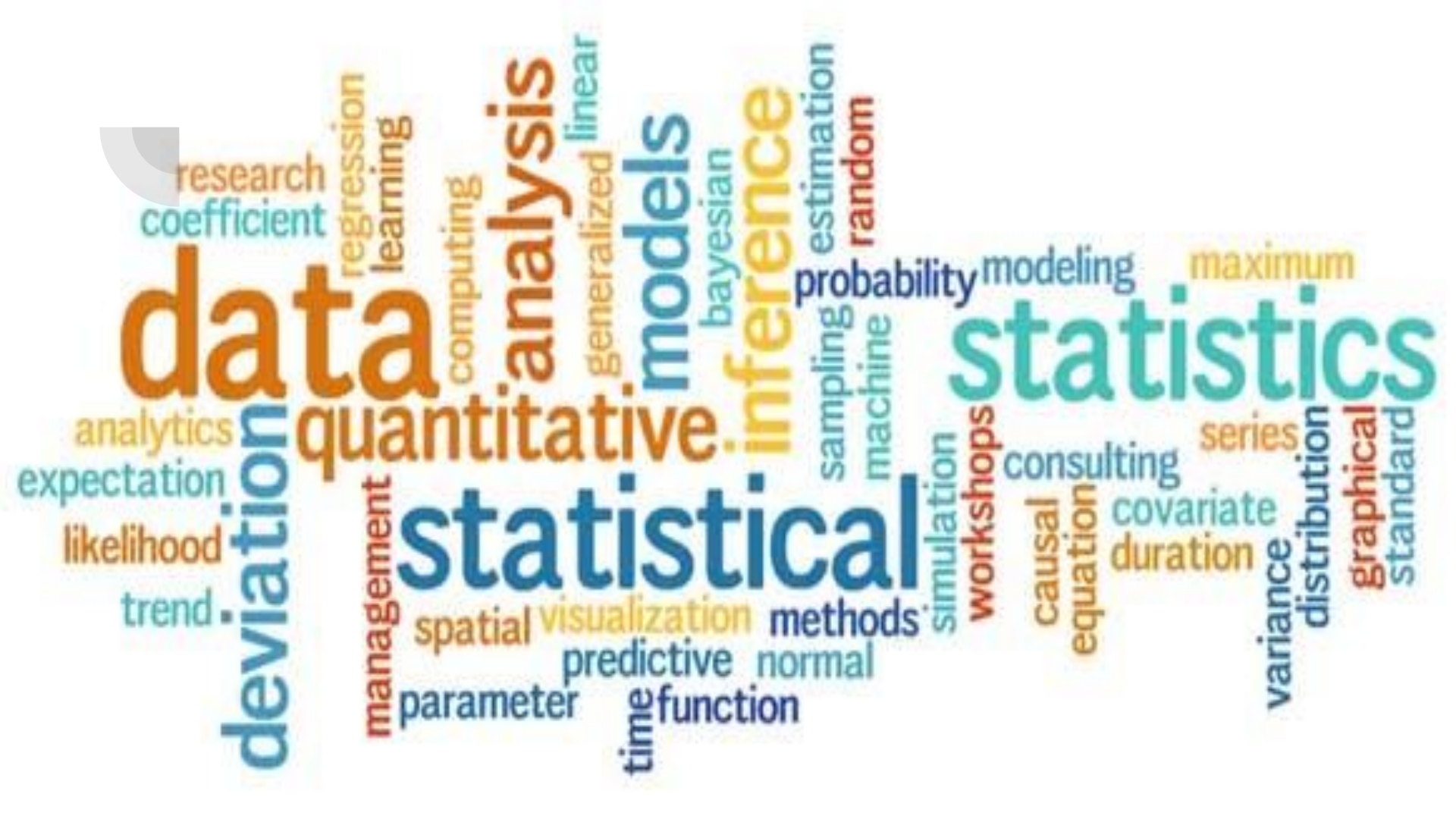
Statistica
A.A. 2019/2020
CdL Scienze Economiche

Prof. Massimiliano Ferrara
Dott. Bruno Antonio Pansera

Lezione n.1


research
coefficient
regression
learning
computing
analysis
linear
generalized
models
bayesian
inference
estimation
random
data
probability
modeling
maximum
statistics
analytics
quantitative
sampling
machine
series
expectation
deviation
statistical
simulation
workshops
consulting
graphical
standard
likelihood
management
covariate
causal
equation
duration
distribution
trend
spatial
visualization
methods
variance
predictive
normal
parameter
time
function

LA STATISTICA

di Trilussa

Sai ched'è la statistica? È na' cosa
che serve pe fà un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che spósa.

Ma pè me la statistica curiosa
è dove c'entra la percentuale,
pè via che, lì,la media è sempre eguale
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perch'è c'è un antro che ne magna due.

HANNO DETTO CHE:

- Le statistiche sono una forma di realizzazione del desiderio, proprio come i sogni- *Jean Baudrillard*
- Non mi fido molto delle statistiche, perché un uomo con la testa nel forno acceso e i piedi nel congelatore statisticamente ha una temperatura media- *Charles Bukowski*
- Se vuoi ispirare fiducia, dai molti dati statistici. Non importa che siano esatti, neppure che siano comprensibili. Basta che siano in quantità sufficiente- *Lewis Carroll*
- Le sole statistiche di cui ci possiamo fidare sono quelle che noi abbiamo falsificato- *Winston Churchill*
- Lo statistico è uno che fa un calcolo giusto partendo da premesse dubbie per arrivare a un risultato sbagliato- *Jean Delacour*
- Certo, certissimo, anzi probabile- *Ennio Flaiano*
- La teoria delle probabilità in fondo non è altro che buon senso ridotto a calcolo- *Simon de Laplace*
- La morte di una persona è una tragedia, la morte di milioni è una statistica- *Josif Stalin*
- Il dubbio non è piacevole, ma la certezza è ridicola- *Voltaire*

