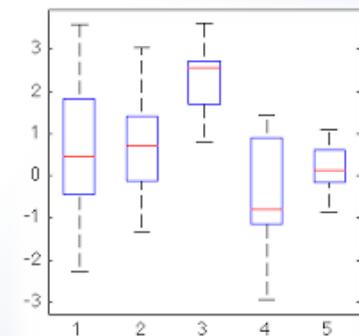
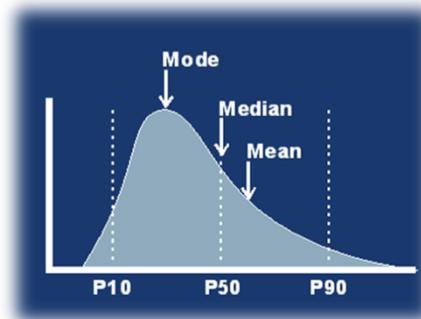




Metodi grafico-numeric per l'analisi statistica di dati geochimici



Principali tipi di dati

Nel discutere i metodi per fornire descrizioni sommarie dei dati, aiuta distinguere i dati in due tipi fondamentali:

```
graph TD; A[Principali tipi di dati] --> B[Dati qualitativi o categorici]; A --> C[Dati quantitativi, numerici o di misura]; B --> D[Quando la caratteristica studiata è classificabile soltanto in categorie e non misurabile numericamente]; C --> E[La caratteristica è misurata in scala numerica e i dati risultanti sono costituiti da un insieme di numeri];
```

Dati **qualitativi** o categorici

Quando la caratteristica studiata è classificabile soltanto in categorie e non misurabile numericamente

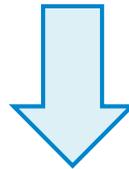
Dati **quantitativi**, numerici o di misura

La caratteristica è misurata in scala numerica e i dati risultanti sono costituiti da un insieme di numeri

Table e grafici per dati quantitativi (numerici)

- Quando si ha un numero elevato di osservazioni (misure) diventa necessario **sintetizzare** i valori, al fine di presentare, analizzare ed interpretare correttamente i risultati.
- Si dividono così dati in **classi** di raggruppamento e si riportano i risultati in forma tabellare, ottenendo la cosiddetta **distribuzione delle frequenze** (assolute, relative, percentuali) delle osservazioni.

Una **distribuzione delle frequenze** è una **tabella** in cui i dati sono organizzati in opportune classi di raggruppamento o categorie.



La tabella sintetizza le principali caratteristiche del dato ma nella sintesi si **perdono le informazioni** che riguardano la singola osservazione (si perde qualcosa nel corso del processo di raggruppamento)

Attenzione nelle fasi di **scelta del numero delle classi** e dell'ampiezza di ciascuna di esse

Scelta del numero delle classi

Il numero opportuno di classi in cui i dati devono essere raggruppati dipende dalla quantità di osservazioni disponibili.

Dimensione campione	Numero intervalli
< 50	5 – 6
50 - 100	6 – 8
> 100	8 – 10

Generalmente il numero di classi dovrebbe variare da un minimo di 5 classi ad un massimo di 15

Per trovare il numero opportuno di intervalli si può utilizzare anche la seguente regola:

Regola di Sturges

$$K = 1 + 3.322(\log_{10} n)$$

n = numero osservazioni

k = numero classi

Costruzione degli intervalli

- Generalmente si richiede che ogni classe abbia la stessa ampiezza, ma non è strettamente necessario.
- Se le classi hanno uguale ampiezza la lunghezza di ciascun intervallo si determina dividendo il **range** dei valori (intervallo dei valori della variabile oggetto di studio) per il **numero di classi** assunto.

Ampiezza intervallo= range/numero di classi

Determinazione degli
estremi delle classi



definire gli estremi delle classi in modo da evitare sovrapposizioni ed attribuzioni ambigue

La distribuzione delle frequenze relative

- Si ottiene rapportando le frequenze assolute al numero totale delle osservazioni; la distribuzione delle frequenze percentuali si ottiene moltiplicando per 100 ciascuna frequenza relativa.

Classi porosità (%)	Frequenza	Frequenza relativa	Frequenza percentuale
20.0 – 25.0	2	0.034	3.4
25.0 – 30.0	13	0.220	22.0
30.0 – 35.0	24	0.407	40.7
35.0 – 40.0	4	0.068	6.8
40.0 – 45.0	11	0.186	18.6
45.0 – 50.0	5	0.085	8.5
Totale	N = 59 osservazioni	1.000	100.0

La distribuzione cumulativa

Classi porosità (%)	Frequenza percentuale cumulativa
20.0	0.0
25.0	3.4
30.0	$25.4 = 3.4 + 22.0$
35.0	$66.1 = 3.4 + 22.0 + 40.7$
40.0	$72.9 = 3.4 + 22.0 + 40.7 + 6.8$
45.0	$91.5 = 3.4 + 22.0 + 40.7 + 6.8 + 18.6$
50.0	$100.0 = 3.4 + 22.0 + 40.7 + 6.8 + 18.6 + 8.5$

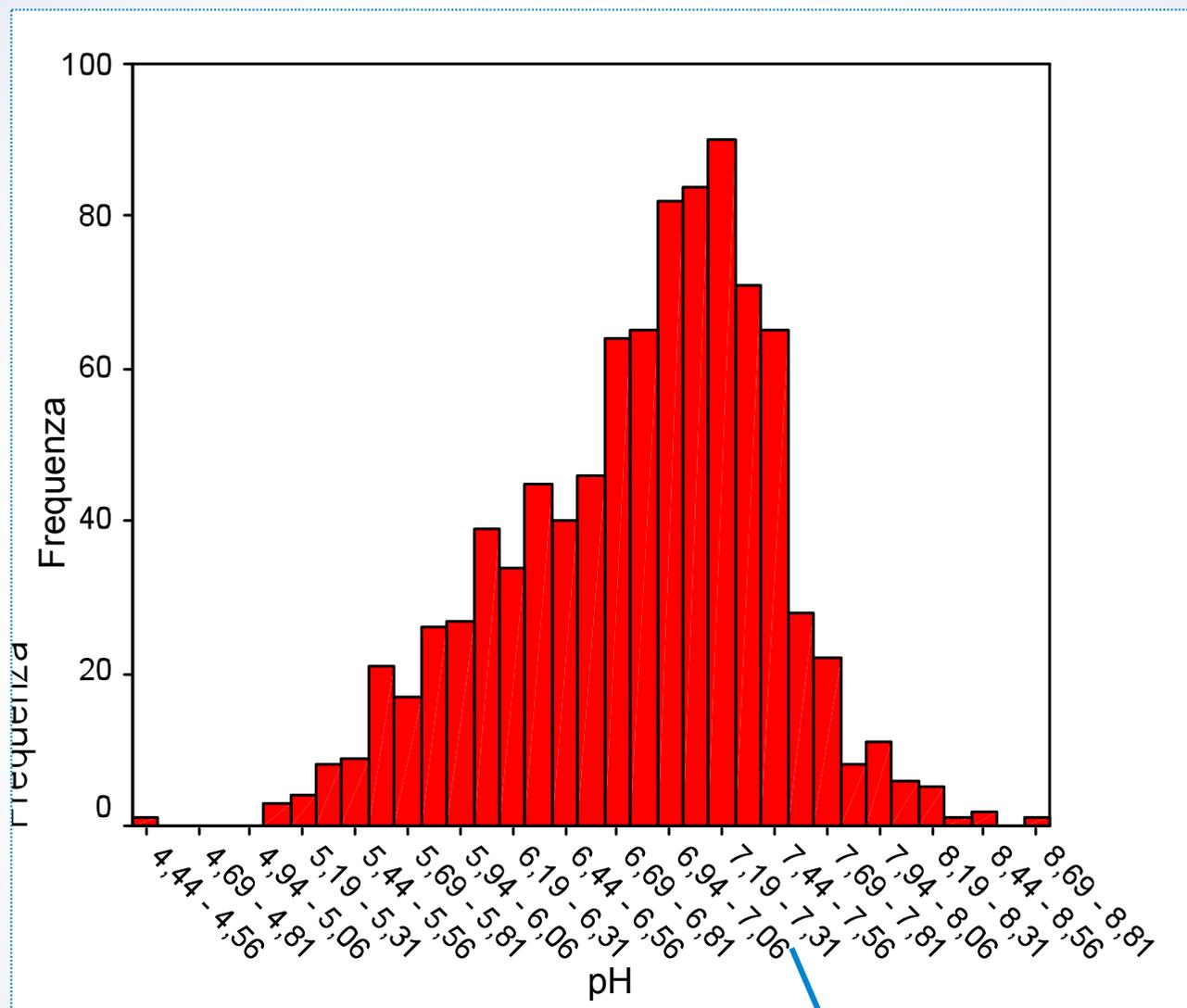
- I valori che compaiono nella prima colonna rappresentano gli [estremi inferiori](#) delle classi di raggruppamento già utilizzate nella costruzione della distribuzione delle frequenze (si è aggiunto l'estremo superiore dell'ultima classe)
- I [valori cumulati](#) sono stati ottenuti sommando le percentuali di osservazioni minori di ciascuno dei suddetti valori estremi (seconda colonna)

