

ISTITUTO SUPERIORE DI SANITÀ

**Elementi di statistica
nelle procedure di calcolo
per il controllo della qualità dell'aria**

Maria Chiara Mura
Laboratorio di Igiene Ambientale

ISSN 1123-3117

Rapporti ISTISAN

01/2

Istituto Superiore di Sanità

Elementi di statistica nelle procedure di calcolo per il controllo della qualità dell'aria.

Maria Chiara Mura

2001, 28 p. Rapporti ISTISAN 01/2

Il processo di elaborazione statistica dei dati provenienti dal monitoraggio dell'inquinamento chimico dell'atmosfera, finalizzato al controllo della qualità dell'aria, è presentato in modelli di procedure al fine di fornire un sintetico strumento di lavoro agli operatori del settore. I modelli di procedure sono modulari ed integrabili. Includono gli elementi di calcolo elementare ed i metodi statistici d'analisi. Gli elementi di calcolo sono sviluppati con metodo d'induzione probabilistica per collegarli ai modelli statistici, che sono alla base dei metodi d'analisi nello studio del fenomeno dell'inquinamento atmosferico anche a fini previsionali. Il rapporto si inserisce nell'attività di aggiornamento e di formazione che fin dagli anni ottanta l'Istituto Superiore di Sanità indirizza agli operatori del settore ambientale.

Parole chiave: Controllo qualità dell'aria, Indici statistici, Modelli procedurali, Parametri statistici

Istituto Superiore di Sanità

Statistical elements in calculation procedures for air quality control.

Maria Chiara Mura

2001, 28 p. Rapporti ISTISAN 01/2 (in Italian)

The statistical processing of data resulting from the monitoring of chemical atmospheric pollution aimed at air quality control is presented. The form of procedural models may offer a practical instrument to the operators in the sector. The procedural models are modular and can be easily integrated with other models. They include elementary calculation procedures and mathematical methods for statistical analysis. The calculation elements have been developed by probabilistic induction so as to relate them to the statistical models, which are the basis of the methods used for the study and the forecast of atmospheric pollution. This report is part of the updating and training activity that the Istituto Superiore di Sanità has been carrying on for over twenty years, addressed to operators of the environmental field.

Key words: Air quality control, Procedural models, Statistical indexes, Statistical parameters

Si ringraziano Sandro Mazzola, Giuseppe Paolangeli e Luigi Pezone del Laboratorio di Igiene Ambientale (Istituto Superiore di Sanità) che hanno collaborato alla presentazione delle procedure grafiche. Inoltre, un particolare ringraziamento a Sandra Salinetti per gli utili suggerimenti dati durante la revisione del testo.

Il rapporto è disponibile online nel sito di questo Istituto: www.iss.it.

INDICE

Introduzione.....	1
1. Modelli di procedure e presentazione dei risultati.....	3
1.1. Procedura d'acquisizione e di validazione	3
1.1. 2. Unità statistiche.....	5
1.2. Procedure di controllo della qualità dell'aria nell'ambiente esterno	6
1.3. Procedure di lavoro: metodologia statistica.....	8
1.4. Forma grafica	10
2. Elementi di statistica.....	14
2.1. Indici di valore centrale.....	14
2.1. 1. Media aritmetica e ponderata.....	14
2.1. 2. Mediana	15
2.2. Percentile.....	15
2.3. Variabilità del carattere	15
2.3. 1. Varianza.....	15
2.3. 2. Deviazione standard	16
2.3. 3. Coefficiente di variazione.....	16
2.3. 4. Errore standard sulla media	16
2.4. Dagli indici statistici ai parametri statistici.....	16
2.5. Curva delle distribuzioni normali	17
2.6. Stima dei parametri.....	19
2.6. 1. Parametri statistici: media e varianza	19
Bibliografia.....	21
Appendice A	25
Tecniche grafiche	25
Appendice B	26
Matrice.....	26
Appendice C	28
Percentile	28

PREMESSA

Il rapporto nasce dal lavoro e dall'esperienza acquisita dall'Istituto Superiore di Sanità nell'ambito di gruppi di lavoro e di corsi di aggiornamento e di formazione diretti agli operatori ambientali nel settore del controllo della qualità dell'aria.

Il lavoro ha preso in considerazione la difficoltà di alcuni di loro a misurarsi con processi di analisi e di elaborazione statistica dei dati provenienti dal monitoraggio in stazione dei parametri chimici dell'atmosfera, per competenze solo parzialmente formate ed in ogni caso non uniformemente distribuite sul territorio, mentre i sistemi informatici di gestione e i programmi di calcolo automatico hanno avuto una crescita esponenziale nell'ultimo decennio. D'altra parte, non si ignora il potenziale sviluppo applicativo dei programmi che si fondano sui metodi matematici della statistica e la loro facile accessibilità.

Per questo motivo, il rapporto introducendo gli elementi statistici, cioè gli indici statistici, come previsti dalla normativa per effettuare la valutazione di superamenti dei contaminanti monitorati, li ridefinisce e li formalizza come parametri statistici per ricollegarli ai modelli statistici ed ai metodi matematici della statistica. Questi ultimi, applicati correttamente, permettono di rappresentare il comportamento del fenomeno che si vuole mettere in evidenza (ritardi di un contaminante sull'altro, periodicità di presentazione di valori massimi, ecc.).

Trattandosi di procedure sia di acquisizione che di elaborazione, l'argomento viene sviluppato mediante schemi o modelli procedurali, presentati in diagrammi a blocchi secondo i metodi e le tecniche tipici della ricerca operativa. Informazione sintetica sui criteri di costruzione di un modello è data, per collocare il rapporto in un più ampio contesto culturale, mentre esempi applicativi potrebbero essere oggetto di successivi rapporti.

INTRODUZIONE

La statistica è la scienza che analizza fenomeni collettivi riguardanti un gran numero di soggetti altrimenti indistinguibili. I suoi metodi sono metodi matematici che, oggi, sono spesso disponibili sotto forma di pacchetti di programmi d'elaborazione dati.

Nell'attuale contesto culturale, caratterizzato da una crescente produzione di dati e dalla necessità di estrarre da questi informazione compatta, facilmente leggibile, affidabile, riproducibile, confrontabile nel tempo e nello spazio, ben si spiegano sia l'attenzione alla statistica, sia la crescente domanda di prodotti automatici da applicarsi ai dati per ottenere informazione con le suddette caratteristiche.

Ed è attraverso questi pacchetti che anche gli operatori di vari settori di lavoro hanno facile accesso alla statistica, la cui applicazione richiede sia una robusta preparazione scientifica di base di tipo matematico-statistico, sia un continuo aggiornamento metodologico. Il fatto merita la massima attenzione per le possibili conseguenze sulla qualità dell'informazione prodotta e spinge gli esperti statistici a pianificare, in forma divulgativa ma rigorosa, le procedure per gli operatori che lavorano in base alla normativa, perché possano trarre il massimo vantaggio dall'approfondimento delle loro conoscenze. Questo tipo di contributo può essere tanto più significativo nel settore di controllo della qualità dell'aria dell'ambiente esterno, ai fini della prevenzione del danno alla salute pubblica.

Il controllo avviene attraverso il monitoraggio continuo di parametri d'inquinanti chimici per i quali, attraverso criteri sanitari, sono stabiliti i livelli massimi di concentrazione ammissibili in atmosfera durante un certo periodo di tempo. Il monitoraggio è effettuato mediante stazioni fisse o mobili, in rete o non, prevalentemente in aree urbane ed industriali. La quantità di dati prodotta ogni anno è nell'ordine di grandezza di milioni (es. 8760 dati orari/anno dal monitoraggio di un solo parametro in una stazione).

Si è nel caso tipico della necessità di utilizzare i metodi della statistica per produrre l'informazione necessaria all'area decisionale. I metodi che si utilizzano sono più o meno complessi in funzione del tipo d'informazione che si vuole ottenere. Si sviluppano a partire dal calcolo della media, della mediana e del percentile fino ai modelli di previsione del comportamento in atmosfera di uno o più inquinanti.

Concepito come strumento di lavoro, questo testo è strutturato per riprodurre le fasi del lavoro dell'operatore in forma di modelli logici di procedure in cascata, nei quali sono opportunamente inseriti gli elementi di statistica, le ipotesi del loro sviluppo e gli elementi di *feedback* (recupero del sistema), che coincide con l'attività decisionale dell'operatore. I modelli sono modulari ed integrabili tra loro e con qualsiasi altro modello.

1. MODELLI DI PROCEDURE E PRESENTAZIONE DEI RISULTATI

1.1. Procedura d'acquisizione e di validazione

In Figura 1 è schematizzato il modello logico (o modello operativo) delle procedure di controllo per la qualità dell'aria, dall'acquisizione del dato alla sua accettazione, come premessa a qualunque successiva applicazione di procedura di controllo. Le tecniche grafiche utilizzate sono spiegate per punti essenziali in Appendice A.