Fonte: Istat

La statistica nella storia

Rodolfo de Cristofaro scrive nella rivista STATISTICA, anno LXII, n. 2, 2002 un saggio sulla STORIA DEL PENSIERO STATISTICO CON ALCUNE OSSERVAZIONI SULL'INSEGNAMENTO DELLA STATISTICA dove indica alcune date da ricordare in merito alla nascita della statistica:

- Nel 1906, R. Benini definiva la "statistica" come un metodo particolare di trattazione dei "fenomeni collettivi", intendendo per collettivi "quei fenomeni suscettivi di variare senza regola assegnabile a tutto rigore". Un altro termine molto usato era quello di "fenomeni di massa"; in altre parole, fenomeni che manifestano una regolarità solo per grandi masse di osservazioni; altrimenti detti: tipici, indeterministici, casuali, statistici.
- La scoperta delle "regolarità statistiche" per grandi masse di osservazioni risale al Seicento. In particolare, nel 1662 J. Graunt pubblicò un saggio, nel quale raggruppò i nati e i morti della città di Londra in gruppi omogenei, scoprendo delle sorprendenti regolarità sul rapporto dei sessi alla nascita, sulla mortalità e così via
- In realtà, dalla Cronaca del Villani, pubblicata nel 1346, risulta che il sacrestano del battistero di Firenze aveva già scoperto l'equilibrato rapporto dei sessi alla nascita con una leggera prevalenza di maschi sulle femmine.

La statistica nella storia

- Riferimenti all'obbligo di "contarsi" si trovano numerosi nella Bibbia (libro dei Numeri). Preoccupazioni censuarie delle autorita politiche per fini militari e fiscali sono comuni all'impero cinese come a quello romano. Il Concilio di Trento (1563) emanò il decreto che obbligava i parroci alla tenuta di registri ordinati per battesimi e matrimoni e dopo mezzo secolo l'obbligo si estese alla registrazione dei morti.
- Con la nascita dei grandi Stati europei, si attribuisce all'analisi statistica dei fenomeni collettivi un interesse pubblico che spinge progressivamente le Nazioni a dotarsi di Istituti "centrali" di Statistica (in Italia, l'ISTAT) deputati per legge alla raccolta, organizzazione e diffusione di dati su popolazione, risorse economiche, etc

La statistica nella storia

I successi conseguiti nella "fisica" avevano fatto ritenere, secondo le parole di Galileo (il padre della scienza sperimentale), che la natura fosse "un libro scritto con un linguaggio matematico". Per questo motivo, per molto tempo è stata fatta una netta distinzione tra le cosìdette "scienze esatte" e tutte le altre scienze. La statistica era considerata appunto una scienza "povera", occupandosi di quei fenomeni non regolati da quelle che un tempo si pensava fossero le "immutabili leggi della natura".

In Inghilterra, lo sviluppo della statistica fu favorito da quel sano empirismo inglese, originato da F. Bacon, che si proponeva di edificare una scienza sperimentale, in polemica con l'aristotelismo e la tradizione alchemica. Un'istanza empirica che, nel Novum Organum del 1620, è concepita da Bacon come l'incontro tra la natura delle cose e la mente dell'uomo (non un'adesione totale e acritica ai dati accidentali dei sensi, senza alcun intervento della ragione umana, come talvolta qualcuno sostiene o intende fare credere).

La statistica nella storia

Nel corso dell'Ottocento, il pensiero statistico si è arricchito di alcuni importanti contributi, legati soprattutto ai nomi di *C.F. Gauss (1777-1855), P.S. de Laplace (1749-1827).*

- Gauss è stato il principale artefice della "teoria degli errori di misura", che può essere considerata una delle più importanti radici storiche della statistica. Secondo alcuni, anzi, gran parte della statistica moderna altro non sarebbe che una teoria degli errori mascherata (nel senso che si utilizzano le stesse ipotesi e gli stessi strumenti di quella teoria con riferimento a una classe più generale di fenomeni). In particolare, a Gauss è dovuta la "curva degli errori accidentali" (detta normale o gaussiana); oltre al "metodo dei minimi quadrati", che si è rivelato essenziale nello studio delle relazioni tra due o più variabili.
- A sua volta, Laplace è stato l'autore del primo importante trattato di "calcolo delle probabilità", in cui si trovano enunciati e dimostrati molti teoremi e nel quale si pongono i fondamenti della così detta "statistica induttiva", utilizzando sia una formula sulla "probabilità delle cause", scoperta nel Settecento dal reverendo T. Bayes (1702-1761), sia altre tecniche di verifica delle ipotesi, già anticipate dai Bernoulli nel Settecento.

La statistica nella storia

Il carattere interdisciplinare della statistica tardò comunque ad affermarsi. Per molto tempo, essa fu confusa con la demografia. Si temeva, infatti, che, perdendo il riferimento a dei fenomeni reali, la statistica non potesse più essere considerata una scienza. Tuttavia, è proprio agli inizi del Novecento che viene esplicitamente riconosciuto che la statistica è una scienza "non soltanto sociale, perché il suo campo è molto più largo" (A. Kaufmann, 1913) e viene affermata la sua autonomia come disciplina metodologica (come abbiamo visto all'inizio, citando Benini).

Secondo *K. Pearson (1857-1936),* l'oggetto di studio della statistica è una moltitudine - ora detta popolazione - di certi oggetti, piuttosto che gli individui di questa popolazione. In altre parole, l'oggetto di studio della statistica sono le "popolazioni" (i "collettivi", secondo una terminologia di origine tedesca) da analizzare mediante lo strumento matematico, con finalità prevalentemente conoscitive

La statistica nella storia

Proprio agli inizi del Novecento, la Fisica riconobbe il carattere statistico dei fenomeni oggetto del proprio studio. Come ha scritto M. Plank (1933), "In ogni legge fisica, anche nella gravitazione e nell'attrazione elettrica, c'è un nucleo di natura statistica, si tratta sempre di leggi di probabilità, basate su valori medi di numerose osservazioni dello stesso tipo e valide nei singoli casi in modo approssimato".