L'istogramma

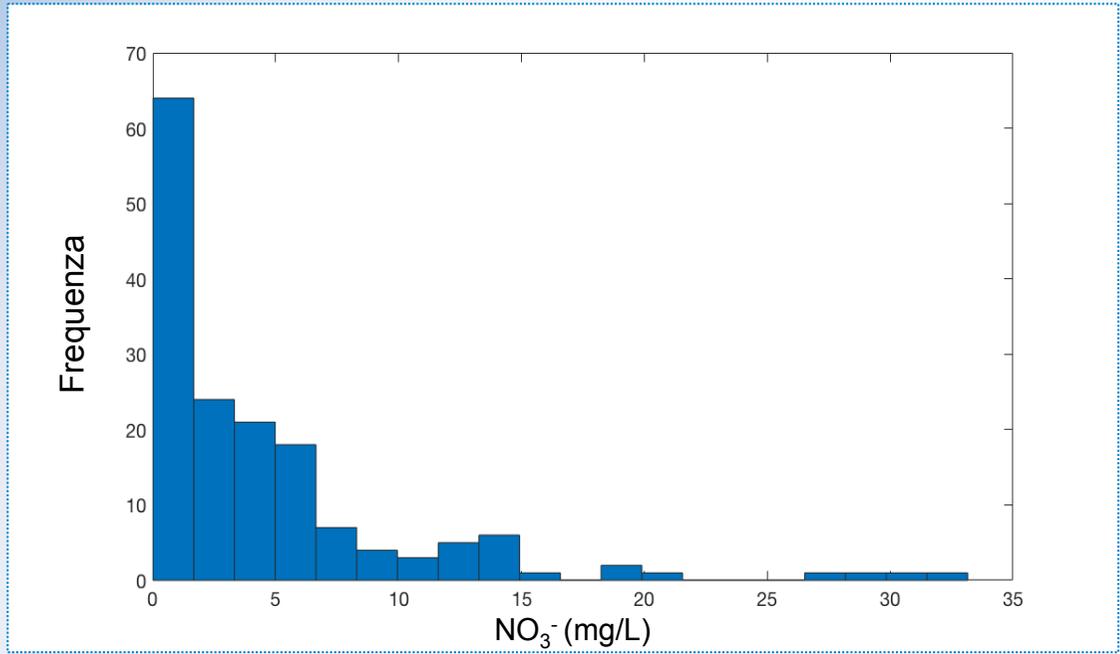
- L'istogramma è un grafico adatto alla rappresentazione di dati numerici che siano stati sintetizzati in tabelle di frequenza (assoluta, relativa o percentuale)
- è un diagramma a **barre verticali** in cui i rettangoli hanno come base gli intervalli in cui sono state raggruppate le osservazioni e come altezza il numero dei dati che cadono in ogni classe.

Asse orizzontale → valori assunti dalla variabile in analisi

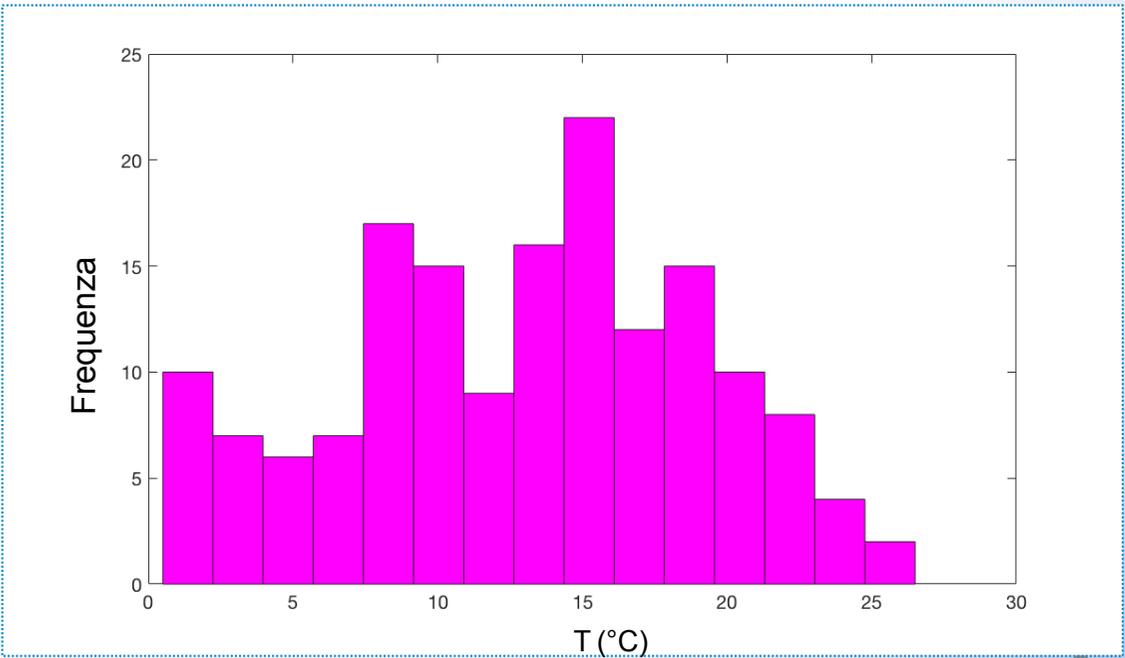
Asse verticale → rappresenta invece il numero, la proporzione o la percentuale di osservazioni che cadono in ciascuna classe



...valori massimo e minimo di ogni classe

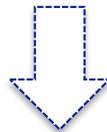


Esempi:
acque superficiali bacino
del fiume Tevere



Sintesi e descrizione dei dati quantitativi

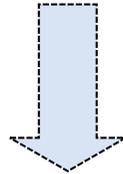
- La rappresentazione dei dati è una fase essenziale della Statistica Descrittiva ma non è esauriente.
- Quando si considerano **dati quantitativi** non è sufficiente presentare adeguatamente i dati, ricavando indicazioni a partire dall'osservazione di tali rappresentazioni.
- Una buona **analisi dei dati** richiede anche che le caratteristiche delle osservazioni siano sintetizzate con **opportune misure** e che tali misure siano interpretate in modo adeguato.



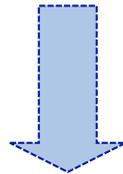
Misure di posizione, variabilità e forma

Le misure di posizione

Nella maggior parte degli insiemi di dati le osservazioni mostrano una tendenza a raggrupparsi attorno ad un valore centrale



risulta quindi in genere possibile selezionare un **valore tipico** per descrivere un intero insieme di dati
si tratta di un valore di **posizione** o di **tendenza centrale**



- media aritmetica
- mediana, moda
- range medio
- media interquartile

La media aritmetica

È la **misura di posizione** più comune. Si calcola dividendo la somma dei valori osservati per il numero totale di misure. Per un insieme di dati che contiene n osservazioni, x_1, x_2, \dots, x_n la media aritmetica si ottiene come:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

- La media è un **punto di equilibrio** tale che le osservazioni più piccole **bilanciano** quelle più grandi
- Il calcolo della media si basa su tutte le n osservazioni dell'insieme di dati
- Poiché il calcolo della media si basa su tutte le osservazioni, tale misura di posizione risulta **influenzata da valori estremi**
- In presenza di valori estremi la media fornisce un quadro **distorto** della struttura dei dati ed è quindi opportuno ricorrere ad altre misure di posizione

La mediana

- La mediana è l'osservazione che, nella **serie ordinata dei dati**, si lascia alla destra il 50% delle osservazioni e a sinistra l'altro 50%
- Quindi il 50% delle osservazioni risulteranno maggiori della mediana e il 50% minori

mediana = osservazione al posto $\frac{n+1}{2}$ nella serie ordinata

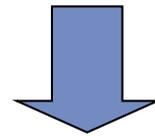
- Se il numero n di osservazioni è **dispari**, la mediana coincide con il valore centrale della serie ordinata
- Se il numero n di osservazioni è **pari**, il valore centrale si trova a metà tra le due osservazioni centrali della serie ordinata