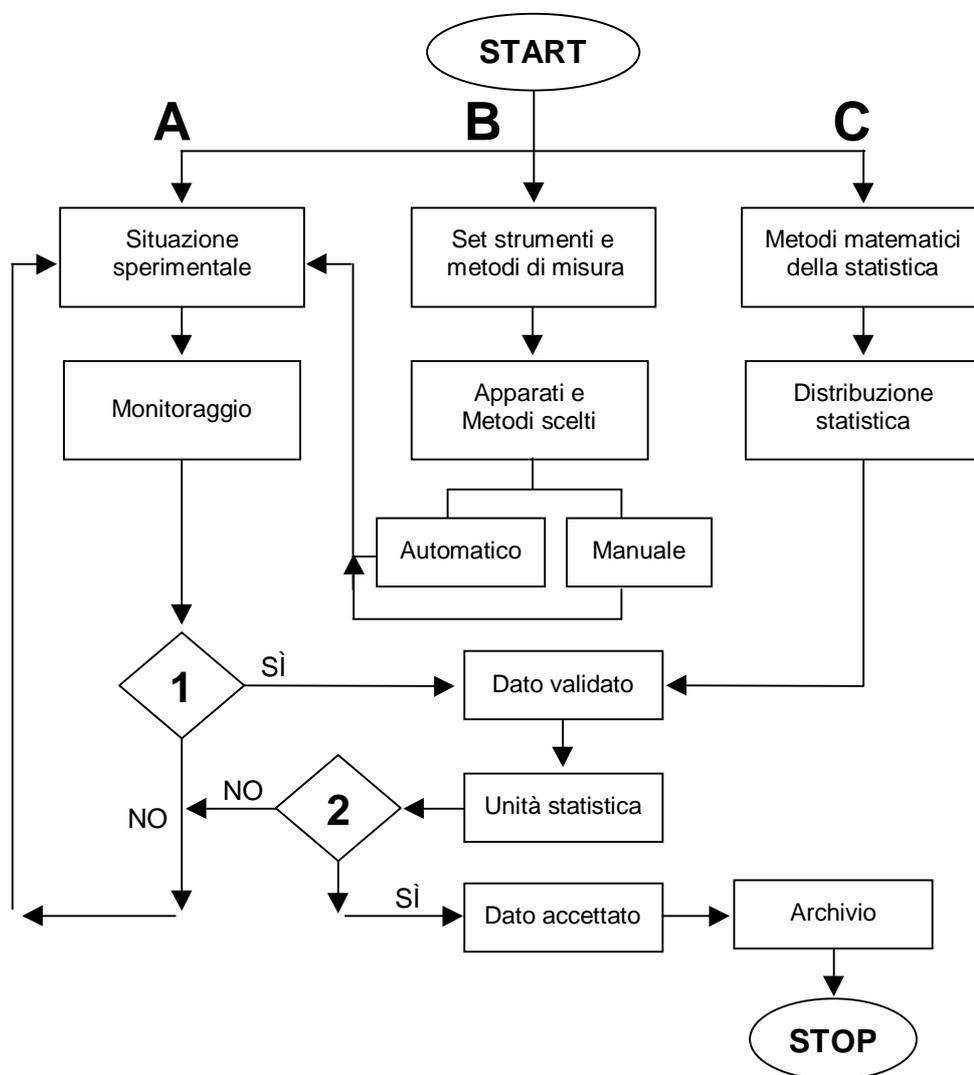


Figura 1. Modello operativo per le procedure di controllo: sulla situazione sperimentale (A), sul set degli strumenti metodi di misura (B), sui metodi matematici della statistica (C)

Il modello logico è costituito da tre sub-modelli connessi tra di loro tramite linee di flusso di procedure che, passo dopo passo, portano all'attivazione dei blocchi: A, B, C.

In Figura 1 sono rappresentate le tre situazioni iniziali, da ognuna delle quali si sviluppa la rispettiva procedura.

Leggendo da sinistra verso destra queste sono: la situazione sperimentale (sub-modello A); il set strumentale con i metodi di misura (sub-modello B); i metodi matematici della statistica, fondati sul modello di probabilità, con riferimento al modello di Gauss-Laplace (sub-modello C).

Definiti e messi a punto gli strumenti ed i metodi d'analisi *ad hoc* per il monitoraggio dell'inquinante (sub-modello B), essi sono applicati alla specifica situazione sperimentale in studio (sub-modello A) per fornire in uscita la caratteristica del fenomeno, o la proprietà del corpo, in forma quantificata. Questo è chiamato dato, intendendo con questo termine sia un insieme di numeri in sequenza in intervalli di tempo prefissati nel caso di registrazione del fenomeno in automatico, sia i risultati delle determinazioni analitiche con metodi manuali.

Sul dato s'innesta la linea di flusso proveniente dal sub-modello C, dove si trovano i metodi matematici della statistica per l'analisi dei dati prodotti.

L'applicazione dei metodi statistici al dato è possibile in quanto questo è una variabile aleatoria del processo fisico che, invece, è misurato, nel senso che tenteremo di spiegare partendo dalla situazione pratica in cui si trova lo sperimentatore e rimandando a testi specializzati l'approfondimento dell'argomento (1).

Effettuando nelle stesse condizioni sperimentali misure ripetute su una stessa grandezza fisica, infatti, si ottengono dati che variano in modo non prevedibile e che, nel loro insieme, sono soggetti ad una certa dispersione non controllabile. Si chiama variabile aleatoria x una grandezza fisica specifica, quella i cui valori sono soggetti a questa dispersione non controllabile (2). In pratica, i dati, altrimenti detti valori osservati, sono i risultati di n misure $x_1, x_2, x_3, \dots, x_n$ effettuate sulla variabile aleatoria x . Una variabile aleatoria, pertanto, può prendere un qualunque valore di un insieme determinato di valori. Alla variabile aleatoria è associata una distribuzione di probabilità. La legge di probabilità di una variabile aleatoria è rappresentata da una funzione che determina la probabilità che una variabile aleatoria prenda un qualsiasi valore dato o appartenga ad un insieme di valori dati.

Ciò detto, i metodi statistici del sub-modello C, che attraverso la linea di flusso si applicano per il trattamento del dato, sono di tipo specifico, nel senso che, per i dati provenienti da registrazioni in automatico e non, possono essere, per esempio, tecniche di ricostruzione di dati, necessarie quando le registrazioni presentano sostanziose lacune (3, 4) o tecniche per la valutazione della rappresentatività del dato (5); per il dato prodotto con metodi manuali sono formule che permettono di associare al dato il suo errore probabile. Tali formule discendono dallo sviluppo della teoria di Hagen (6) e si riconducono, quindi, al modello di Gauss, noto come modello della distribuzione normale. E' interessante notare come, a partire dall'analisi dell'errore associato alla misura, si sia sviluppata la teoria che ha portato alla definizione del modello di probabilità di Gauss.

I controlli indicati disegnati nelle rispettive losanghe, rappresentano i punti cruciali della procedure che il modello esegue. Questi riguardano l'attività dell'operatore che, ordinatamente, effettua: nel *check 1* (= controllo 1 in losanga 1), la validazione del dato con tecniche di 'esame a vista del guasto del dato'; nel *check 2* (= controllo 2 in losanga 2), l'accettazione dei risultati ottenuti dall'applicazione della procedura statistica. Se entrambi i *check* sono positivi, il dato è archiviato per successive elaborazioni. Se il *check 1* dà risposta negativa, cioè il dato non è validato, si riattiva l'intera procedura su un'altra situazione sperimentale; se i metodi statistici applicati sul dato validato non soddisfano il *check 2*, si riattiva la procedura del blocco C.

In Figura 1, non sono dettagliati i controlli sul blocco B che verificherebbero il funzionamento della strumentazione e la bontà di risposta del metodo di rilevamento, in quanto l'argomento esula dalla presente trattazione.

Il sub-modello C è in un solo blocco, definito come i metodi matematici della statistica. La procedura generalizzata d'applicazione di tali metodi sarà successivamente presentata ed analizzata.

Quando il dato è accettato, esso costituisce l'elemento di input per altre procedure o processi quali, per esempio, d'elaborazione matematica. Per fare ciò, si toglie lo stop disegnato nella Figura 1, la linea di flusso rimane aperta così da potersi innestare su altre procedure.

Elementi progettuali di questo modello, come degli altri che saranno presentati, sono:

- la modularità, garantita dallo stop (stop logico), che permette al modello di vivere di vita autonoma;
- l'integrabilità con altri modelli procedurali, resa possibile dalla linea di flusso che, come già detto, rimane aperta quando si toglie lo stop.

Modelli con queste caratteristiche sono strumenti operativi molto flessibili che è sempre possibile sia modificare al fine di soddisfare esigenze non previste al presente, sia automatizzare: loro peculiarità è quella di produrre un dato affidabile per qualunque controllo e/o ricerca.