Come ha scritto I. Scardovi (1980, pag. 5): "la scoperta della variabilità naturale, del suo ruolo e della sua genesi è la chiave metodologica del pensiero scientifico venuto con la crisi della Weltanschauung deterministica e con l'imporsi - anche nelle "scienze della materia" - di un'immagine statistica della realtà. Momento sperimentale e momento statistico, più che a distinguere metodologie diverse, valgono ormai a denotare fasi e metodi della ricerca, criteri e tecniche di controllo delle ipotesi."

Introduzione alla statistica

Per quale motivo si effettuano analisi statistiche?

QUASI TUTTO È SOGGETTO A VARIAZIONI: Viviamo in un mondo variabile, ma nell'ambito della variabilità che osserviamo si possono riscontrare quadri prevedibili. Usiamo la statistica per individuare e analizzare questi quadri.

La variabilità naturale può complicare l'individuazione di quadri generali. Per esempio, gli studiosi hanno stabilito che il fumo aumenta il rischio di sviluppare un tumore ai polmoni. Sappiamo, tuttavia, che non tutti i fumatori si ammaleranno di tumore ai polmoni e che non tutti i non fumatori non svilupperanno un tumore ai polmoni. Pertanto, confrontando soltanto un fumatore con un non fumatore, rischiamo di trarre conclusioni errate.

La statistica aiuta a individuare quadri generali anche quando la natura non sempre segue questi quadri generali

Introduzione alla statistica

Falsi positivi e di falsi negativi

Nel corso dell'ultimo secolo il pianeta Terra si stava riscaldando. Gli ecologi stanno cercando di comprendere se le popolazioni di piante e di animali siano state in qualche modo interessate da questo riscaldamento globale. Se disponiamo di informazioni a lungo termine sulla distribuzione delle specie e sulla temperatura in alcune aree del pianeta, allora possiamo verificare se gli spostamenti geografici delle specie coincidono con i cambiamenti climatici. Informazioni di questo tipo, tuttavia, possono essere estremamente complesse. Senza metodi statistici appropriati potremmo non essere in grado di determinare il reale impatto della temperatura sulla distribuzione delle specie, oppure potremmo pensare che esista una correlazione dove in realtà questa non c'è.

La statistica aiuta a non trarre conclusioni errate.

Introduzione alla statistica

- Statistica: raccolta di metodi e strumenti matematici atti ad organizzare una o più serie di dati che descrivono una categoria di fatti
- È la scienza che studia i fenomeni collettivi o di massa.
- La statistica insegna a individuare i modi in cui un fenomeno si manifesta, a descriverlo sinteticamente, e a trarne da esso conclusioni più generali di fenomeni più ampi.

Introduzione alla statistica

In che modo la statistica aiuta a comprendere il mondo naturale?

- Definizione del piano sperimentale
- Raccolta dei dati
- Organizzazione e visualizzazione dei dati
- Riassunto dei dati
- Statistica inferenziale (deduttiva)

Introduzione alla statistica

Definizione del piano sperimentale

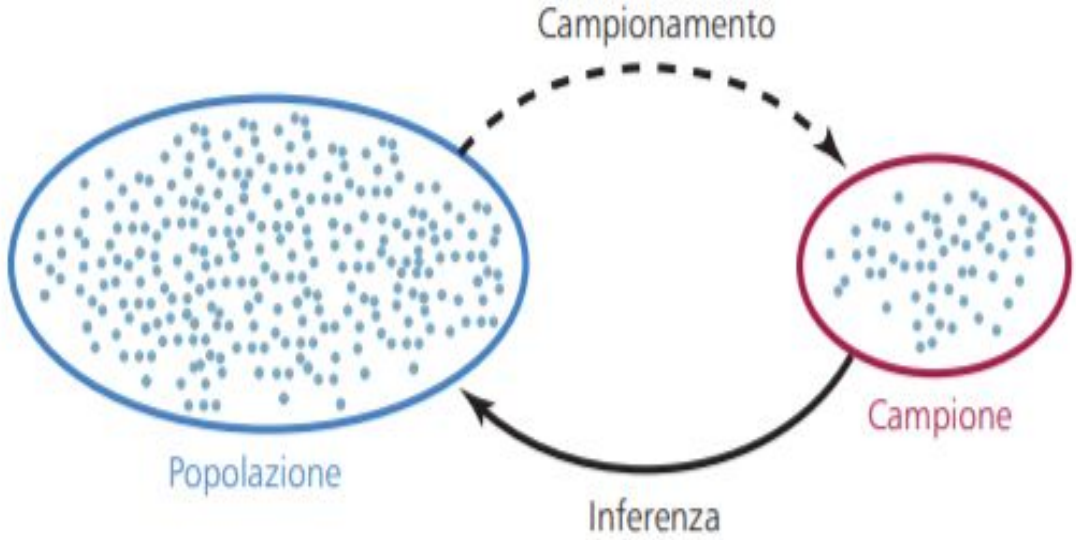
Usiamo la statistica come guida per definire il piano sperimentale in modo da ottenere il giusto tipo di dati.

Prima di allestire un esperimento ricorriamo alla statistica per determinare la mole di dati necessari per verificare la nostra idea e per evitare che fattori estranei possano indurci in errore. Supponiamo di voler condurre un esperimento sui fertilizzanti per verificare che l'azoto incrementa la crescita delle piante. Se includiamo un numero insufficiente di piante, potremmo non essere in grado di determinare se l'azoto ha un effetto sulla crescita delle piante e l'esperimento sarebbe inutile. D'altra parte, lo studio di un numero di piante troppo elevato costituirebbe uno spreco di tempo e di risorse.

