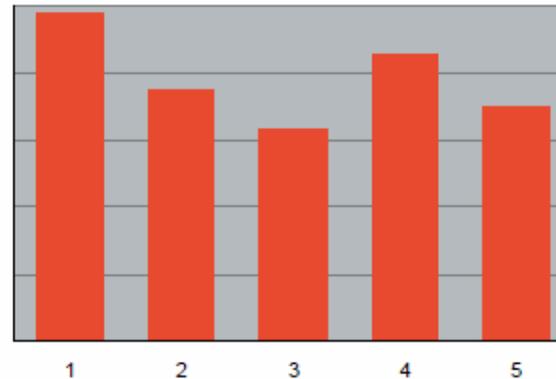
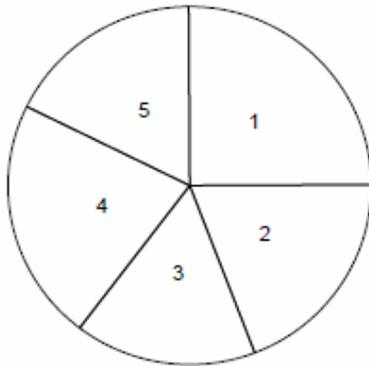
Tipo di grafico	Livello di misurazione dei caratteri
Grafici a barre o ortogrammi: -a colonne (verticale) -a nastri (orizzontale)	Distribuzioni semplici (serie): caratteri qualitativi; mutabili sconnesse, mutabili rettilinee
Diagrammi circolari o areogrammi circolari: -a spicchi proporzionali -a spicchi uguali	Distribuzioni semplici (serie): caratteri qualitativi; mutabili sconnesse, mutabili rettilinee
Diagrammi in coordinate polari	Distribuzioni semplici (serie): caratteri qualitativi; mutabili cicliche
Cartogrammi, mappe tematiche	Distribuzioni semplici (serie): caratteri qualitativi mutabili sconnesse riferite a luoghi, territori ecc. (serie territoriali e spaziali)
Istogrammi e poligono di frequenza	Distribuzioni semplici (seriazioni): caratteri quantitativi: variabili (continue o semicontinue) divise in classi di valori
Diagrammi in coordinate cartesiane ortogonali a canne d'organo	Distribuzioni (seriazioni): caratteri quantitativi: discrete Distribuzioni semplici: serie temporali riferite a fenomeni discreti
Diagrammi in coordinate cartesiane ortogonali (poligoni di frequenza e curve di frequenza)	Distribuzioni (seriazioni): caratteri quantitativi: continue e semicontinue Distribuzioni semplici: serie temporali riferite a fenomeni continui o semicontinui
Ideogrammi o pictogrammi o diagrammi a figure simboliche	Per distribuzioni semplici e doppie e qualsiasi tipo di carattere. Di carattere divulgativo ma non scientifico

## La scelta della rappresentazione grafica

I 2 grafici sottostanti rappresentano la stessa distribuzione. Qual è più chiaro? Quale settore del diagramma circolare è maggiore?

Per la maggior parte delle persone è più facile confrontare segmenti piuttosto che angoli.

Nel diagramma circolare i settori numero 1 e 4 sembrano identici, mentre nel diagramma a barre è evidente la differenza



Fonte: UNECE, Making Data Meaningful Part 2: A guide to presenting statistics, United Nations, Geneva 2009.

E' opportuno rappresentare la stessa distribuzione con più grafici per individuare quello che meglio rappresenta il messaggio che si vuole veicolare.

# E' importante saper elaborare i dati

Basta veramente solo sommare le modalità ( $x_i$ ) e dividere per N???