Il dato prodotto è, generalmente, archiviato su opportuni supporti, predisposti per la sua trattazione informatica. Per l'archiviazione è fondamentale disegnare, secondo opportune specifiche, i modelli di data base. La loro realizzazione tecnica è, poi, funzione dell'attualità della strumentazione che è, peraltro, in continua evoluzione (7, 8).

1. 1. 2. Unità statistiche

I dati prodotti secondo la procedura in Figura 1 sono variabili aleatorie di un processo continuo, rilevate, come si vedrà, con modalità definite. In questo senso coincidono con le unità statistiche. Ogni variabile aleatoria corrisponde al rilevamento di un inquinante; l'insieme di n variabili aleatorie è rappresentabile sotto forma di tabella, in cui ogni elemento è un'unità statistica. La tabella è un primo efficace strumento di lavoro, perché compatta in un archivio organico i dati, facilitandone la lettura e predisponendone la gestione anche ai fini elaborativi.

Nella fattispecie si preferisce chiamare la tabella con il termine di matrice per introdurre il concetto che in matrice gli elementi possono non essere unità statistiche ma, per esempio, risultati di processi d'elaborazione statistiche sui quali si sviluppano ulteriori processi di calcolo finalizzati all'ottimizzazione di modelli predittivi (9-11). Una matrice che abbia n righe e m colonne si dice di dimensione n x m e si scrive in forma compatta: $M(c_{i,j})$, dove $c_{i,j}$ è l'elemento generico e i pedici i e j sono, rispettivamente, l'indice di riga e di colonna (Appendice B).

In Tabella 1 i dati ottenuti dal monitoraggio di tre variabili aleatorie, rispettivamente il monossido di carbonio (CO), il biossido d'azoto (NO₂) e l'ozono (O₃), sono rappresentati sotto forma di matrice di dimensione 3 x m.

Tabella 1 . Rappresentazione delle unità statiche in forma di matrice 3 x m

Variabile aleatoria	Unità statistica						
CO	C _{1,1}	C _{1,2}	C _{1,3}	C _{1,4}	C _{1,5}	C _{1,m}
NO ₂	C _{2,1}	C _{2,2}	C _{2,3}	C _{2,4}	C _{2,5}	C _{2,m}
O ₃	C _{3,1}	C _{3,2}	C _{3,3}	C _{3,4}	C _{3,5}	C _{3,m}

1. 2. Procedure di controllo della qualità dell'aria nell'ambiente esterno

Completato il processo di cui in Figura 1, si attiva la procedura secondo la quale si effettua il controllo della qualità dell'aria per l'ambiente esterno. La procedura è riportata nella Figura 2 nella forma più generale e compatta, essendo stato seguito il criterio di minimizzare il numero delle definizioni massimizzandone il loro contenuto; è disegnata in forma integrabile, in pratica, aperta alla precedente (Figura 1) alla quale, togliendo lo stop, si ricollega, diventandone un suo modulo. Disegnando lo start, la procedura può vivere di vita autonoma. Input della procedura in Figura 2 è l'output della precedente riportata in Figura 1. L'elemento input (dati statistici) sembrerebbe ridondante in questa procedura di tipo integrabile; tuttavia, è stato necessario disegnarlo per mettere in evidenza sia la possibile autonomia della procedura, sia il fatto che è su questo input che si costruisce il campione statistico.

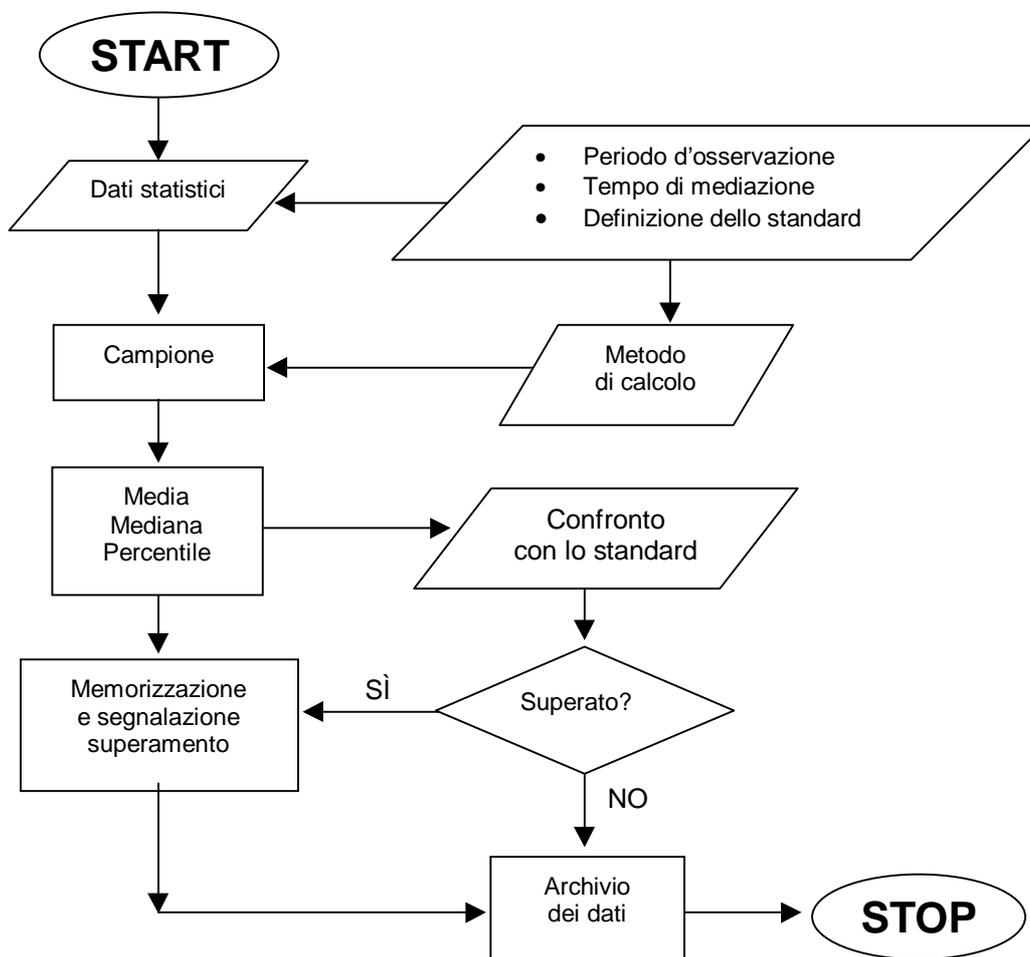


Figura 2. Modello di procedure per il controllo di qualità dell'aria

Facendo riferimento ai valori di concentrazione di un singolo inquinante, gli elementi della procedura, ordinatamente, sono:

- a) i dati in ingresso, o, per quanto detto precedentemente, le unità statistiche delle concentrazioni di un identificato inquinante;
- b) il suo tempo di mediazione¹ ed il suo periodo d'osservazione²;
- c) i metodi di calcolo, specifici per l'inquinante in oggetto;
- d) il valore limite di riferimento, relativo all'inquinante monitorato, per esempio lo standard di qualità dell'aria;
- e) i dati in uscita.

Sul campione è effettuata l'elaborazione dei dati dell'inquinante. Gli elementi di calcolo che si debbono applicare per elaborare i campioni di ciascun inquinante rilevato, allo stato attuale, sono: la media, la mediana e il percentile. Il risultato dell'elaborazione produce un numero che è confrontato con il valore limite in atmosfera di quell'inquinante. La losanga, nella quale avviene il confronto, prevede come uscita sia la risposta di superamento negativa, sia quella positiva. Nel primo caso il dato è archiviato, nel secondo caso è inviato in area decisionale e, successivamente, archiviato con segnalazione di guasto dello stato dell'atmosfera rispetto all'inquinante elaborato matematicamente. Le linee di flusso che collegano i vari blocchi evidenziano la procedura.

E' interessante notare che la costruzione del campione statistico e la sua dimensione (che, notoriamente, rappresenta il momento cruciale per qualsivoglia indagine statistica, in quanto dalla sua rappresentatività dipende la bontà di risposta del trattamento statistico) è, in questo caso, guidata dai criteri, definiti dalla normativa, che tengono conto del tempo di mediazione e del periodo d'osservazione.

Come nella procedura riportata in Figura 1, anche in questo caso la procedura in Figura 2 si chiude su un archivio di dati disponibili per successive elaborazioni quali, primariamente, quelle mirate alla valutazione della distorsione dei dati (5), alla ricostruzione dei dati mancanti (3, 4, 12), alla rappresentazione dell'evoluzione del fenomeno nel tempo e nello spazio. I dati dell'archivio costituiscono serie temporali: cioè, ogni dato della serie è un valore di concentrazione, per esempio, oraria, giornaliera, mensile, ecc., equidistanziato temporalmente dal successivo e dal precedente.