Il calcolo della mediana è influenzato dal numero delle osservazioni, non dalla grandezza dei valori (non è influenzato da valori estremi)

La moda

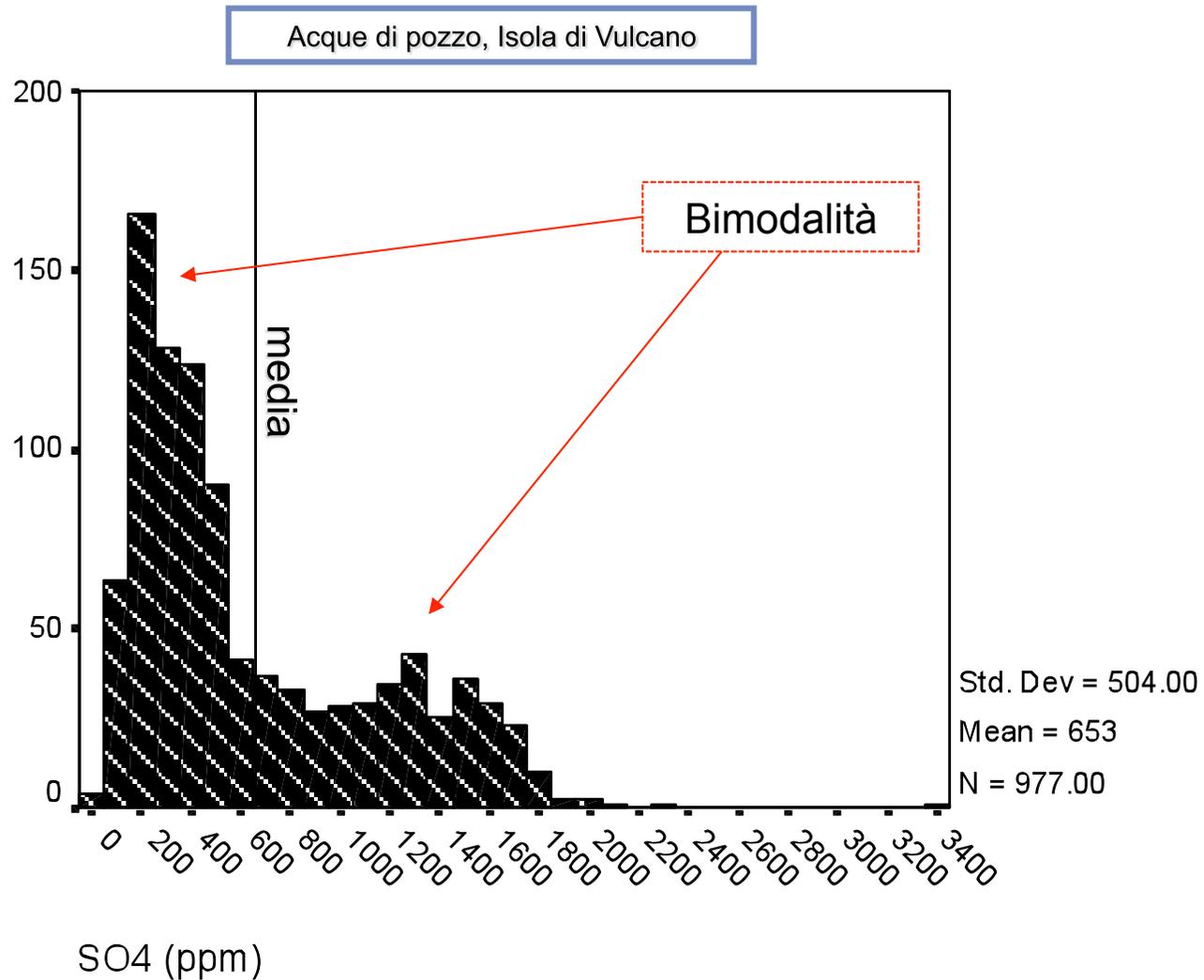
- La moda è **il valore più frequente** di un insieme di dati e la sua determinazione non è influenzata da valori estremi
- si tratta di una misura di posizione usata solo per scopi descrittivi poiché è caratterizzata da maggiore variabilità rispetto alle altre misure di posizione (piccole variazioni in un insieme di dati possono far variare in modo consistente la moda)
- Nella serie ordinata di valori:

10.0 20.6 **28.6 28.6** 29.4 29.5 29.9 30.1 **30.5 30.5** 32.1 32.2 32.4 33.0

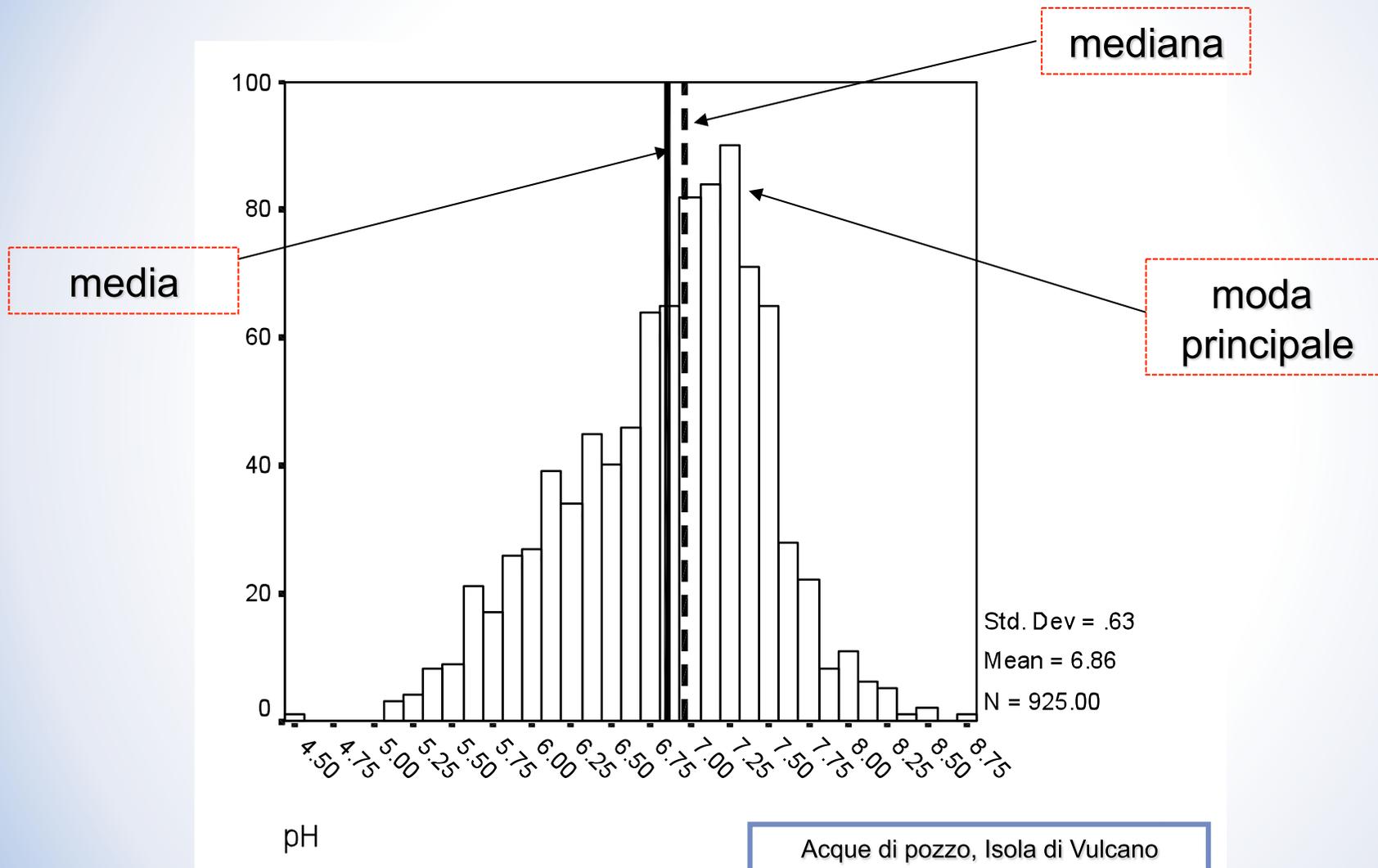


vi sono due valori **“più tipici”** o due mode
un insieme di dati di questo tipo si dice **bimodale**

Istogramma con più mode



Media, moda e mediana: confronto



Range medio

- È dato dalla media tra la più piccola e la più grande delle osservazioni di un insieme di dati
- Si calcola sommando il valore più piccolo e quello più grande e dividendo il risultato per 2:

$$\text{range medio} = \frac{x_{(\text{più piccola})} + x_{(\text{più grande})}}{2}$$