Introduzione alla statistica

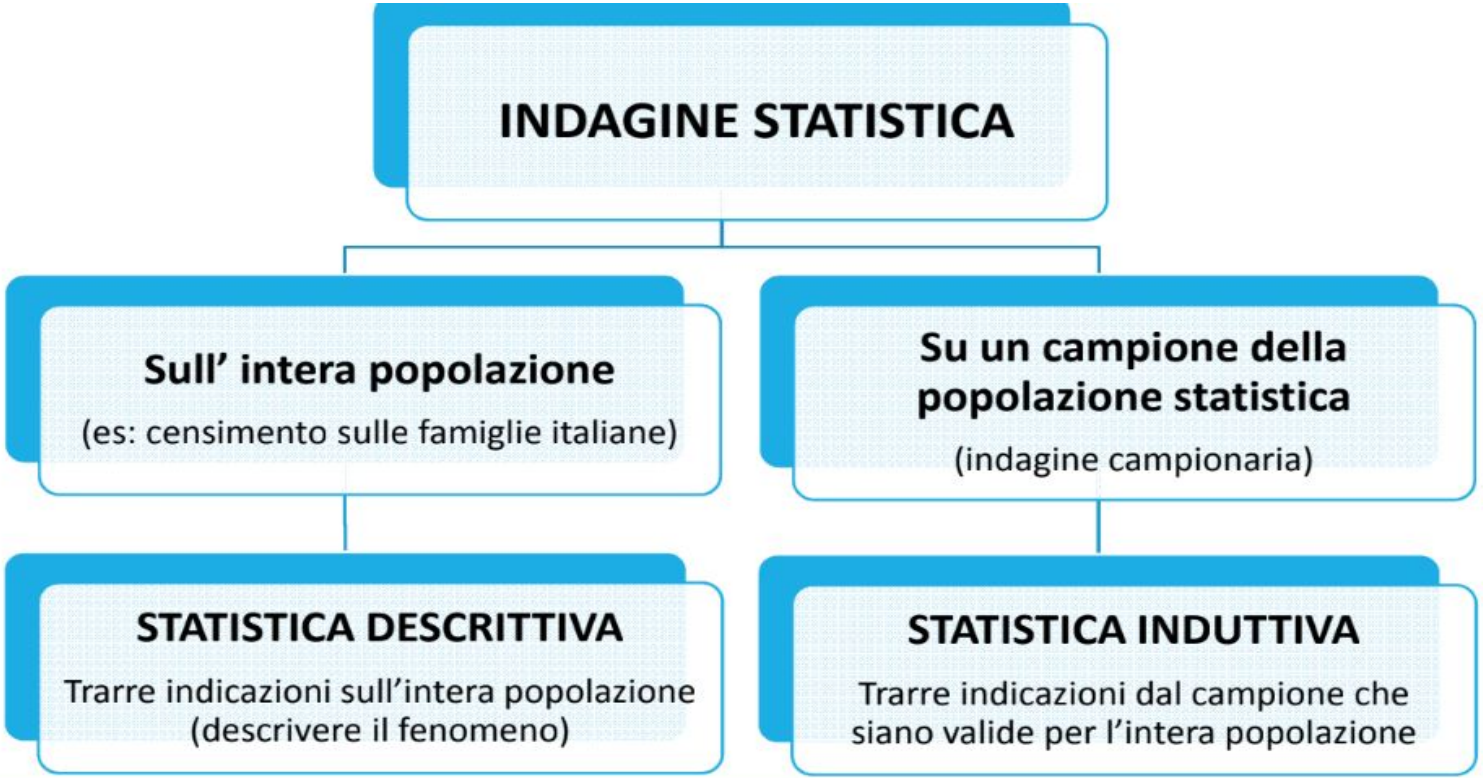
Raccolta dei dati

Gli studiosi raccolgono campioni rappresentativi di una popolazione, applicano le statistiche descrittive per caratterizzare i campioni raccolti e ricorrono poi alla statistica inferenziale per trarre conclusioni relative alla popolazione originale.