All'analisi delle serie temporali è rivolta da vari anni l'attenzione del mondo scientifico internazionale a partire da quello economico e da quello meteorologico, dove sono stati messi a punto quei metodi matematici di analisi che hanno permesso di delineare classi di modelli noti come modelli di previsione. Questi metodi, da ormai oltre venti anni, si studiano e si applicano anche sulle serie temporali provenienti dal monitoraggio in continuo degli inquinamenti chimici dell'atmosfera (13, 14).

Lo sviluppo di queste metodologie statistiche si rivela particolarmente vantaggioso quando si dispone unicamente dei dati di monitoraggio, perché sui soli dati di concentrazione in atmosfera si possono studiare e caratterizzare i loro comportamenti (modelli comportamentali) quali, per esempio: le tendenze e l'accumulazione (15), la quantificazione del ritardo di una variabile su altra/e (14), il periodo di ritorno di valori estremi (17). La variabile che descrive il fenomeno è chiamata variabile di stato, o variabile endogena, del sistema o del modello che descrive il fenomeno.

¹ Il tempo di mediazione corrisponde all'intervallo su cui la funzione continua (registrazione) è matematicamente integrata. Un'integrazione su un intervallo temporale di un'ora dà il valore di concentrazione media dell'inquinante monitorato in un'ora.

² Il periodo d'osservazione è il tempo relativo al periodo in cui l'inquinante è monitorato.

Variabili atmosferiche, cause inquinanti e altre entità sono variabili esogene del modello che si utilizzano, se disponibili, per ottimizzarlo, ma che non sono strettamente necessarie per la descrizione del fenomeno. Nelle serie temporali, infatti, è scritta la storia presente e futura del fenomeno perché ogni elemento che le compone ha memoria del precedente e contiene elementi che possono contribuire ad indicare lo stato futuro.

La bontà di risposta del modello dipende dalla capacità di scegliere il campione, di scrivere le equazioni e i vincoli matematici, di risolvere il sistema di equazioni, di valutarne le azioni concomitanti, e, soprattutto di minimizzare l'errore di adattamento (9-11). Per fare ciò si devono definire lo spazio matematico e scegliere i metodi matematici della statistica assieme ai test specifici di valutazione dei risultati, tenendo presente che non sono i dati che si devono adattare ai metodi, ma i metodi ai dati (18). Si parla, in generale, di messa a punto di metodologie statistiche su dati sperimentali (19-22). Se il risultato dell'applicazione matematico-statistica è soddisfacente, tale cioè che il fenomeno è descritto in forma di un modello che ne spiega il comportamento, il metodo è provato in altre situazioni sperimentali e, successivamente, validato.

L'ampia disponibilità sul mercato di software che traducono metodi matematici incoraggia ad applicarli per costruire modelli. Ciò si rivela vantaggioso quando viene delineato un modello concettuale con opportuni criteri³.

1. 3. Procedure di lavoro: metodologia statistica

La prova d'applicazione di selezionati metodi statistici a campioni (statistici) e la scelta del metodo di elaborazione – più adeguato all'elaborazione del tipo di dati in studio – sono disegnate nella procedura in Figura 3 con gli stessi criteri di modularità e d'integrabilità adottati nelle precedenti figure. Operativamente, togliendo lo start, la procedura in Figura 3, può essere messa in cascata, sia con il modello di procedure per il controllo della qualità dell'aria, sia con il modello operativo riportato in Figura 1, sia con qualsiasi altra procedura o sub-procedura.

I blocchi principali del modello sono: il campione, i metodi statistici d'analisi, i risultati e la loro accettazione. Attraverso una linea di flusso, uscente dal blocco metodi statistici, un metodo fra quelli predisposti è, ordinatamente, applicato al campione n-esimo per produrre un risultato che è statisticamente valutato per essere accettato o respinto. Nel caso in cui sia accettato, anche il metodo statistico adottato viene accettato e provato su un altro campione, se esistente; nel caso in cui il risultato non sia soddisfacente, la procedura prevede la possibilità di utilizzare un altro metodo statistico, se esistente, fra quelli predisposti. Se non esistono altri metodi, il risultato ottenuto dall'applicazione precedente viene rigettato assieme al metodo.

Come nelle precedenti presentazioni dei modelli, il *check* ("OK?" in losanga) è il momento cruciale della procedura e quindi è devoluto all'attività dell'operatore, unico che può decidere fino a quanto il prodotto ottenuto può scostarsi da quello atteso. Per fare ciò, prima di qualunque

³ Criterio base per fare un modello, prima concettuale e successivamente matematico, è quello di distinguere il superfluo dall'essenziale. Un modello è 'buono' in funzione della bontà di risposta che dà al quesito posto (problema della pertinenza). La bontà di risposta del modello cresce al crescere di tutti gli aspetti di rilievo – richiesti dall'ipotesi di lavoro – che riesce ad evidenziare ed al decrescere degli aspetti di cui non si vuole tenere conto nell'analisi del sistema reale. Non esiste un'unica soluzione e possono andare bene svariati modelli: evidentemente, si dovrà scegliere tra questi, a posteriori, quello che riproduce con maggiore approssimazione il comportamento del fenomeno. Un buon modello, quindi, dovrebbe essere costruito in base al principio della logica concreta, principio secondo il quale tanto più un concetto è ricco di determinazione, tanto più è attivo.

applicazione di metodo, l'operatore stabilisce il livello di probabilità al quale può accettare o respingere il risultato.

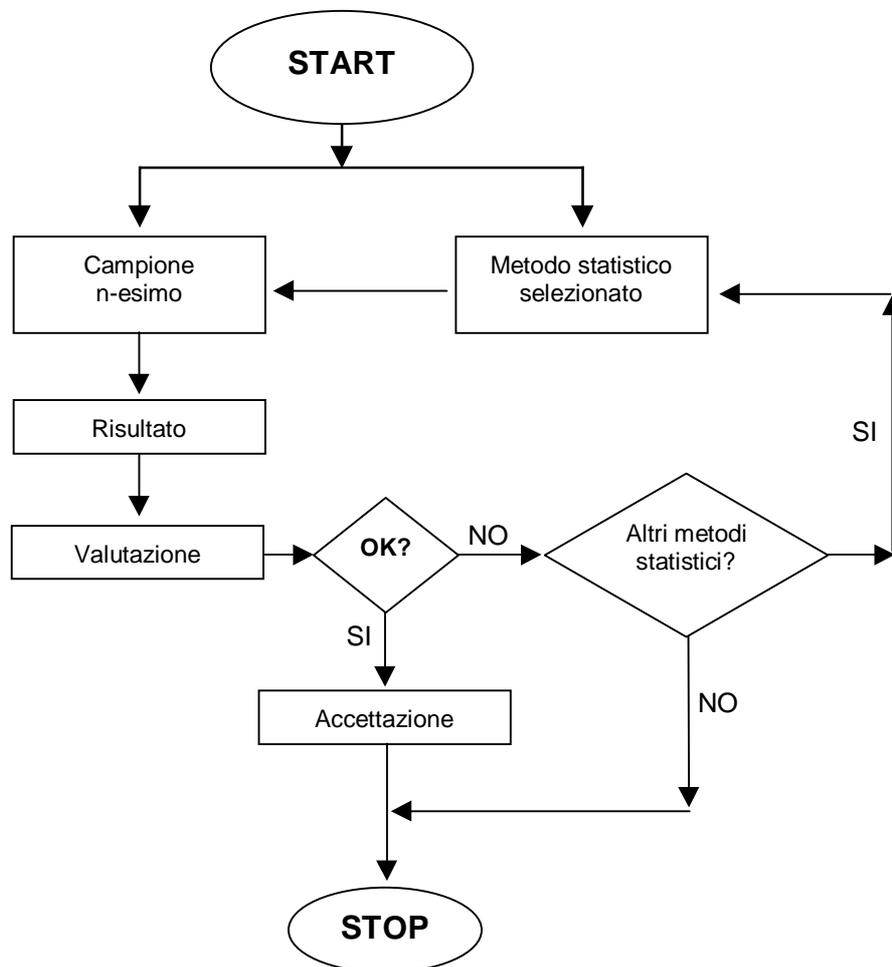


Figura 3. Modello di procedure per l'applicazione dei metodi statistici

Per effettuare la valutazione si predispongono test di significatività⁴ specifici, sia del metodo d'analisi utilizzato, sia della dimensione del campione. Il livello di probabilità d'accettazione, va da sé, è funzione degli obiettivi che ci si propone.

⁴ Tutti i test statistici si basano sul seguente principio fondamentale: si deve comparare una popolazione sconosciuta con una popolazione conosciuta od ipotetica. Per effettuare la comparazione si pone l'ipotesi zero, secondo la quale le due popolazioni coincidono. L'ipotesi viene posta in funzione di essere rigettata, cioè: le popolazioni non coincidono, ma una differisce dall'altra secondo un certo livello di probabilità. Per l'approfondimento dell'argomento si rimanda a testi di inferenza statistica, tema che in questa sede viene solo implicitamente introdotto.