- Il **range medio** è una misura adeguata e semplice per caratterizzare l'intero insieme di dati ma deve essere usato con **precauzione**
- Infatti è una misura che si basa solo sulla più piccola e la più grande delle osservazioni ed è quindi influenzata dalla possibile presenza di valori estremi

I quartili

- Si tratta di un insieme di misure dette, in generale, **quantili**.
- Sono misure di posizione “**non centrale**” di ampio uso e si utilizzano quando si sintetizzano (si descrivono) le caratteristiche di ampi insiemi di dati quantitativi
- Ricordiamo che la **mediana** è un valore che divide a metà la serie ordinata delle osservazioni
- I **quartili** sono misure descrittive che dividono i dati ordinati in quattro parti
- Altri quartili usati di frequente sono i **decili**, che dividono i dati ordinati in dieci parti e i **percentili** che dividono i dati ordinati in cento parti

I quartili

Il primo quartile Q_1 è il valore tale che il 25% delle osservazioni è più piccolo di Q_1 e il 75% è più grande di Q_1

Q_1 = osservazione di posto $(n+1)/4$ nella serie ordinata di valori

Il terzo quartile Q_3 è il valore tale che il 75% delle osservazioni è più piccolo di Q_3 e il 25% è più grande di Q_3

Q_3 = osservazione di posto $3(n+1)/4$ nella serie ordinata di valori

La media interquartile

- È una misura di sintesi che viene utilizzata per evitare i problemi che possono sorgere in presenza di valori estremi
- È data dalla media tra il primo e il terzo quartile dell'insieme di dati

$$\text{media interquartile} = \frac{Q_1 + Q_3}{2}$$

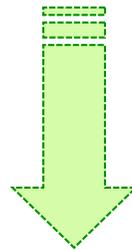
- La media interquartile, in quanto media di Q_1 e Q_3 non è influenzata da valori anomali visto che dal calcolo le misure inferiori a Q_1 e superiori a Q_3 sono escluse

Misure di sintesi come la mediana e la media interquartile, che non sono influenzate da valori anomali (outliers) si dicono “robuste”

Misure di variabilità

La **variabilità** è associata alla dispersione che presentano i dati:

- due insiemi di dati possono differire sia nel valore delle misure di tendenza centrale che nella variabilità intorno a tali statistiche.
- due insiemi di dati possono differire nel valore delle misure di tendenza centrale ma avere una variabilità simile.



- Range
- Range interquartile
- Varianza
- Deviazione standard
- Coefficiente di variazione

Il range

- Il range (o intervallo di variazione) è la differenza (in valore assoluto) tra il valore più grande e quello più piccolo di una serie di osservazioni:

$$x_{\text{massimo}} - x_{\text{minimo}}$$

- Si tratta di una misura della **dispersione totale** dell'insieme di dati
- È una misura molto semplice da calcolare ma il limite consiste nel fatto che non tiene conto di come i dati si distribuiscono effettivamente tra il valore massimo e quello minimo. In molti casi è una misura non adeguata della variabilità

Il range interquartile

- È la differenza tra il terzo ed il primo quartile di un insieme di dati:

$$|Q_3 - Q_1|$$

- è una misura di variabilità che tiene conto di cosa accade ai valori relativi al 50% delle osservazioni che occupano le posizioni centrali; **non è pertanto influenzata da valori estremi**
- sia il range che il range interquartile non tengono conto di come le osservazioni si distribuiscono e si **“addensano”** intorno ad una misura di tendenza centrale come ad esempio la media.

Varianza

- La varianza campionaria è approssimativamente data dalla media dei quadrati degli scarti di ciascuna osservazione dalla media (valore atteso)
- In un campione di n osservazioni x_1, \dots, x_n , la varianza campionaria s^2 è data da:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

Deviazione Standard

- Si definisce scarto quadratico medio (o deviazione standard):

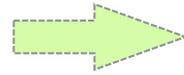
$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

- La deviazione standard corrisponde alla radice quadrata della varianza.

Esempio di calcolo (campione n = 17 osservazioni)

A= [32.2 29.5 29.9 32.4 30.5 30.1 32.1 35.2 10.0 20.6 28.6 30.5 38.0 33.0 29.4 37.1 28.6]

$$\bar{x} = 29.86$$



Media aritmetica

x_i	$(x_i - \bar{x}_{medio})$	$(x_i - \bar{x}_{medio})^2$
32.2	+2.34	5.4756
29.5	-0.36	0.1296
...
28.6	-1.26	1.5876
totale	0	658.5592

- Calcolare le differenze tra ciascuna osservazione e la media
- Elevare al quadrato ciascuna differenza
- Sommare le differenze così ottenute
- dividere il totale per n-1
- Per calcolare s si estrae la radice quadrata della varianza

$$S^2 = 658.55/16 = 41.16$$

$$s = 6.42$$

- Nel calcolo della varianza campionaria le differenze tra ciascuna osservazione e la media sono elevate al quadrato, quindi né la varianza né la deviazione standard possono essere negativi.
- L'unico caso in cui s^2 ed s possono essere uguali a zero è quando non c'è variabilità nei dati, cioè quando tutte le osservazioni nel campione mostrano valori uguali. In tal caso anche il **range** ed il **range interquartile** sono uguali a zero.

I dati quantitativi sono però naturalmente variabili, cioè non costanti, perché ogni fenomeno o processo naturale è sempre caratterizzato da una varietà di valori

Poiché i dati quantitativi sono **variabili** diventa importante sintetizzare i dati non solo per mezzo delle **misure di posizione**, che forniscono una sintesi del valore dei dati, ma anche per mezzo delle **misure di variabilità** che ne sintetizzano la dispersione

L'unità di misura della varianza coincide con il quadrato dell'unità di misura dei dati (ppm al quadrato, % al quadrato, etc.); per tale motivo la principale misura di dispersione è la deviazione standard (scarto quadratico medio), il cui valore è espresso nelle unità di misura originarie dei dati.

Capire la variabilità dei dati...

- Quanto più i dati sono dispersi, tanto maggiori saranno range, range interquartile, varianza e deviazione standard
- Quanto più i dati sono concentrati o omogeni, tanto minori saranno il range, il range interquartile, la varianza e la deviazione standard
- Se le osservazioni sono tutte uguali (non c'è variabilità) il range, il range interquartile, la varianza e lo scarto quadratico medio sono tutti uguali a zero
- Nessuna delle misure di variabilità (range, range interquartile, varianza e scarto quadratico medio) può essere negativa.

Coefficiente di variazione

Il coefficiente di variazione è uguale alla deviazione standard divisa per la media aritmetica, moltiplicato per 100 (è espresso in %).

Deviazione
standard

$$CV = \left(\frac{s}{|\bar{x}|} \right) \times 100$$

Media (valore
assoluto)

Il **coefficiente di variazione CV** misura la dispersione nell'insieme dei dati relativamente alla media:

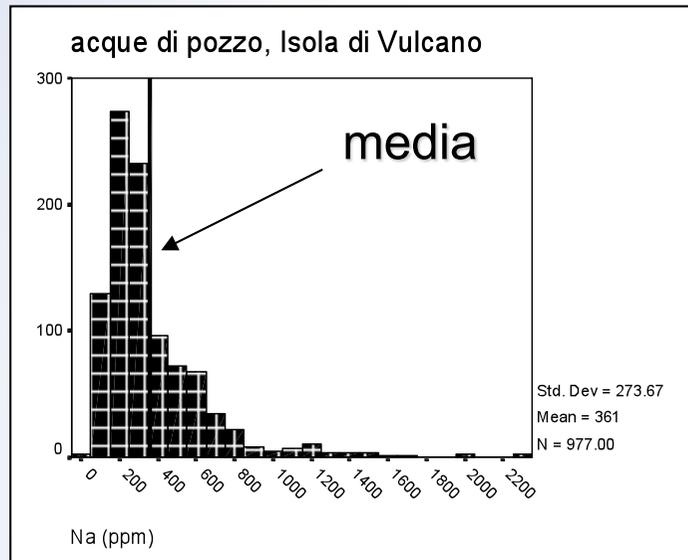
- il coefficiente di variazione è quindi una **misura relativa**, espressa in percentuale
- è particolarmente utile quando si devono confrontare le variabilità di due o più insiemi di dati espressi in unità diverse
- è utile anche quando si confrontano due o più insiemi di dati che sono misurati nella stessa unità di misura ma che differiscono nella dimensione tanto da rendere poco significativo il confronto delle sole deviazioni standard