**Problema:** Maria vuole sapere quanto tempo trascorre normalmente al telefono. Dal sito web del suo operatore trascrive la durata in minuti di 12 chiamate scelte casualmente (1,7,4,1,2,4,3,48,3,5,3,6).  
Analizzare i dati e calcolare la durata media delle chiamate

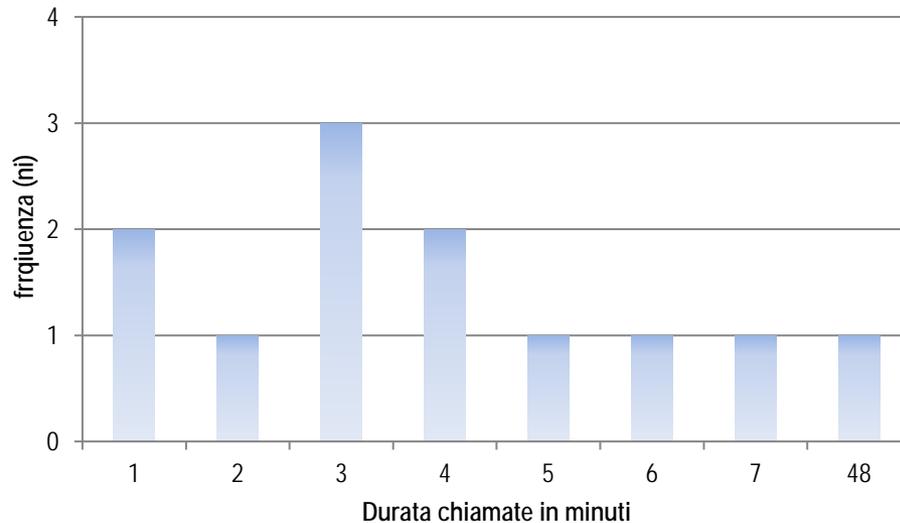


Tabella 1 - Distribuzione delle telefonate di Maria

Durata telefonate in minuti ( $x_i$ )	frequenza ( $n_i$ )	$x_i \cdot n_i$
1	2	2
2	1	2
3	3	9
4	2	8
5	1	5
6	1	6
7	1	7
48	1	48
Totale	12	87
media	7,25	



Il dato medio di 7,25 minuti per telefonata può essere considerato realistico?

# E' importante saper elaborare i dati

Basta veramente solo sommare le modalità ( $x_i$ ) e dividere per N???

**Esempio: Nella tabella sono elencati gli esiti del test di Statistica di 50 studenti**

Esiti modalità ( $x_i$ )	numero studenti frequenze ( $n_i$ )	$x_i \cdot n_i$
0	15	0
15	2	30
17	3	51
18	5	90
22	2	44
26	5	130
27	7	189
28	4	112
39	7	273
	50	919

somma $x_i \cdot n_i$		919
media aritmetica	=919/50	18,4
media aritmetica PRESENTI	=919/(50-15)	26,3

E' corretto affermare che il risultato medio al test è 18,4?  
Il dato più utile è quello medio dei soli presenti (che hanno ottenuto un punteggio) ESCLUDENDO gli esiti=0 perché non valutati

# **Per sintetizzare i dati NON BASTA calcolare solo la Media Aritmetica!!!!**

Effettuando il calcolo della media si ottiene un valore di sintesi (cioè un unico valore). Ma questo valore è REALMENTE rappresentativo del fenomeno???? Occorre valutare la capacità di sintesi della media aritmetica: più è alta la variabilità, meno informativa risulta la media aritmetica



**Misure di tendenza  
centrale: mediana,  
moda, quantili**



**Studiare la variabilità**

# Analisi esplorativa dei dati

Dopo aver effettuato la rilevazione dei dati, si passa alla descrizione delle caratteristiche di interesse.

Per l'analisi esplorativa dei dati sono necessarie tre informazioni:

1. la tendenza centrale o posizione, che indica dove i dati sono concentrati;
2. la variabilità, cioè la «distanza» tra i dati o rispetto ad un valore rappresentativo;
3. la forma della distribuzione, più o meno simmetrica rispetto ad un valore centrale.

I dati vengono elaborati per ricavare gli elementi caratteristici, i tratti essenziali dei fenomeni oggetto di studio, dalla massa delle informazioni.