1. 4. Forma grafica

L'operatore ha il primo grossolano impatto conoscitivo sia della concentrazione di un inquinante in atmosfera che del suo andamento temporale, in forma di percezione visiva, osservando il segnale prodotto dall'analizzatore che ne effettua il monitoraggio. Quando questo avviene in forma analogica, il segnale è rappresentato da una curva continua in un riferimento cartesiano – $RC(0, x, y)$ ⁵ – dove l'asse delle x è il tempo e quello delle y la concentrazione.

Il segnale in forma digitale è costituito da una sequenza temporale di numeri che è conveniente posizionare in un $RC(0, x, y)$; unendo i punti con una linea continua si riproduce il segnale.

Il segnale riportato in RC si chiama grafico. Il grafico rappresenta l'andamento della concentrazione media (es. oraria) dell'inquinante nel periodo temporale prefissato (un mese, un anno, sei mesi, ecc.). Sull'asse delle ascisse (x) è mostrato il periodo temporale, sull'asse delle ordinate (y) il valore di concentrazione media.

Il grafico è lo strumento preliminare a qualunque successiva analisi. Anche in questa forma iniziale – si noti che in questa fase i dati non sono né accettati, né validati – il grafico dà la chiave di lettura del monitoraggio avvenuto, evidenziando, per primo, eventuali guasti o interruzioni (grafico con buchi) in corrispondenza dell'interruzione del segnale, o picchi dei livelli di concentrazione anomali rispetto a quelli attesi.

A titolo esemplificativo, sono riportati alcuni grafici dei livelli di concentrazione degli inquinanti, rilevati in continuo da analizzatori automatici a tempo di integrazione di un'ora durante il mese di gennaio del 1994 in una stazione urbana.

Per ogni inquinante è stata disegnata anche la sua tendenza di comportamento nel periodo in studio, approssimando i dati, sottostanti i relativi grafici, con polinomi matematici di grado superiore al primo. La dimensione prevista di valori di concentrazione per l'insieme di ciascun inquinante, assumendo pari a 31 giorni il periodo di sua registrazione oraria, era di 744 dati.

Nelle Figure 4 e 5 sono riportati, rispettivamente, i grafici del monossido di carbonio (CO) e dell'ozono (O_3); nelle Figure 6 e 7, quelli del metano (CH_4).

Il CO (Figura 4) mostra un andamento con cicli ben definiti e con tendenza stazionaria intorno alla media mensile ($m = 3,9$).

L' O_3 (Figura 5) presenta anch'esso una certa ciclicità, con distacco netto tra comportamento notturno e diurno con tendenza ad un abbassamento dei livelli di concentrazione nella seconda decade del mese, periodo durante il quale persistono bassi livelli di concentrazione durante le ore di insolazione.

⁵ Il riferimento cartesiano si ottiene introducendo nello spazio degli assi ortogonali, aventi una comune origine, su ognuno dei quali viene fissata la rispettiva unità di misura. Quando la peculiarità della rappresentazione lo permette, è opportuno fissare la stessa unità di misura su ogni asse che, se uguale all'unità, fa parlare di riferimento cartesiano monometrico. Il riferimento cartesiano associa ad ogni punto dello spazio una n -pla ordinata di numeri, ognuno dei cui elementi corrisponde ordinatamente agli assi introdotti. Nel caso in cui si abbiano i soli assi ortogonali x ed y , caso bidimensionale, detto RC il riferimento che stabilisce la corrispondenza tra il generico punto P dello spazio, si ha:

$$P \xleftrightarrow{RC} (x, y)$$

cioè, ad ogni punto del piano viene associata una coppia ordinata di numeri e viceversa.

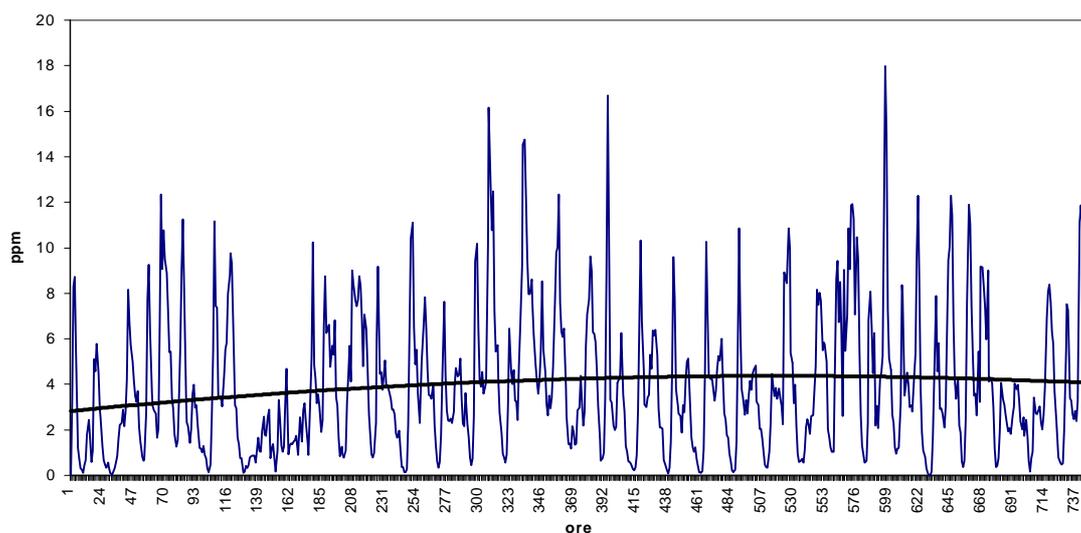


Figura 4. Trend orario delle concentrazioni di CO (gennaio 1994)

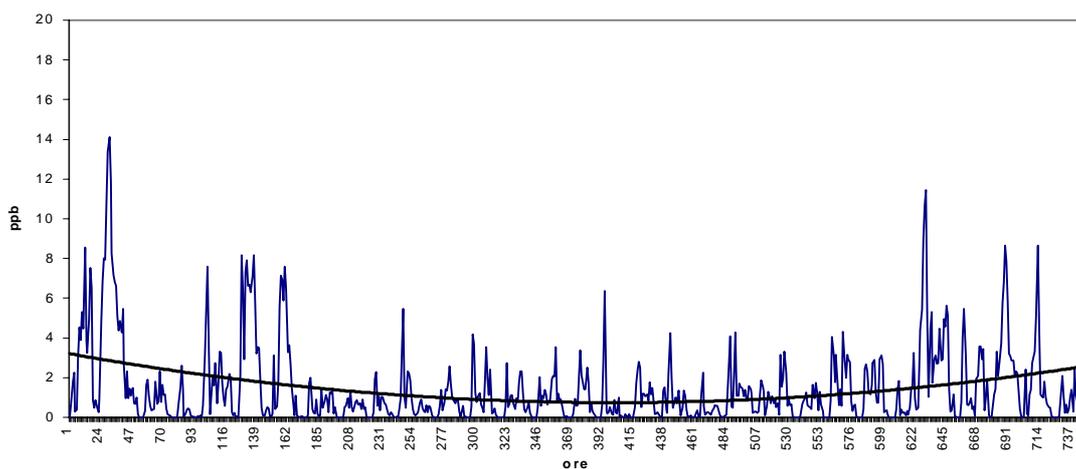


Figura 5. Trend orario delle concentrazioni di O₃ (gennaio 1994)

Nella Figura 6 il grafico del CH₄ è discontinuo con larghe serie di buchi. E' impossibile capire la tendenza di comportamento dell'inquinante durante il mese. E' erroneo tracciare qualunque linea di tendenza che, nella figura, è presente al solo scopo di evidenziare a quale conclusione porterebbe la sua interpretazione. Se, infatti, si accettasse, si dovrebbe concludere che l'inquinante presenta una tendenza mensile di persistenza sostenuta nei giorni estremi del mese con una tendenza verso bassi livelli durante la parte centrale del mese.

Praticamente, in questo caso, il grafico mette in evidenza una sospensione della registrazione, dovuta, probabilmente, a motivi tecnici legati all'apparecchiatura di rilevamento.

Dal punto di vista statistico la situazione è critica e suggerisce l'approfondimento di specifici studi di *data filling* (3, 4).