La forma di una distribuzione di frequenza

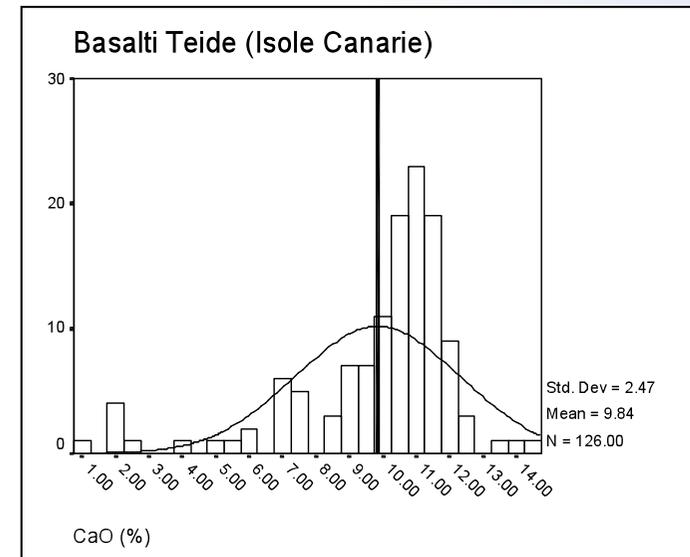
- Lo studio della forma della distribuzione di frequenza (chiamata anche **curva di densità**) dei dati fornisce utili indicazioni sui processi naturali all'origine della generazione dei valori
- La distribuzione dei dati può essere simmetrica o asimmetrica (obliqua)
- Per descrivere la forma di una distribuzione di frequenza è sufficiente confrontare la media con la mediana

- **media > mediana** **asim. positiva (obliquità destra)**
- **media = mediana** **simmetria**
- **media < mediana** **asim. negativa (obliquità sinistra)**

Esempi



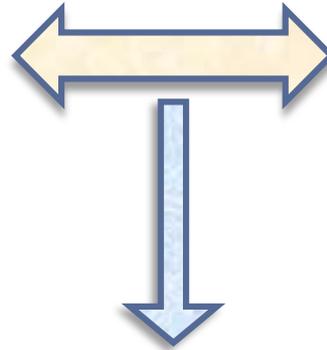
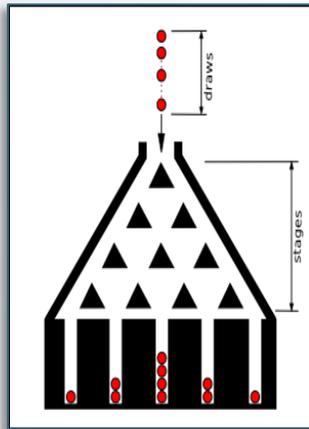
asimmetria positiva (obliquità destra): la media mostra valori maggiori della mediana



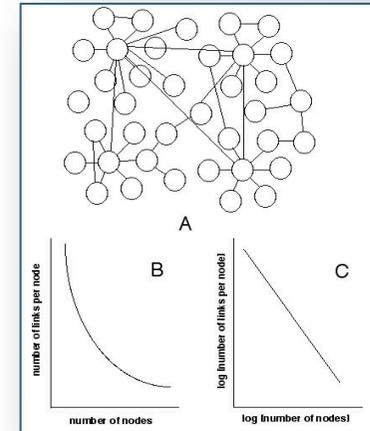
asimmetria negativa (obliquità sinistra)

Principali distribuzioni di frequenza utilizzate in geochimica

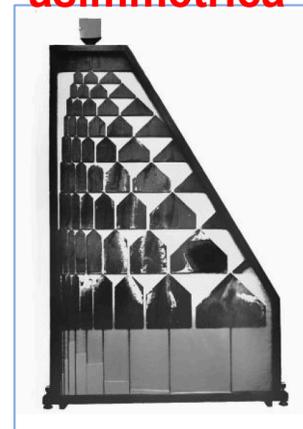
Distribuzione Gaussiana



Legge di Potenza (distribuzione a code pesanti)



Lognormale asimmetrica

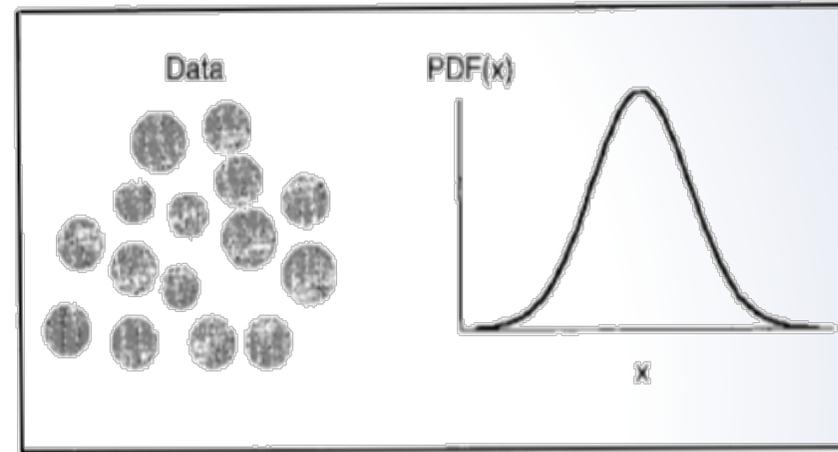


- Lo studio della distribuzione di frequenza di valori di concentrazione è in grado di fornire utili informazioni sulle dinamiche che riguardano il sistema in esame

Distribuzione normale

- La distribuzione Normale (o Gaussiana) è stata nel corso degli ultimi duecento anni il modo più utilizzato per descrivere i dati sperimentali di molte discipline scientifiche.
- Secondo il modello di probabilità normale i dati sono costituiti da **molti valori nei pressi di un valore medio**, con **pochi valori** che sono con uguale probabilità **più piccoli e più grandi**.

“Normal”



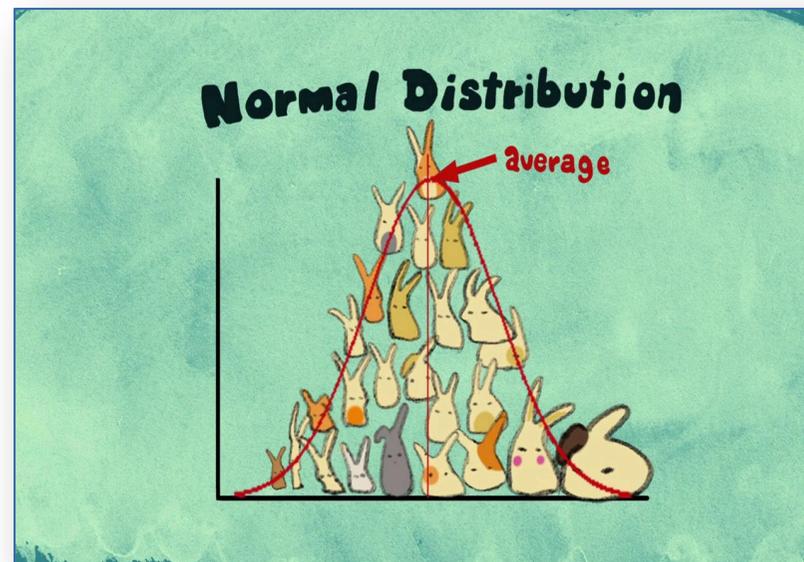
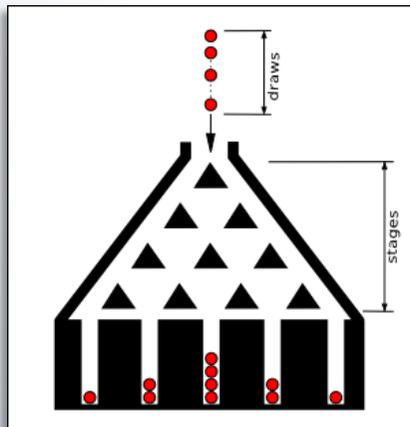
Distribuzione normale (x rappresenta la misura del diametro dei cerchi).

La funzione di probabilità di densità $PDF(x)$, cioè la probabilità che una qualsiasi misurazione ha di avere un valore compreso tra x e $x+dx$, è rappresentata in questo caso dalla classica curva a “campana” di Gauss.