## Misure di tendenza centrale

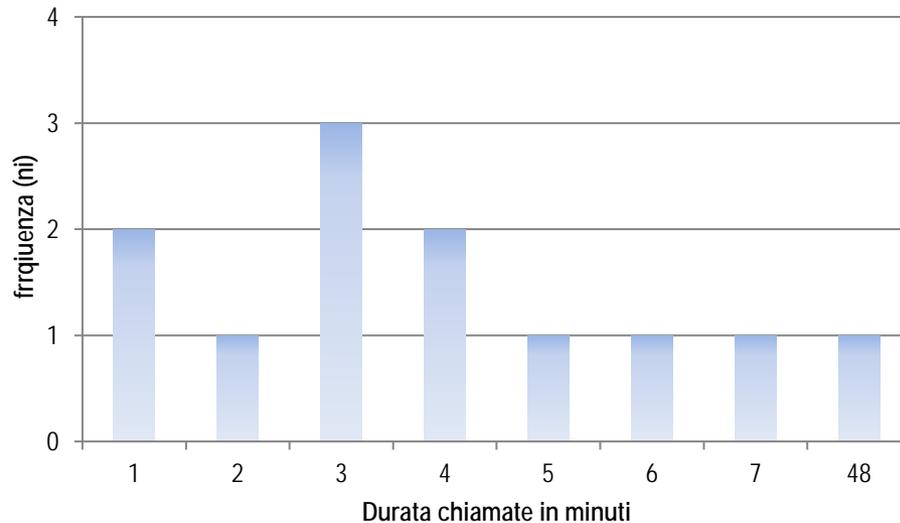
Le misure di tendenza centrale o posizione, descrivono i tratti essenziali dei fenomeni che stiamo studiando. Le più diffuse sono:

1. la **media aritmetica** (semplice o ponderata);
2. la **mediana (Me)**, l'osservazione che occupa la posizione centrale in una distribuzione ordinata in senso crescente, Se  $N$  è dispari,  $Me$  si trova al posto  $(N+1)/2$ ; se  $N$  è pari,  $Me$  è la media delle osservazioni che occupano i posti  $N/2$  e  $N/2+1$ ;
3. la **moda**, l'osservazione con frequenza più alta;
4. i quantili.

# E' importante saper elaborare i dati

Basta veramente solo sommare le modalità ( $x_i$ ) e dividere per N???

**Problema:** Maria vuole sapere quanto tempo trascorre normalmente al telefono. Dal sito web del suo operatore trascrive la durata in minuti di 12 chiamate scelte casualmente (1,7,4,1,2,4,3,48,3,5,3,6).  
Analizzare i dati e calcolare la durata media delle chiamate



Durata telefonate in minuti ( $x_i$ )	frequenza ( $n_i$ )	$x_i \cdot n_i$
1	2	2
2	1	2
3	3	9
4	2	8
5	1	5
6	1	6
7	1	7
48	1	48
Totale	12	87
Media	7,25	
Mediana	3,5	

Il dato mediano di 3,5 minuti per telefonata è più «robusto» perché NON è sensibile ai valori estremi o anomali.

Se la telefonata di 48 minuti fosse stata di 5 minuti la media sarebbe 3,7 minuti mentre la mediana rimarrebbe a 3,5

La moda è 3 minuti

# La variabilità

(la vera ragione per cui esiste la Statistica)

## Il paradosso del pollo di Trilussa

Per ironizzare, la Statistica è spesso associata alla poesia  
*..te tocca un pollo all'anno/ e, se nun entra nelle spese tue/  
T'entra nel la statistica lo stesso/ perché c'è un antro che ne magna  
due...*

Costruiamo una tabella di esempio

Polli (xi)	Consumatori (ni)
0	9
10	1
	N=10

La media aritmetica (cfr. Trilussa) di 1 pollo a testa deriverebbe da  $(0 \cdot 9 + 10 \cdot 1) / 10 = 1$  ma non è rappresentativa perché il 90% dei soggetti consuma fino a 0 polli..... questo è un caso particolare: la massima variabilità. Occorre approfondire....

## Definizione di variabilità

La variabilità è l'attitudine del fenomeno quantitativo ad assumere diverse modalità o meglio la tendenza di ogni singola osservazione ad assumere valori differenti rispetto al valore medio.