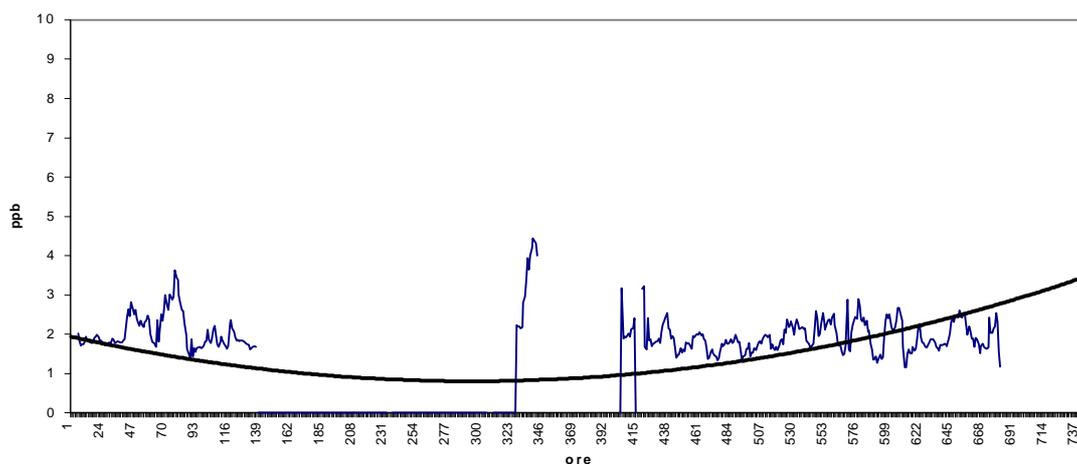


Figura 6. Registrazione oraria delle concentrazioni di CH₄ (gennaio 1994)

In Figura 7, è rappresentato l'andamento dello stesso inquinante per un periodo dello stesso mese (18-27 gennaio 1994) in cui i dati disponibili erano continui. In questo periodo, analogamente ad altri periodi del mese di registrazioni senza buchi, si possono estrapolare informazioni sul comportamento del CH₄ che, dalla grafica in esame, presenta una tendenza stazionaria con oscillazioni intorno alla media ($m = 1,9$).

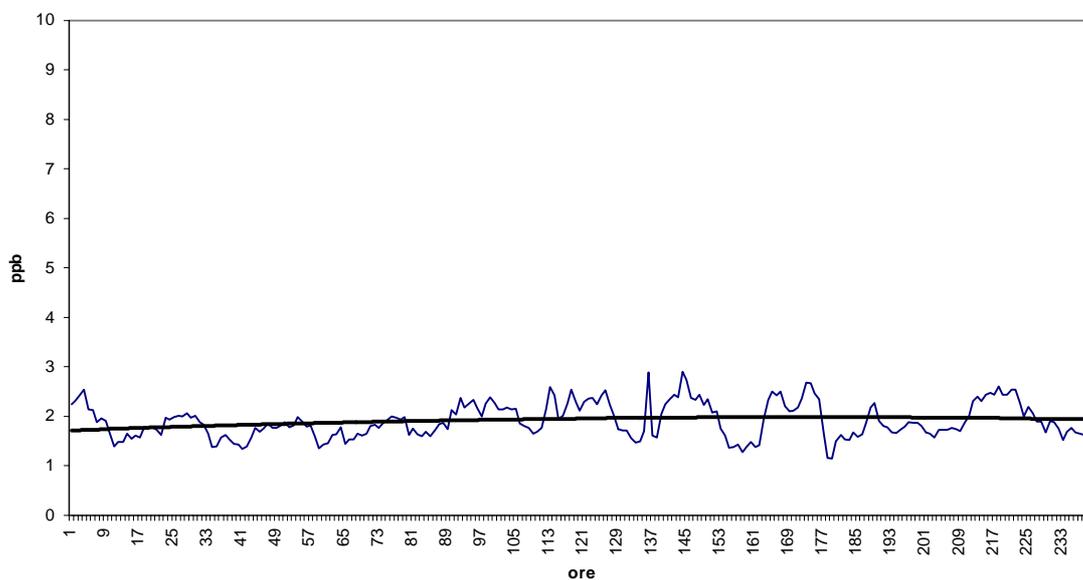


Figura 7. Trend orario delle concentrazioni di CH₄ (18-27 gennaio 1994)

Come risulta da quanto finora detto, la rappresentazione grafica è per l'operatore un semplice ma potente strumento di lavoro che gli permette sia di tradurre in forma immediatamente visibile ed interpretabile l'andamento di una serie storica di dati, o valori di concentrazioni di inquinanti, sia di effettuare il confronto spazio-tempo reale tra le serie temporali di valori di concentrazione di contaminanti e tra le serie temporali di altre variabili aleatorie registrate in loro concomitanza quali, per esempio, quelle meteorologiche; infine, è necessario sottolineare che l'analisi 'a vista da esperto' del trend del grafico è fortemente consigliata quando si devono scegliere i metodi di calcolo per effettuare previsioni, metodi la cui applicazione è condizionata dalla struttura matematica dei dati che sostengono il grafico.

Periodicità, stazionarietà, tendenza, persistenza, ecc., degli inquinanti, infatti, possono essere individuate dall'analisi dei grafici e poste come ipotesi di lavoro per lo sviluppo metodologico e l'applicazione dei metodi di calcolo (23), al fine di verificare in forma oggettiva le ipotesi di lavoro.

Nei dati che sostengono questo tipo di grafici è, infatti, contenuta tutta l'informazione sul fenomeno. Si tratta di estrarre metodologicamente informazione da questi dati, per conoscere il fenomeno nelle sue varie manifestazioni. I metodi matematici della statistica, fondati sul calcolo della probabilità, permettono quest'estrazione.

2. ELEMENTI DI STATISTICA

I metodi matematici della statistica sono fondati su un'esplicita formalizzazione dei dati mediante un modello concettuale o matematico. La formalizzazione consiste, in generale, nel rappresentare la matrice dei dati (Appendice B) in un determinato spazio matematico. Le proprietà strutturali di tale spazio condizionano, conseguentemente, i metodi di analisi che possono essere applicati allo studio del fenomeno.

Lo schema di formalizzazione, alla base della formulazione dei criteri statistici da utilizzarsi per il controllo della qualità dell'aria, è di tipo probabilistico. L'assunzione è conforme con la natura dei dati ottenuti nel monitoraggio, in quanto variabili aleatorie di un processo aleatorio continuo alle quali è associata una certa probabilità di presentarsi in un intervallo che può assumere infiniti valori compresi tra zero ed uno.

Si noti che uno spazio vettoriale non probabilizzato porrebbe le fondamenta per una statistica descrittiva, in contrapposizione a quella probabilistica; in uno spazio vettoriale non probabilizzato, il risultato dell'applicazione dei metodi descrive la posizione del fenomeno.

Media, mediana e percentile, strumenti di calcolo per effettuare il controllo della qualità dell'aria sulla base dei dati del monitoraggio di un inquinante, possono inizialmente essere applicati indifferentemente nell'uno o nell'altro spazio, in quanto, come risultato dell'applicazione, si ottiene la posizione del fenomeno. Ma se si vogliono ottenere altre informazioni quali, per esempio, la forma della distribuzione dei dati che sostengono il fenomeno, l'affidabilità di un risultato ottenuto, lo scostamento del fenomeno dalla norma, ecc., si deve ipotizzare che questi strumenti di calcolo si muovono in uno spazio probabilistico.

In questa ottica, questi strumenti di calcolo verranno collegati, con metodo di induzione, alla curva di probabilità.

2.1. Indici di valore centrale

2.1.1. Media aritmetica e ponderata

La media aritmetica, m , di un campione è data da:

$$m = \sum_{i=1}^n x_i/n$$

ovvero dalla sommatoria (indicata con il simbolo Σ) per i che va da 1 a n , delle modalità del carattere osservato nel collettivo in esame, divisa per il numero n delle modalità, x essendo la modalità del carattere (concentrazione), osservata nell' i -esima unità del collettivo in esame, mentre i è l'indice di posizione che descrive l'intero collettivo di ampiezza n .

Se si vuole tenere conto esplicitamente della frequenza v con cui ciascuna modalità si presenta, si ottiene l'espressione, analoga alla precedente:

$$m = \sum_{j=1}^k x_j v_j/n$$

dove k è il numero delle modalità differenti assunte dalla variabile (concentrazione) x nel collettivo in questione e j è il numero della frequenza relativa a quel carattere. Tale media aritmetica si chiama media ponderata.

La media aritmetica corrisponde, in termini geometrici, al baricentro dei punti-massa rappresentativi della distribuzione considerata. Può essere, quindi, assunta come indice di posizione.

2.1. 2. Mediana

Si definisce mediana quel valore tale che la somma delle frequenze relative alle modalità non inferiori a x è pari al 50% del totale e, simmetricamente, la somma delle frequenze relative alle modalità non superiori a x è pari al 50%.

Si definisce concentrazione mediana c_m il valore centrale della sequenza ordinata. Tale valore divide la sequenza ordinata in parti uguali.

Detto i l'indice che scorre i numeri, detto n il numero dei dati, la concentrazione mediana c_m occupa, rispettivamente il posto:

- per n pari, $i = n/2$;
- per n dispari, $i = (n+1)/2$.