Distribuzione normale

Una distribuzione normale in geochimica è considerata come:

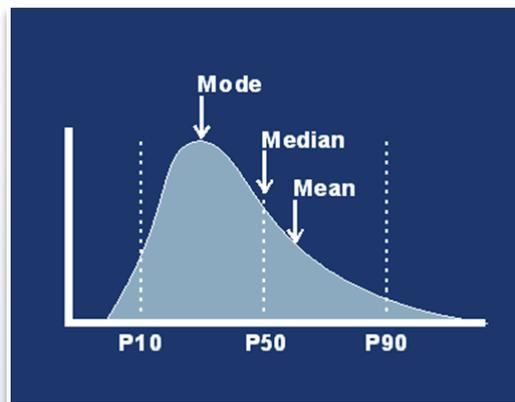
- risultato di **processi di mixing**,
- dominata da **effetti di tipo additivo** (*Allegre & Lewin, 1995*)
- in altre parole, deriva dalla somma di numerose componenti (processi o variabili) indipendenti.



Distribuzione lognormale

Una distribuzione lognormale in geochimica è considerata come:

- dominata da **effetti di tipo moltiplicativo**.
- deriva dalla prodotto di numerose componenti (processi o variabili) indipendenti.
- una serie sequenziale di **processi casuali di diluizione** rappresenta un esempio calzante del ruolo degli effetti moltiplicativi e conduce sempre a comportamenti log-normali.



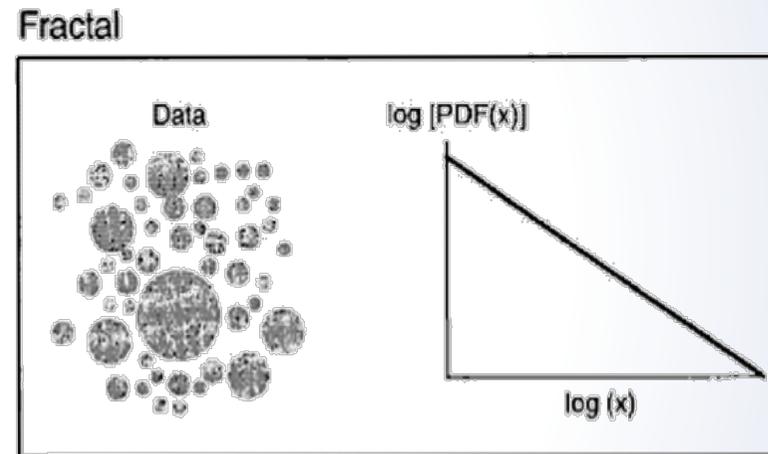
Distribuzione frattale

Una distribuzione frattale (distribuzione di Pareto o “a code pesanti”) in geochimica è considerata come:

- Dominata da **dinamiche di feedback** interdipendenti.
- Si tratta di **sistemi auto-simili**, nel senso che le parti più piccole sono copie ridotte delle parti più grandi. Un albero è un tipo di frattale, esso è costituito da un numero sempre maggiore di rami sempre più piccoli e non vi è alcun significato nel considerare il diametro medio di questi rami. (es: catena montuosa o un arcipelago ma può essere anche un processo nel tempo, costituito da fluttuazioni di ampiezza sempre minore o un insieme di numeri derivati da un processo sperimentale.)
- I dati di un oggetto frattale, vale a dire i numeri che lo costituiscono, consistono di alcuni valori di grandi dimensioni, **molti valori medi e un gran numero di piccoli valori**.

Distribuzione frattale

- La funzione di probabilità di densità PDF (x), nel caso di un processo frattale non è quindi normale. Generalmente si tratta di una linea retta su un grafico $\log[\text{PDF}(x)]$ contro $\log(x)$ che indica la presenza di una “legge di potenza”
- Dato che la media non è stazionaria ma può aumentare o diminuire all'aumentare dei dati, per caratterizzare le misure frattali in modo significativo viene utilizzato un parametro **chiamato Dimensione frattale (D)**, legato alla pendenza della retta.



Distribuzione frattale (x rappresenta la misura del diametro dei cerchi).

L'analisi esplorativa dei dati

Le tre caratteristiche principali dei dati sono posizione, variabilità e forma della distribuzione di frequenza. Mediante il **diagramma a scatola**, basato su **5** numeri di sintesi, è possibile effettuare una analisi esplorativa dei dati molto utile

I cinque numeri di sintesi:

x_{minimo} , Q_1 , **mediana (Q_2)**, Q_3 , x_{massimo}

a partire da questi numeri si possono ottenere tre misure di posizione (mediana, media interquartile, range medio) e due misure di variabilità (range interquartile e range) che consentono di avere una idea della forma della distribuzione di frequenza

I 5 numeri di sintesi in caso di simmetria...

- La distanza tra Q_1 e la mediana è uguale alla distanza tra Q_3 e la mediana
- La distanza tra x_{minimo} e Q_1 è uguale alla distanza tra Q_3 e x_{massimo}
- Mediana, media interquartile e range medio sono uguali (e coincidono con la media dei dati)

I 5 numeri di sintesi in caso di asimmetria...

- In distribuzioni oblique a destra la distanza tra Q_3 e x_{massimo} è maggiore della distanza tra x_{minimo} e Q_1
- In distribuzioni oblique a destra la mediana e la media interquartile sono più piccole del range medio
- In distribuzioni oblique a sinistra la distanza tra x_{minimo} e Q_1 è maggiore della distanza tra Q_3 e x_{massimo}
- In distribuzioni oblique a sinistra il range medio è più piccolo della mediana e della media interquartile

A partire dai 5 numeri di sintesi e dalle misure di posizione che da questi si possono derivare è possibile capire qualcosa della forma della distribuzione di frequenza dei dati...

Diagramma a scatola e baffi (box-plot)

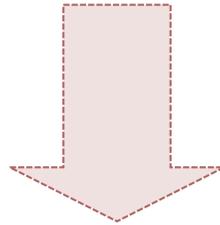
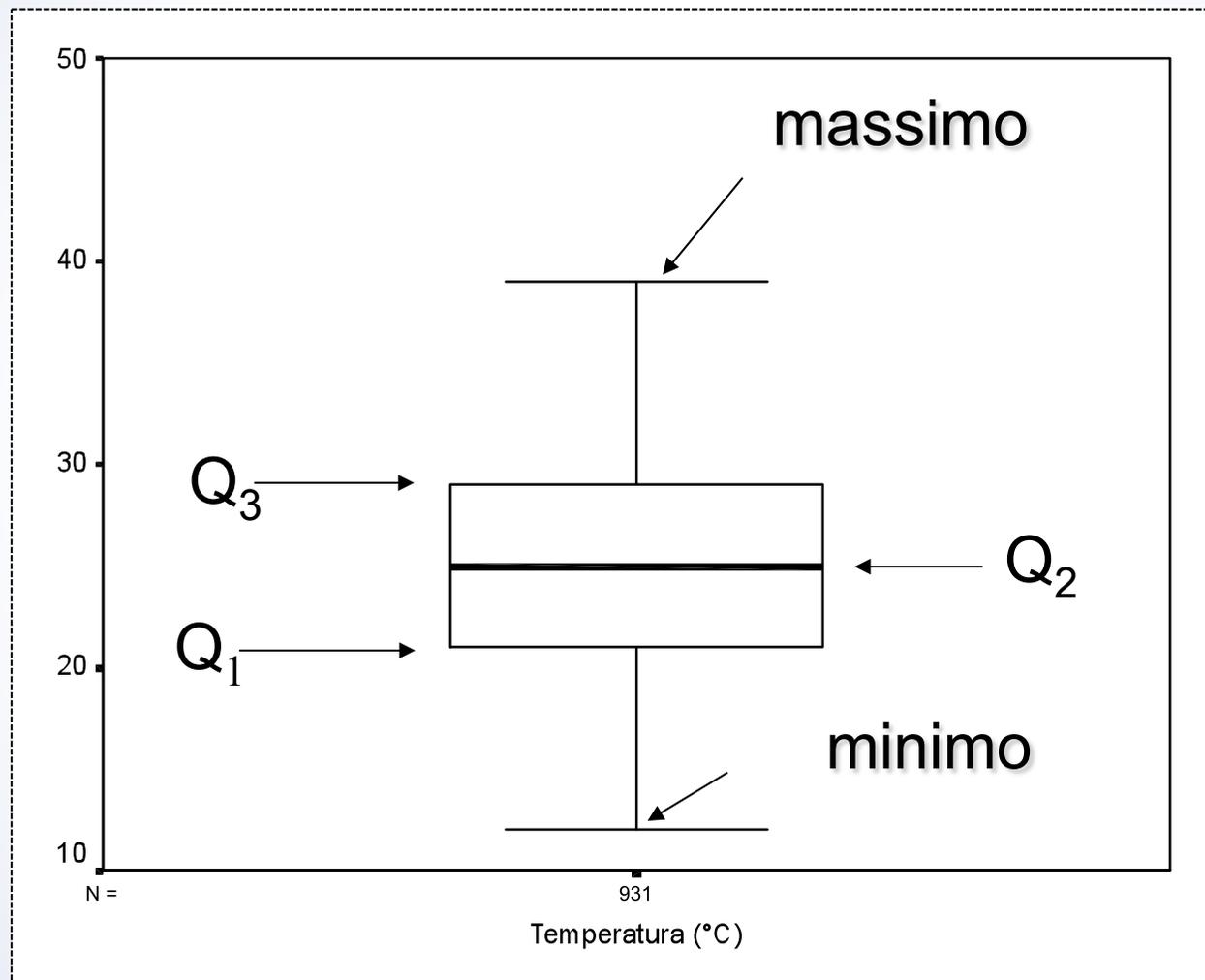


Diagramma che fornisce una **rappresentazione grafica** dei dati sulla base dei soli 5 numeri di sintesi.