Introduzione alla statistica

Raccolta dei dati



Introduzione alla statistica

Raccolta dei dati

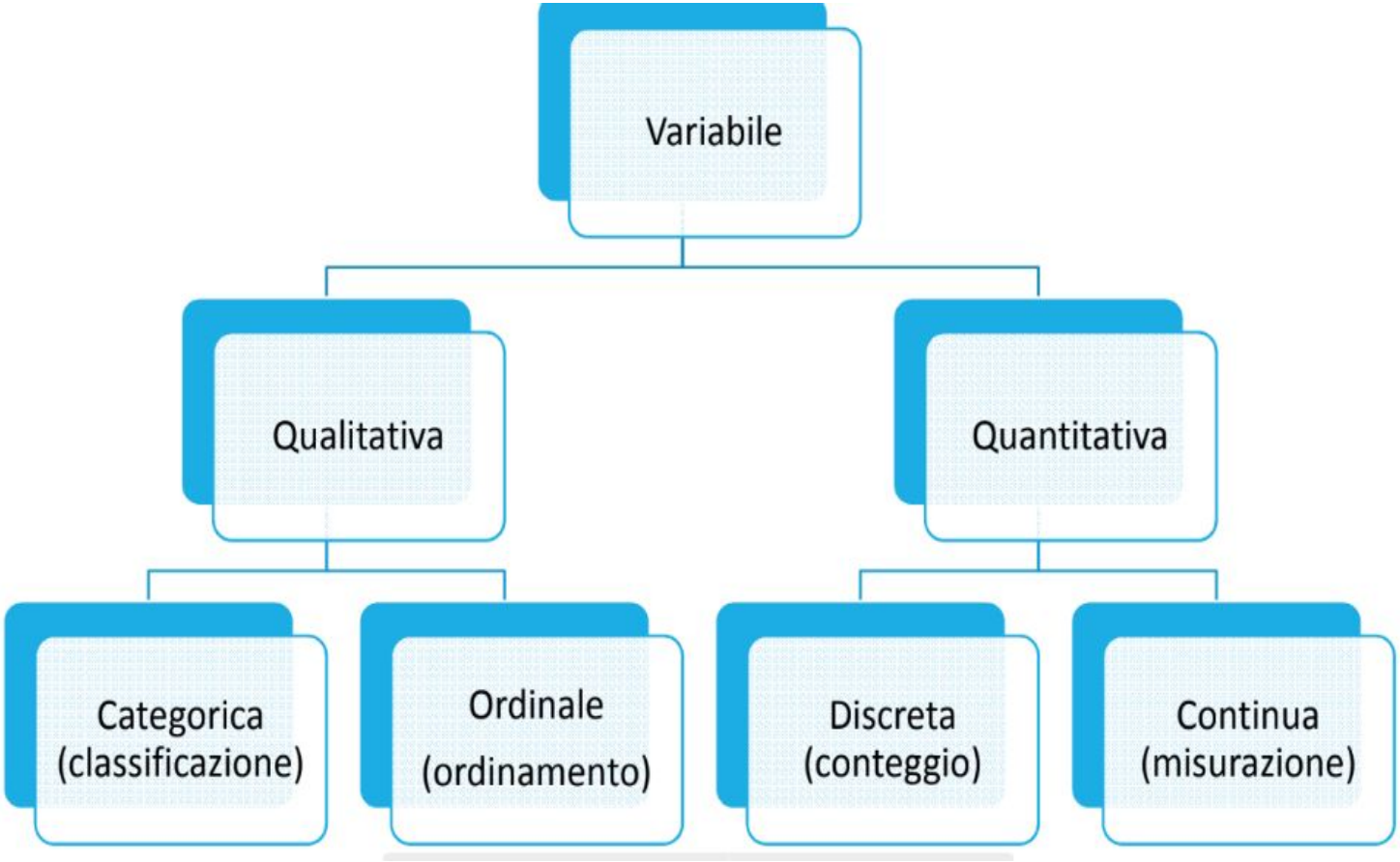
- **Popolazione statistica:** insieme degli elementi a cui si riferisce l'indagine statistica
- **Unità statistica:** ogni elemento della popolazione statistica, la minima unità della quale si raccolgono i dati
- **Campione statistico (sample):** un qualsiasi insieme di unità statistiche prese da tutta la popolazione. Un campione è dunque un sottoinsieme di misurazioni selezionate dalla popolazione

Raccolta dei Dati

Variabili

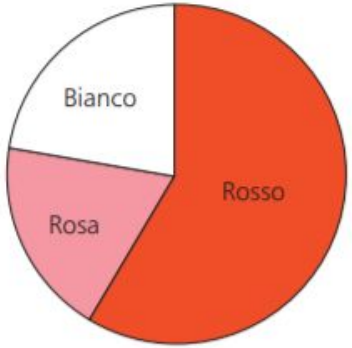
In statistica si usa il termine variabile per indicare un carattere misurabile di un individuo o di un sistema. Alcune variabili sono espresse in scala metrica, come la temperatura massima giornaliera (un valore numerico basato sulla precisione del termometro utilizzato) oppure in valore numerico (un numero intero: 0, 1, 2, 3, ...). Questi valori prendono il nome di variabili quantitative. Le variabili quantitative che esprimono soltanto numeri interi vengono dette variabili distinte, mentre le variabili che possono comprendere anche valori frazionari vengono indicate come variabili continue. Altre variabili hanno come valore determinate categorie, come il gruppo sanguigno nell'uomo. Le variabili di questo tipo vengono indicate come variabili qualitative. Le variabili qualitative dotate di un ordine naturale, come il voto finale in Statistica (da 0 a 30), vengono definite variabili ordinali.

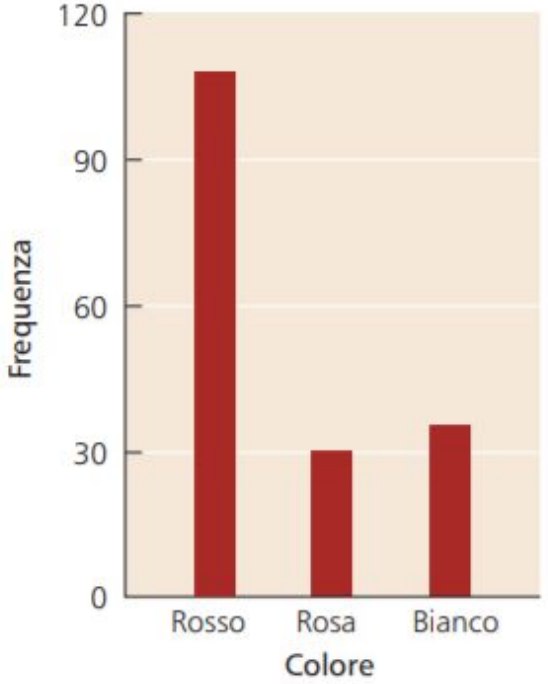
Raccolta dei Dati



Organizzazione e visualizzazione dei dati

Tabella B1 I colori della poinsettia		
COLORE	**FREQUENZA**	**PROPORZIONE**
Rosso	108	0,59
Rosa	34	0,19
Bianco	40	0,22
Totale	**182**	**1,0**





Organizzazione e Visualizzazione dei Dati

Ogni riga rappresenta un'*unità statistica*

Ogni colonna rappresenta una *variabile*

N.	Sesso	Titolo di studio	Età	Peso	N. Ricoveri
1	M	Licenza media inferiore	36	65	3
2	F	Laurea	45	70	1
…	…	…	…	…	…
N	F	Diploma	60	55	6

Riassunto dei Dati

Una statistica è una quantità numerica calcolata da dati, mentre le statistiche descrittive corrispondono a quantità che descrivono quadri generali sotto forma di dati. Le statistiche descrittive permettono di paragonare direttamente diversi insiemi di dati e di comunicare in maniera concisa le caratteristiche dei dati raccolti.