La mediana è un quantile q_a di ordine $1/2$ (corrispondente al 50%):

- per $a = 1/4$ si hanno i quartili: $1/4, 2/4, 3/4$;
- per $a = 1/10$ si hanno i decili: $1/10, 2/10, 3/10, 4/10, 7/10, 8/10, 9/10$;
- per $a = 1/100$ si hanno i centili: $1/100, 2/100, \dots, 97/100, 98/100, 99/100$.

2. 2. Percentile

Dato un numero N totale di campionamenti, ordinati i valori delle concentrazione in modo crescente, si definisce k -esimo percentile c_k (2), il valore di concentrazione che nella sequenza occupa il posto:

$$k \times n / 100$$

Esso soddisfa le condizioni:

- a) la Σ della frequenza associata ai valori di concentrazione minori o uguali alla concentrazione c_{i-1} , deve essere $< k \times n / 100$;
- b) la Σ della frequenza associata ai valori di concentrazione maggiori o uguali a c_{i-1} , deve essere $> k \times n / 100$ (Appendice C).

2. 3. Variabilità del carattere

2. 3. 1. Varianza

La variabilità del carattere, ovvero, l'attitudine della concentrazione ad assumere modalità differenti nel collettivo osservato è la varianza.

La varianza è rappresentata dalla formula:

$$\sigma^2 = \sum_{i=1}^n (x_i - m)^2 / n$$

dove x_i è il generico elemento che descrive il collettivo, n il numero degli elementi e m è la media aritmetica:

$$m = \sum_{i=1}^n x_i / n$$

Esplicitando le frequenze si ottiene:

$$\sigma^2 = \sum_{j=1}^n (x_j - m)^2 v_j / n$$

dove v_j è la frequenza associata a x_j .

2. 3. 2. Deviazione standard

Estraendo la radice quadrata della varianza si ottiene la deviazione standard, indicata dal simbolo σ . Per definizione:

$$\sigma = +\sqrt{\sigma^2}$$

esplicitando:

$$\sigma = +\sqrt{\left[\sum_{i=1}^n (x_i - m)^2 / n \right]}$$

2. 3. 3. Coefficiente di variazione

Indicato con il simbolo v , il coefficiente di variazione è dato dal rapporto tra la deviazione standard e la media:

$$v = \sigma / m$$

e permette, tra l'altro, di confrontare rapidamente la variabilità della distribuzione di due o più serie di concentrazioni.

2. 3. 4. Errore standard sulla media

L'errore standard sulla media è dato da: σ / n

2. 4. Dagli indici statistici ai parametri statistici

I valori che rappresentano efficacemente i dati iniziali sono chiamati indici statistici. Un indice statistico è tanto più efficace quanto:

- più riesce a riassumere il contenuto informativo dei dati iniziali con la minore perdita d'informazione;
- meglio si presta ai calcoli ed ad ulteriori test.

Gli indici statistici più efficaci sono la media e la varianza: il massimo contenuto di informazione si ha utilizzando l'una e l'altra.

Da quanto detto, segue che gli indici statistici che sintetizzano molte caratteristiche delle distribuzioni statistiche o di variabili statistiche sono quelli:

- di *posizione*: ordine di grandezza di N valori;
- di *dispersione*: variabilità degli stessi N valori.

Gli indici statistici (media, mediana, percentile) sono indici di posizione. Tramite essi è sempre possibile dare una prima informazione sul fenomeno.

Attraverso uno schema di formalizzazione è possibile trasformare gli indici statistici in parametri⁶ statistici (Figura 8).

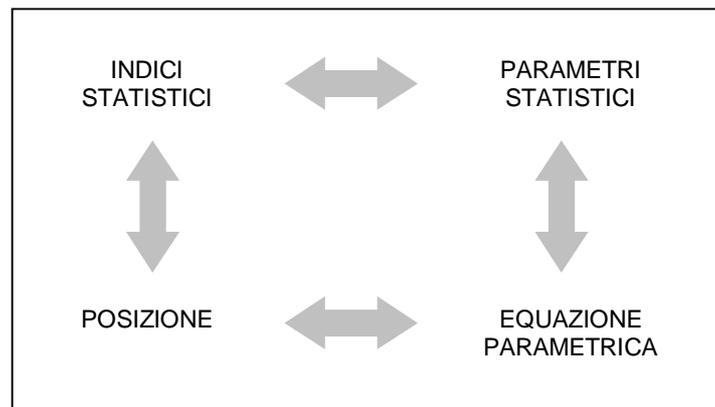


Figura 8. Schema di formalizzazione: dagli indici statistici ai parametri statistici

2. 5. Curva delle distribuzioni normali

Sia:

$$y = f(x) = (1/\sigma\sqrt{2\pi}) e^{-1/2 [(x-\mu)/\sigma]^2} \quad [1]$$

la funzione esplicita definita in RC con:

x = variabile indipendente;

y = variabile dipendente;

$\pi = 3,14$, costante;

e = 2,718281 base dei logaritmi naturali, anche chiamati neperiani;

μ = costante arbitraria, detta media in una popolazione;

σ = deviazione standard (o scarto tipo).

⁶ In matematica, si chiama parametro ogni costante arbitraria che compare nell'equazione di una curva e al cui variare si ottiene la famiglia di curve. L'equazione si chiama equazione parametrica. Le equazioni parametriche esprimono l'andamento di una o più funzioni per mezzo di uno o più parametri o variabili ausiliarie. Detto p il parametro, l'equazione $y = f(p)$ esprime la coordinata y del punto sulla retta di dipendenza del parametro p. La media aritmetica viene assunta come parametro dell'equazione.

La funzione è continua con x che varia da meno infinito a più infinito e in corrispondenza a tale variazione la y assume valori tali che in RC si disegna una curva a forma di campana.

La funzione è completamente caratterizzata dai due parametri μ e σ ordinatamente media e scarto tipo, che, rispettivamente, determinano la posizione e l'ampiezza della curva in RC. Dal momento che i parametri dell'equazione sono arbitrari, esiste un'infinità di curve.

In Figura 9 è riportata una famiglia di curve, osservando ognuna delle quali si vede che più σ aumenta, più la curva si appiattisce. Inoltre, risulta che tanto più piccolo è il valore della deviazione standard, tanto più i valori di concentrazione sono intorno alla media μ . La deviazione standard è quindi determinante come misura di dispersione.

Se poniamo:

$$(x-\mu)/\sigma = c$$

la funzione precedente è [1] diventa:

$$y = f(c) = (1/\sqrt{2\pi}) e^{-c^2/2} \quad [2]$$

cioè, una funzione standardizzata con media nulla e scarto tipo unitario.

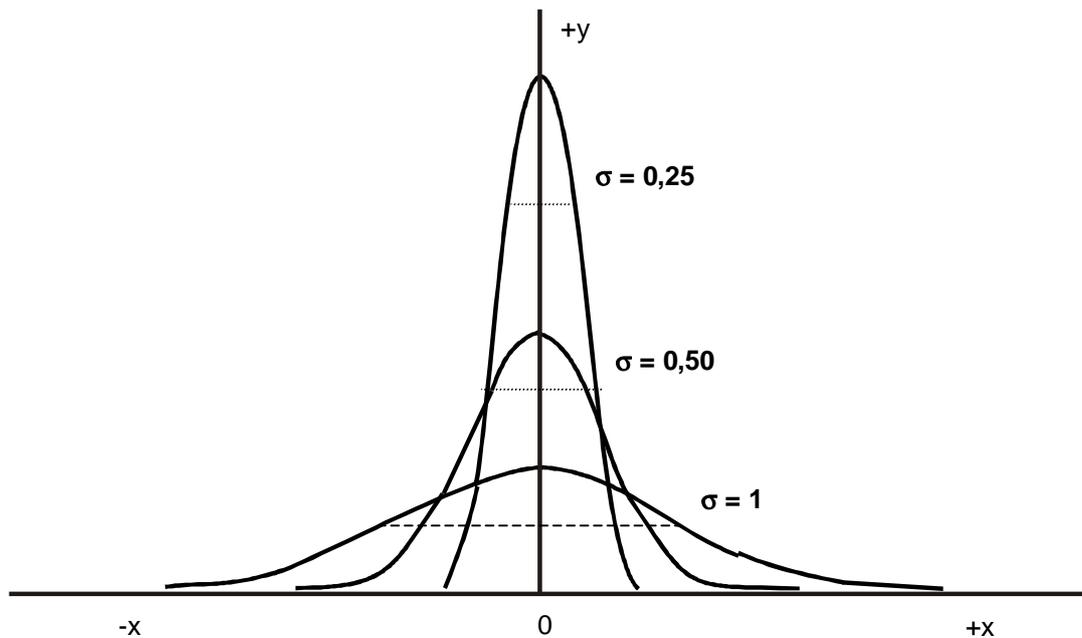


Figura 9. Distribuzione normale con differenti scarti tipo

Osservando la Figura 9, si nota che la curva

- 1) è simmetrica rispetto all'asse delle x ;
- 2) presenta il suo massimo in $x = 0$.