Esempio di diagramma a scatola

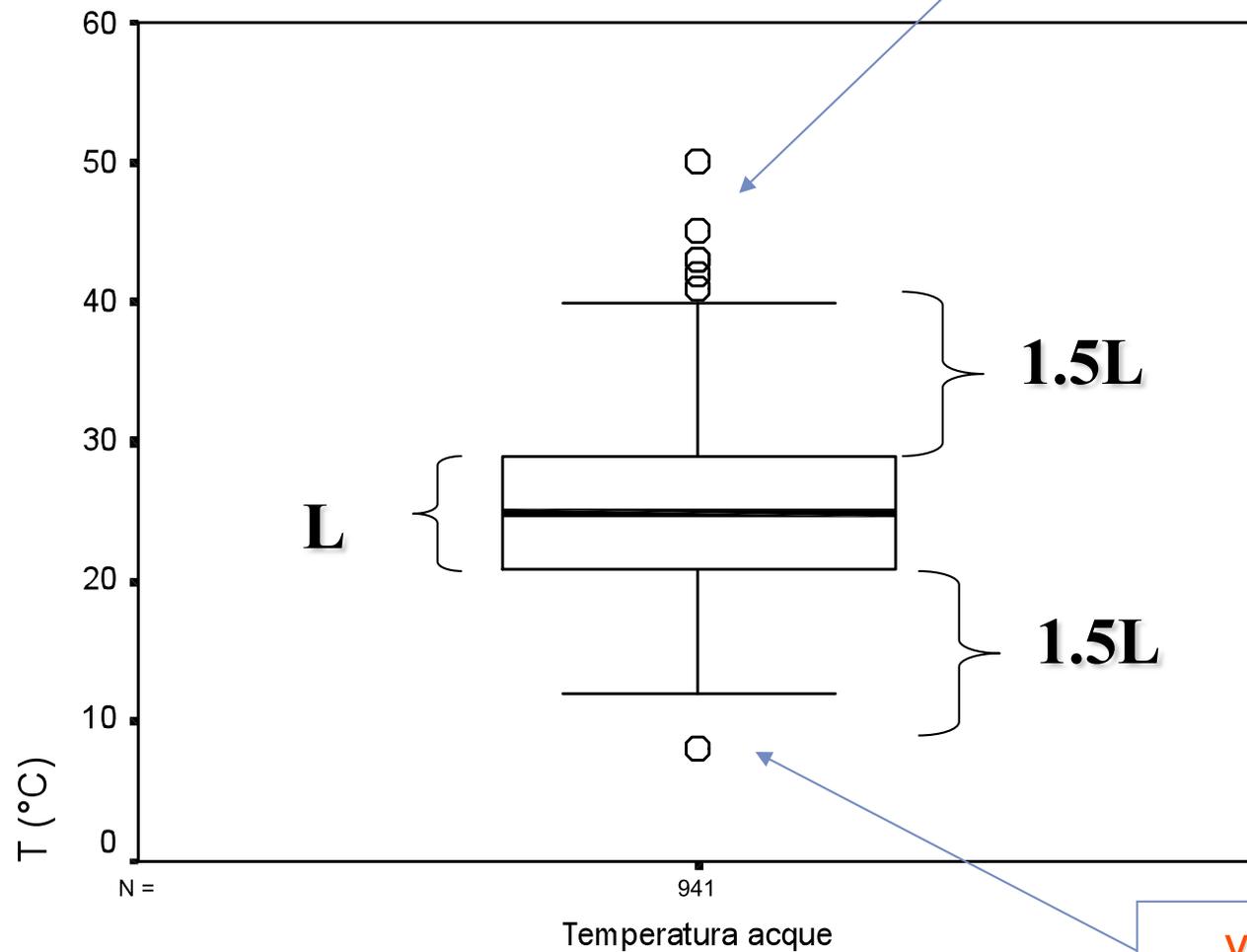


- La scatola contiene il **50%** delle osservazioni della distribuzione
- Il **25%** dei dati più piccoli è rappresentato da una linea (baffo) che collega il lato inferiore della scatola (Q_1) con x_{minimo}
- Il **25%** dei dati più grandi è rappresentato da una linea (baffo) che collega il lato superiore della scatola (Q_3) con x_{massimo}

Il diagramma a scatola e l'individuazione dei valori anomali

- Si considera la differenza $L = Q_3 - Q_1$
- Si riporta il segmento di lunghezza $1.5L$ sia verso l'alto che verso il basso, partendo, rispettivamente da Q_3 e Q_1
- Si verifica la presenza di valori maggiori o minori dei valori limite così definiti; se presenti, tali valori sono anomali (outliers)
- I nuovi baffi congiungeranno Q_3 e Q_1 con i nuovi valori di massimo e minimo non considerati anomali (sono i valori più basso e più alto, entro $1.5L$)

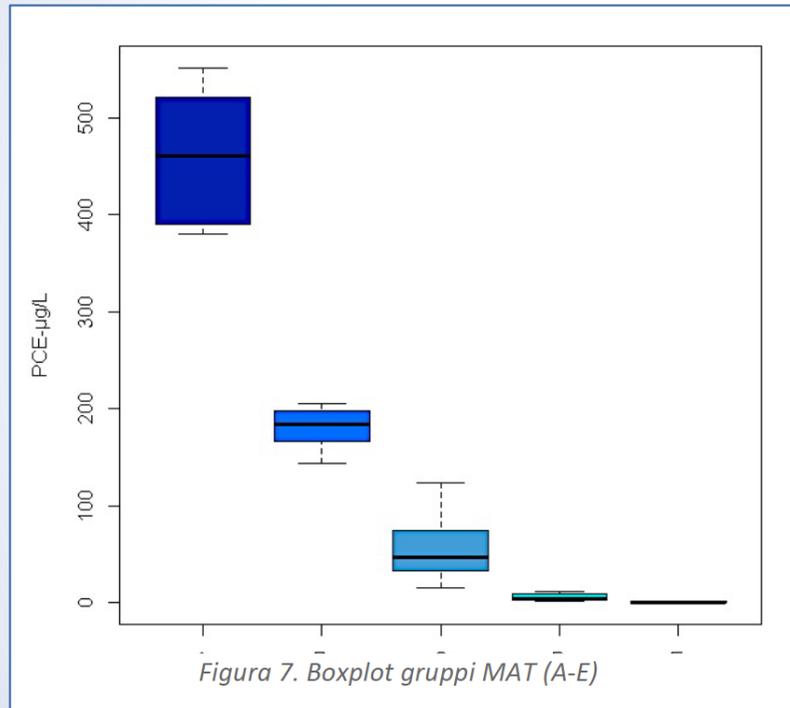
pozzi Isola di Vulcano



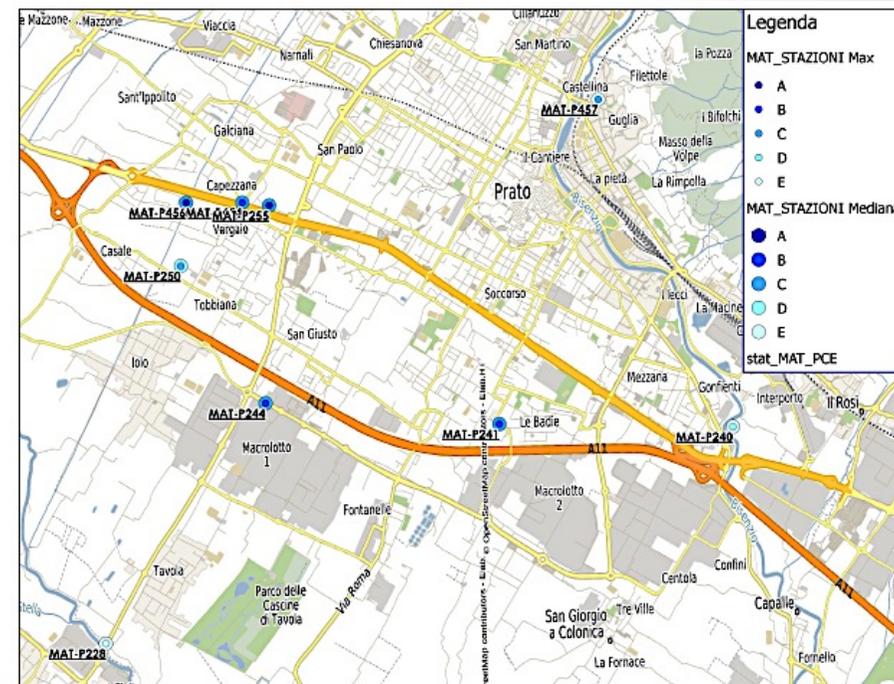
valori
anomali

valore
anomalo

Confronto tra i valori della stessa variabile per gruppi di dati: Percloroetilene corpo Idrico di Prato