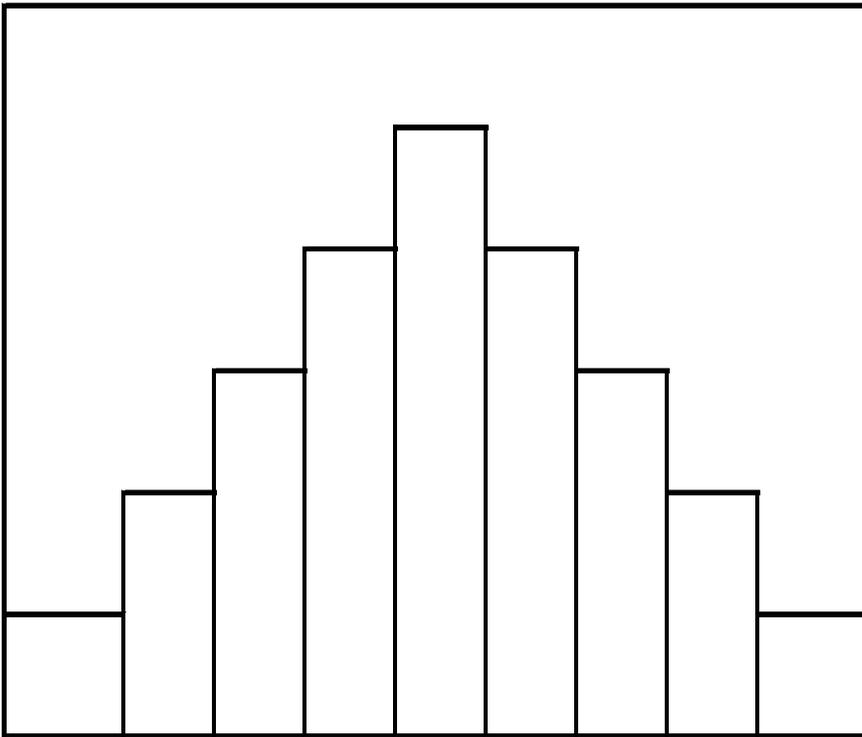
La variabilità cresce all'aumentare della «diversità» tra le osservazioni

Si definisce **scarto** la differenza tra la singola osservazione e la media

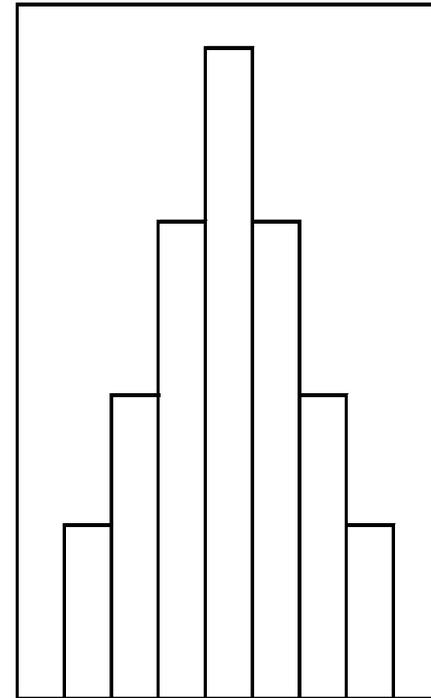
$$scarto = x_i - \mu$$

# Variabilità - Dispersione

A



B



**Le due figure hanno la stessa media ma una dispersione diversa. Qual è la media meno rappresentativa?**

# Indici di variabilità

Misurano la variabilità di una distribuzione di dati.

Si calcolano gli scostamenti medi, che si ottengono determinando gli scarti tra le modalità del carattere e la sua media aritmetica (**dispersione**).

La dispersione si misura con lo **scarto quadratico medio** (o deviazione standard se applichiamo la correzione dividendo per  $n-1$  ).