Descrizione Dei Dati Qualitativi

Le variabili qualitative vengono tipicamente descritte come proporzioni. In altre parole, si allestiscono tabelle con le proporzioni delle osservazioni per ciascuna categoria. Per esempio, la terza colonna nella Tabella iporta le proporzioni tra piante di poinsettia per ogni categoria di colore e il diagramma a torta riporta una loro rappresentazione grafica.

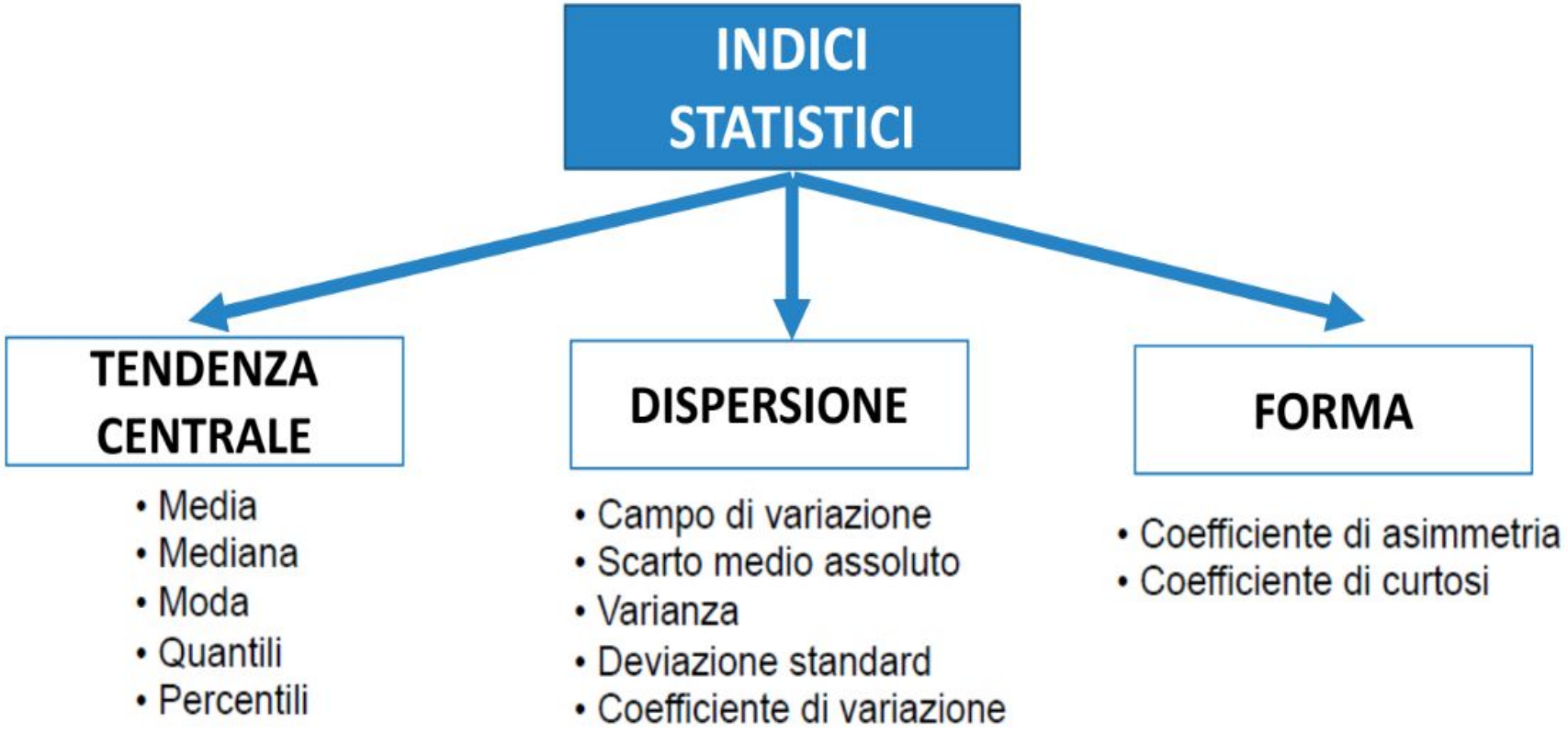
Riassunto dei Dati

Descrizione Dei Dati Quantitativi

Per i dati quantitativi si inizia spesso calcolando il **valore medio** o la **media** del campione. Questo termine piuttosto familiare corrisponde semplicemente alla somma di tutti i valori del campione diviso per il numero di osservazioni effettuate sul campione stesso

Questi valori quantitativi vengono indicati complessivamente come misurazioni del centro. Altre misure del centro comunemente usate sono la **mediana** – il valore che si trova letteralmente al centro del campione di dati – e la **norma** (o **moda**) – ovvero il valore caratterizzato dalla massima frequenza.

Indici Statistici



Tendenza Centrale: Media, moda e mediana

In un'indagine statistica, dopo aver tabulato e rappresentato graficamente i dati relativi ad un fenomeno, occorre sintetizzare la molteplicità di informazioni raccolte, analizzarle ed effettuare dei confronti con fenomeni analoghi. Il primo passo che si compie è, solitamente, l'individuazione dei valori medi statistici, in quanto essi hanno la caratteristica di rappresentare tutto l'insieme dei dati e di essere compresi tra il più piccolo ed il più grande dei valori raccolti.