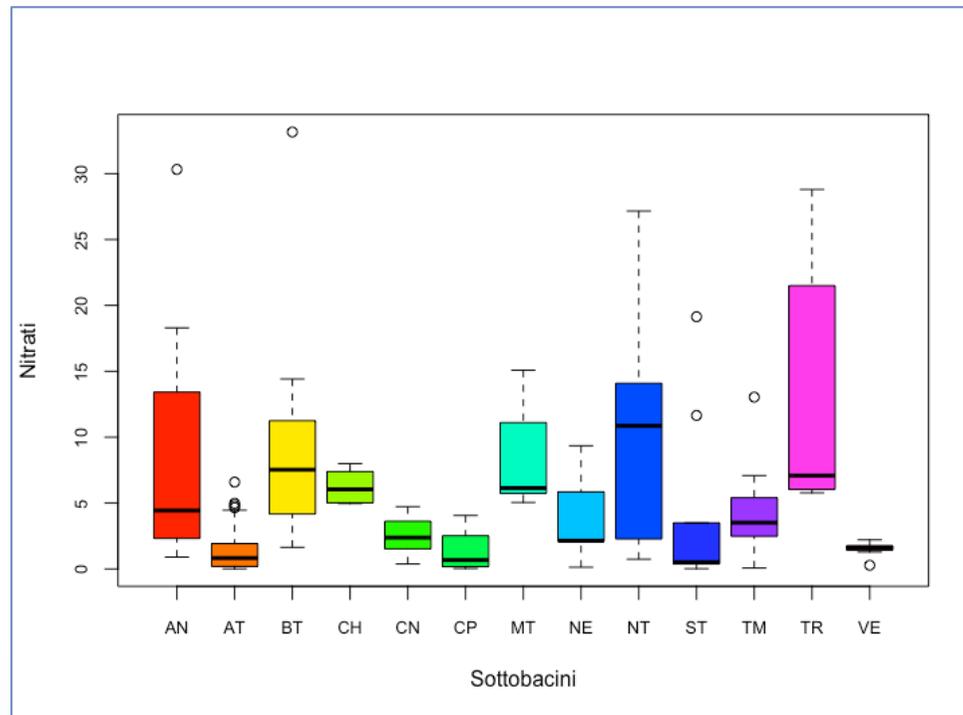
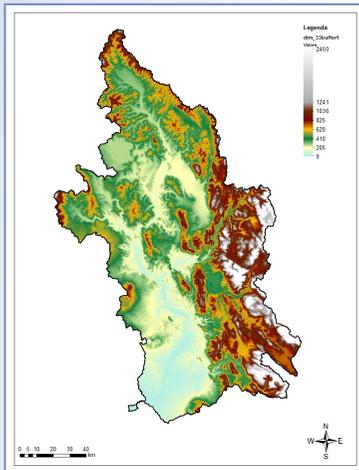
PCE-MAT	NumObs	Minimum	Maximum	Mean	Median	Variance
A	8	380	552	460,1	461,5	4725
B	8	144	206	180,5	183,5	430,9
C	96	15	124	55,4	46,8	748
D	32	1,4	11	5,469	4,4	11,49
E	27	0,1	1,2	0,67	0,7	0,0652



Confronto tra i valori della stessa variabile per gruppi di dati:

Acque superficiali del bacino del fiume Tevere

Nitrati

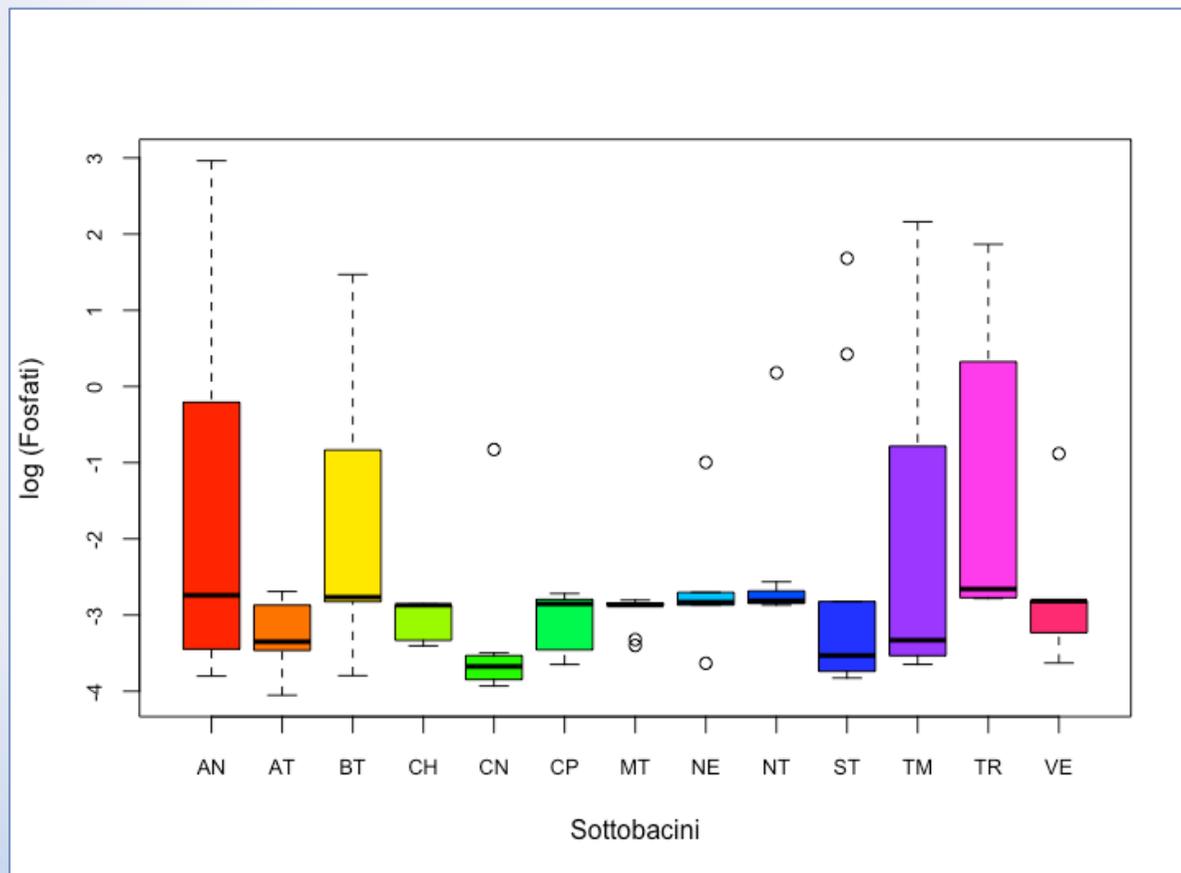


Nome Sottobacino	Sigla
Alto Tevere	AT
Chiascio	CH
Nestore e Trasimeno	NT
Topino Marroggia	TM
Medio Tevere	MT
Chiani e Paglia	CP
Basso Tevere	BT
Nera	NE
Corno e Nera a Monte del Velino	CN
Velino	VE
Salto e Turano	ST
Aniene	AN
Tevere area urbana Roma	TR

Confronto tra i valori della stessa variabile per gruppi di dati:

Acque superficiali del bacino del fiume Tevere

Fosfati



Nome Sottobacino	Sigla
Alto Tevere	AT
Chiascio	CH
Nestore e Trasimeno	NT
Topino Marroggia	TM
Medio Tevere	MT
Chiani e Paglia	CP
Basso Tevere	BT
Nera	NE
Corno e Nera a Monte del Velino	CN
Velino	VE
Salto e Turano	ST
Aniene	AN
Tevere area urbana Roma	TR