Gli indici si distinguono in:

1. Indici di variabilità assoluta, che sono espressi nella stessa unità di misura del fenomeno osservato e non sono né valutabili (*è tanta o poca?*) né confrontabili (*è più o meno variabile?*)
2. Indici di variabilità relativa, che prescindono dall'unità di misura e consentono confronti

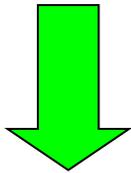
# Indici di variabilità

Sono indici di **variabilità assoluta**

1. Campo di Variazione
2. Differenza Interquartilica
3. Scarto Quadratico Medio
4. Varianza (è il quadrateo dello s.q.m.)
5. Devianza (è il numeratore della varianza)

# Scarto quadratico medio o deviazione standard

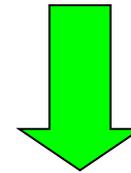
**DISTRIBUZIONE  
SEMPLICE**



$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_N - \mu)^2}{N}} =$$

$$= \sqrt{\frac{\sum_{i=1}^s (x_i - \mu)^2}{N}}$$

**DISTRIBUZIONE  
PONDERATA**



$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 n_1 + (x_2 - \mu)^2 n_2 + \dots + (x_s - \mu)^2 n_s}{N}} =$$

$$= \sqrt{\frac{\sum_{i=1}^s (x_i - \mu)^2 n_i}{N}}$$

## Variabilità - Scarto Quadratico Medio

$X_i$ (anni)	$(x_i - \mu)$	$(x_i - \mu)^2$
20	-2	4
21	-1	1
22	0	0
23	1	1
24	2	4
Totale		10

$$\sigma = \sqrt{\frac{\sum_{i=1}^s (x_i - \mu)^2}{N}} = \sqrt{\frac{10}{5}} = \sqrt{2} = 1,4$$

## Variabilità - Scarto Quadratico Medio - Caso ponderato

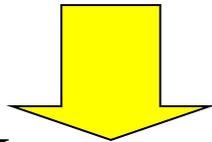
$X_i$	$n_i$	$(x_i - \mu)$	$(x_i - \mu)^2$	$(x_i - \mu)^2 n_i$
20	1	-2	4	4
21	2	-1	1	2
22	4	0	0	0
23	2	1	1	2
24	1	2	4	4
Totale	10			12

$$\sigma = \sqrt{\frac{\sum_{i=1}^s (x_i - \mu)^2 n_i}{N}} = \sqrt{\frac{12}{10}} = \sqrt{1,2} = 1$$

# Variabilità - Varianza

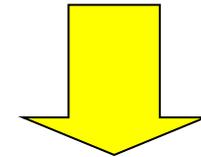
Elevando al quadrato lo scarto quadratico medio si ottiene la **varianza**

DISTRIBUZIONE  
SEMPLICE



$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

DISTRIBUZIONE  
PONDERATA



$$\sigma^2 = \frac{\sum_{i=1}^s (x_i - \mu)^2 n_i}{N}$$

# Indici di variabilità

Sono indici di variabilità **relativa**

1. Indice di variabilità relativo alla media  
(V.a./ $\mu$ ) .Esempio Coefficiente di Variaz
2. Indice di variabilità relativo al massimo  
(variano tra 0 e 1)
3. Concentrazione

# Laboratorio statistico

## Esercitazione su analisi esplorativa dei dati

Tabella 1 - Principali statistiche sui dati delle telefonate di Maria

Durata telefonate in minuti ( $x_i$ )	frequenza ( $n_i$ )	$x_i * n_i$	$x_i - \text{media}$	$(x_i - \text{media})^2$	$((x_i - \text{media})^2) * n_i$
1	2	2	-6,25	39,0625	78,125
2	1	2	-5,25	27,5625	27,5625
3	3	9	-4,25	18,0625	54,1875
4	2	8	-3,25	10,5625	21,125
5	1	5	-2,25	5,0625	5,0625
6	1	6	-1,25	1,5625	1,5625
7	1	7	-0,25	0,0625	0,0625
48	1	48	40,75	1660,5625	1660,5625
Totale	12	87			1848,25
Media	7,25	=MEDIA(num1; num2; .....)			
Mediana	3,5	=MEDIANA(num1; num2; .....)			
Varianza	154,02	=VAR.P(num1; num2; .....)			
Devianza	1848,25				
sqm	12,41	=DEV.ST.P(num1; num2; .....)			



**Formule in Excel**