Esistono vari tipi di medie e quelle più utilizzate sono la media aritmetica, la moda e la mediana. Esse hanno delle caratteristiche diverse tra cui la più evidente è che la media aritmetica è una media di calcolo mentre la moda e la mediana sono medie di posizione, come si vedrà mediante opportune esemplificazioni.

Tendenza Centrale: Media

Definizione (Chisini)

Data una variabile X si definisce **media** il valore $\bar{x}$, intermedio tra il min ed il max delle modalità x_i, $i = 1, n$, che, rispetto ad una funzione sintetica delle osservazioni f ne lascia inalterato il valore:

$$f(x_1, x_2, \ldots, x_n) = f(\bar{x}, \bar{x}, \ldots, \bar{x})$$

Tendenza Centrale: Media

Media di una popolazione: somma di tutti i valori delle variabili della popolazione diviso il numero di unità della popolazione (N)

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

Dove:
-N = numero elementi popolazione
-*Xi* =i-esima osservazione della variabile Xi

Media di un campione: somma di tutti i valori delle variabili di un sottoinsieme della popolazione diviso il numero di unità di tale campione (*n)*

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Tendenza Centrale: Media- Esempio

Dato il seguente set di misurazioni di livello di espressione dei geni:

55.20	18.06	28.16	44.14	61.61	4.88	180.29	399.11	97.47	56.89	271.95	365.29	807.80

Media della popolazione:

$$\mu = \frac{\sum_{i=1}^{13} 55.20 + 18.06 + 28.16 + 44.14 + 61.61 + \ldots + \ldots + 807.80}{13} = \frac{2390{,}85}{13} = 183.9115$$

Media del campione (55.20; 18.06; 28.16; 44.14):

$$\bar{X} = \frac{55.20 + 18.06 + 28.16 + 44.14}{4} = \frac{145.56}{4} = 36.39$$

La media di qualsiasi campione $\bar{X}$ può essere molto diversa da quella dell'intera popolazione μ .

Più è numeroso il campione, più la media del campione sarà vicina a quella della popolazione.

Tendenza Centrale: Media- Proprietà

- μ è sempre compresa tra il min ed il max delle modalità della variabile (*internalità*);
- la somma degli scarti da μ, $x_i - \mu$, $i = 1, 2, \ldots, n$ è nulla (la media costituisce il baricentro di una distribuzione di frequenza)
- se X ha media aritmetica μ, la variabile $a + bX$, $a, b \in \mathbb{R}$ ha media aritmetica $a + b\mu$ (*linearità*)
- se una popolazione è suddivisa in h sottogruppi di numerosità $n_1, n_2, \ldots, n_h$ al cui interno la variabile X presenta medie $\mu_1, \mu_2, \ldots, \mu_h$, la media complessiva è la media aritmetica delle medie, ciascuna con la propria frequenza assoluta (*associatività*):

$$\mu = \frac{\mu_1 n_1 + \mu_2 n_2 + \ldots + \mu_h n_h}{n_1 + n_2 + \ldots + n_h}.$$

- μ è l'unico valore che rende minima la somma degli scarti al quadrato (proprietà caratterizzante la m.a.)

Tendenza Centrale: Media Ponderata

Media ponderata di una popolazione: si assegna ad ogni variabile un peso; si sommano tutti i valori delle variabili, moltiplicate per il peso, e si divide il numero ottenuto per la somma dei pesi

$$\mu = \frac{\sum_{i=1}^{N} p_i X_i}{\sum_{i=1}^{N} p_i}$$

Tendenza Centrale: Media aritmetica di una distribuzione in classi

Il calcolo della media aritmetica di una distribuzione in classi richiede un procedimento più laborioso, in quanto è necessario trovare, preliminarmente, per ciascuna classe, il corrispondente valore centrale. Successivamente, si moltiplica ciascun valore centrale per la rispettiva frequenza assoluta; i prodotti ottenuti si addizionano ed il risultato si divide per il totale delle frequenze.

Tendenza Centrale: Media aritmetica di una distribuzione in classi

Classe di altezze	Freq. assoluta f_i	Valore centrale della classe	Prodotto $f_i * x_i$
151-155	4	(151+155)/2=153	4*153= 612
156-160	9	(156+160)/2=158	9*158= 1422
161-165	15	(161+165)/2=163	15*163= 2445
166-170	7	(166+170)/2=168	7*168= 1176
171-175	8	(171+175):2=173	8*173= 1384
176-180	3	(176+180)/2=178	3*178= 534
181-185	3	(181+185)/2=183	3*183= 549
186-190	1	(186+190)/2=188	1*188= 188
Totale	50		8310

Tendenza Centrale: Media aritmetica di una distribuzione in classi

La media aritmetica di una distribuzione in classi si calcola addizionando i prodotti delle frequenze assolute f_i per i corrispondenti valori centrali x_i di ciascuna classe e dividendo la somma ottenuta per il totale delle frequenze.

$$X_m = (4*153+9*158+15*163+7*168+8*173+3*178+3*183+1*188)/ 50 =$$

$$= (612+1422+2445+1176+1384+534+549+188)/ 50 =$$

$$= 8310/50 = 166,2 \text{ cm}$$

è l'altezza media della distribuzione in classi di altezze assegnata.

Tendenza Centrale: Moda

- La moda è il valore più frequente di una distribuzione, o meglio, la modalità più ricorrente della variabile (cioè quelle a cui corrisponde la frequenza più elevata).

962	1005	1003	768	980	965	1030	1005	975	989	955	783	1005

La moda di questo campione è 1005 in quanto compare ben 3 volte.

- Caratteristiche:
  - viene utilizzata solamente a scopi descrittivi, perché **è meno stabile e meno oggettiva delle altre misure di tendenza centrale.**
  - Per individuare la moda di una distribuzione si possono usare gli istogrammi,
  - Può differire nella stessa serie di dati, quando si formano classi di distribuzione (intervalli) con ampiezza differente.
  - Per individuare la moda entro una classe di frequenza, non conoscendo come i dati sono distribuiti, si ricorre all'ipotesi della ripartizione uniforme.

Tendenza Centrale: Distribuzione Unimodali, Bimodali e k-modali

Una distribuzione può presentare più mode:

- **Distribuzioni unimodali:** distribuzioni di frequenza che hanno una sola moda, ossia un solo un punto di massimo (che rappresenta sia il massimo relativo che i massimo assoluto);

- **Distribuzioni bimodali o k-modali**: distribuzioni di frequenza che presentano due o più mode, ossia che hanno due (ok) massimi relativi;
  - *Esempio:* misurando le altezze di un gruppo di giovani in cui la parte maggiore sia formata da femmine e la minore da maschi si ottiene una distribuzione bimodale, con una moda principale ed una secondaria.

Mediana

- La mediana è il valore che occupa la posizione centrale in un insieme ordinato di dati.
- E' una misura robusta, in quanto poco influenzata dalla presenza di dati anomali.
- Caratteristiche:
  - si ricorre al suo uso quando si vuole attenuare l'effetto di valori estremi;
  - in una distribuzione o serie di dati, ogni valore estratto a caso ha la stessa probabilità di essere inferiore o superiore alla mediana.

Mediana

La mediana è una media di posizione e, come la moda, non è influenzata dai valori estremi.

Essa ha la caratteristica di dividere in due parti uguali la successione di dati, pertanto si può definire come quel dato per il quale esistono tanti valori inferiori quanti superiori ad esso. Inoltre, la mediana divide l'istogramma della distribuzione in due aree uguali e, nell'ogiva delle frequenze cumulate essa corrisponde all'ascissa del punto la cui ordinata è 1/2 ovvero il 50%.

Mediana: Definizione

Si definisce **mediana** *Me* il valore della variabile che bipartisce la distribuzione *ordinata* delle modalità. Se $x_1 \leq x_2 \leq \ldots \leq x_n$, allora

$$Me = \begin{cases} \dfrac{x_{n/2} + x_{n/2+1}}{2} & \text{se } n \text{ è pari} \\ x_{(n+1)/2} & \text{se } n \text{ è dispari} \end{cases}$$

Per variabili continue discretizzate e variabili continue, la mediana è il valore x t.c. $F(x) = 1/2$.

Mediana: Calcolo

Per calcolare la mediana di un gruppo di dati, bisogna:

1. disporre i valori in ordine crescente oppure decrescente e contare il numero totale *n* di dati;
2. se il numero (*n*) di dati è dispari, la mediana corrisponde al valore numerico del dato centrale, quello che occupa la posizione *(n+1)/2*;
3. se il numero (*n*) di dati è pari, la mediana è stimata utilizzando i due valori centrali che occupano le posizioni *n/2* e *n/2+1*:
   a. con poche osservazioni, come mediana viene assunta la media aritmetica di queste due osservazioni intermedie;
   b. con molte osservazioni raggruppate in classi, si ricorre talvolta alle proporzioni.

Mediana: Esempio

Consideriamo il seguente campione:

96	78	90	62	73	89	92	84	76	86

1. Ordiniamo i campioni in ordine crescente:

62	73	76	78	**84**	**86**	89	90	92	95

2. Dal momento che il numero di campioni è pari *(n=10)* la mediana è calcolata come la media dei due elementi centrali:

$$mediana = \frac{84 + 86}{2} = 85$$

Mediana: Proprietà

- E' caratterizzata dal fatto di minimizzare la somma degli scarti assoluti $\sum_{i=1}^{n} |x_i - x|$.
- Tiene conto solo dell'ordinamento delle osservazioni, limitandosi a considerare la modalità dell'elemento centrale. E' dunque particolarmente resistente alla presenza di valori atipici, eccezionali, errati (le medie, ovviamente, no!). Ma proprio per questo è estremamente sensibile a modifiche indotte nel corpo centrale della distribuzione.
- Ha un costo computazionale elevato in presenza di un numero elevato di osservazioni (richiede algoritmi di ordinamento efficienti)

Quartili

- I quantili sono una famiglia di misure, a cui appartiene anche la mediana, che si distinguono a seconda del numero di parti uguali in cui suddividono una distribuzione.
- I quartili ripartiscono la distribuzione in 4 parti di pari frequenza, dove ogni parte contiene la stessa frazione di osservazioni:
  - *Il **primo quartile** è definito come il numero q1 per il quale il 25% dei dati statistici è minore o uguale a q1.*
  - *Il **secondo quartile** è definito come il numero q2 per il quale il 50% dei dati statistici è minore o uguale a q2. Il secondo quartile corrisponde alla mediana*
  - *Il **terzo quartile** è definito come un numero q3 per il quale il 75% dei dati statistici è minore o uguale a q3.*

Quartili- Esempio

Studio che esamina i tempi d'attesa al ristorante in un campione di 10 clienti

Dati ordinati:

Q2 = Mediana

58.6 59.0 59.3 59.4 62.7 62.8 63.7 65.4 67.3 68.1

La mediana è pari a 62,75

Si considera la metà inferiore dei dati, ovvero tutti i valori inferiori alla mediana e su questo sottoinsieme di dati si calcola la mediana, il valore trovato è Q1

Q1

58.6 59.0 59.3 59.4 62.7

Si considera la metà superiore dei dati, ovvero tutti i valori superiori della mediana e su questo sottoinsieme di dati si calcola la mediana il valore trovato è Q3

62.8 63.7 65.4 67.3 68.1

Q3