Tale massimo (y , μ/c) coincide con il valore medio, la mediana e la moda (massima frequenza). Si dimostra che:

$$\mu = 1/\sqrt{2}$$

è l'errore quadratico medio, mentre differenziando due volte la curva otteniamo due punti d'inflexione simmetrici, e, precisamente:

$$\begin{aligned}x_1 &= +1/2^{-1/2}\pi \\x_2 &= -1/2^{-1/2}\pi\end{aligned}$$

Conosciuta la media μ , la curva può essere facilmente costruita utilizzando la tavola dei logaritmi.

La funzione (1), o la sua standardizzata, ottenute sviluppando teorie fondate sul calcolo delle probabilità (2, 6), sono note come curva di Gauss-La Place o curva delle distribuzioni normali, intendendosi per convenzione, che l'attributo normale sta ad indicare che si tratta di un tipo di distribuzione che si presenta più frequentemente di altre.

Proprietà della curva: media, moda e mediana coincidono.

Proprietà tipica delle distribuzioni normali è che i casi osservati cadono:

- il 68% nell'intervallo: media $\pm\sigma$;
- il 95% nell'intervallo: media $\pm 2\sigma$;
- il 99% nell'intervallo: media $\pm 3\sigma$.

2. 6. Stima dei parametri

Nello studio di una popolazione possono verificarsi i due seguenti casi:

- a) Data una popolazione (*universo*) con una certa distribuzione di probabilità, pur non conoscendo né varianza né media, è possibile conoscerne l'andamento.
Infatti, ripetendo n volte le osservazioni, si ottiene un campione casuale. Facendo in modo che le variabili casuali siano indipendenti ed identicamente distribuite, si ottiene il modello teorico e sulla base di queste informazioni si calcolano i parametri.
- b) Data una popolazione senza distribuzione di probabilità (popolazione *finita*) è possibile studiare il fenomeno non partendo dal modello con probabilità conosciuta, ma considerando un solo campione. La scelta del campione deve essere casuale; si parla, in questo caso d'indagine campionaria.

Si dimostra che tutte le popolazioni di campioni possono essere assunte come finite.

2. 6. 1. Parametri statistici: media e varianza

In una certa distribuzione di probabilità, si supponga che μ e σ^2 siano, rispettivamente il valore medio e la varianza (non conosciuti) della popolazione.

Sia:

$$m = \sum_{i=1}^n x_i/n$$

la media campionaria. S'indica con m la stima della media della popolazione (si ricorda che m nella statistica descrittiva è la media aritmetica); dove m indica il valore medio di un esperimento con n prove.

Sia M la media della distribuzione di m (descrittiva):

$$M(x)_m = \sum_{i=1}^n M(x_i)/n = \mu$$

Si dice che x è una stima non distorta, in quanto alcuni valori possono essere μ , ma in media sono μ .

Analogamente per la varianza:

$$\text{var}(x) = \sigma_x = \sum_{i=1}^n \text{var}(x_i)/n^2 = n\sigma^2/n^2 = \sigma^2/n$$

Supponendo che la distribuzione sia normale, questa risulta tanto più concentrata quanto più è grande il numero d'osservazioni. Ovvero, aumentando la precisione diminuisce la varianza.

Si evidenzia che la varianza di un estimatore non distorto è importante in quanto rappresenta la distanza quadratica media tra l'estimatore ed il parametro che si vuole stimare (μ).

Infine, confrontando la varianza rispetto alla mediana, indicata con π e la varianza rispetto alla media μ , la mediana risulta uno stimatore meno buono della media, essendo:

$$\mu\sigma^2/2 < \sigma^2/N$$

BIBLIOGRAFIA

1. Lessi O. *Corso di calcolo delle probabilità*. Padova: Metria Editore; 1990.
2. Aivazian S. *Etude statistique des dependences*. Mouscu: Editions Mir; 1970.
3. Efron B. Better bootstrap confidence intervals. *J. Amer. Statist. Ass.* 1987;82:170-85.
4. Hanna, SR. Confidence limits for air quality model evaluation, as estimated by Jackknife Resampling Methods. *Atmos. Environ.* 1989;23:1385-98.
5. Mura MC, Falleni F, Miguel Garcia JA, Valero F. A scatter Plot and QQ plot qualitative analysis for the assessment of tropospheric ozone by a surface urban station. *Boll. Geof.* 1999;18:39-46.
6. Mura MC. *La teoria di Hagen alla base della distribuzione normale dell'errore casuale e della sua propagazione*. Roma: Istituto Superiore di Sanità; 1988. (Rapporti ISTISAN 88/41).
7. Mura MC, Sellitri S, Soggiu ME. Concentrazioni di inquinanti chimici nell'atmosfera di Roma. Organizzazione ed analisi statistica dei dati. In: *Atti del II Convegno Nazionale di Fisica dell'Ambiente, ricerca monitoraggio prevenzione*. Cosenza, 19-22 ottobre 1995. p. 429-45.
8. Mura MC, Sellitri S, Soggiu ME. Metodi Statistici e strumenti informatici per facilitare l'interpretazione dell'inquinamento chimico dell'atmosfera. Alcuni esempi su Roma. In: *Atti del Convegno "L'informazione statistica per il governo dell'ambiente"*. Roma, 29-30 gennaio 1996. Supplemento dell'annuario statistico nazionale. *Ann. Stat.* 1996;10:361-8.
9. Pindyck RS, Rubinfeld DL. *Econometric models and economics forecast*. Singapore: McGraw-Hill Company; 1981.
10. Hamilton JD. *Econometria delle Serie Storiche*. Bologna: Monduzzi Editore; 1995.
11. Johnston J. *Econometrica*. Milano: Franco Angeli Editore; 1996.
12. Trenberth KE. Signal versus noises in the Southern oscillation. *Mon. Weather Rev.* 1984;112:326-32.
13. Box GPE, Jenkins GM. *Time series analysis: forecasting and control*. Revised edition. San Francisco: Holden-Day; 1976.
14. Jenkins GM, Watts DG. *Spectral analysis and its application*. San Francisco: Holden-Day; 1968.
15. Hernandez E, Mura MC, Bertolaccini MA. Modello predittivo delle concentrazioni di ossido di carbonio in una zona industrializzata. *Acqua Aria* 1987;6:713-6.
16. Mura MC, Fuselli S, Garcia Miguel JA, Valero F. Benzene, monossido di carbonio ed ozono nell'atmosfera di un'area di Roma-est. Relazioni statistiche in uno studio preliminare. *Boll. Geof.* 1995;22(1-2):61-70.
17. Hernandez E, Mura MC, Garcia R, Valero F. Valori estremi di aerosol piombo nell'atmosfera di Madrid. *Acqua Aria* 1986;7:685-7.
18. Valero F, Mura MC. Day-time and night-time concentrations of airborne iron particles in the atmosphere of Madrid. *Ann. Ist. Super. Sanità* 1991;27:281-8.
19. Mura MC, Jimenez JL, Munoz P, Gonzalez J. Daily surface ozone maximum concentrations at Tailarte, Canary islands. *Ann. Ist. Super. Sanità* 1995;31:363-7.
20. Mura MC, Fuselli S, Garcia Miguel JA, Soggiu ME, Valero F. Stochastic approach to study behaviour of atmospheric pollutants in the urban area of Rome. *Sci. Total Environ.* 1995;171:151-4.
21. Mura MC, Gori Giorgi C, Pellegrini MF. Il modello spettrale per caratterizzare gli inquinanti gassosi dell'atmosfera. *Ann. Ist. Super. Sanità* 1990;26: 209-14.

22. Mura MC, Bielli G, Garcia Miguel JA, Soggiu ME, Valero F. Monossido di carbonio nell'atmosfera di una zona di Roma. *Boll. Geof.* 1994;17:55-62.
23. Blackmann R B, Tukey JN. *The measures of power spectra, from the point of view of communications engineering.* New York: Dover Publications; 1958.

APPENDICI

APPENDICE A

Tecniche grafiche

Il disegno dei modelli e delle procedure, riportati nelle figure, è effettuato con strumenti tipici dell'ingegneria che permettono di seguire il flusso logico delle operazioni che si devono effettuare. Tale flusso è rappresentato dalla direzione delle frecce sulle linee che collegano i blocchi.

Si tratta di procedure schematizzate con tecnica del diagramma di flusso (o *flow chart*) – tipica della ricerca operativa – tramite la quale le attività e/o le funzioni vengono descritte inquadrando in figure geometriche piane, ognuna delle quali ha un preciso significato simbolico.

Nel disegno del *flow chart* merita particolare attenzione il controllo logico delle fasi della procedura (geometricamente una losanga): questo può essere rappresentato da una domanda cui generalmente si dà risposta binaria, sì o no, in base alla quale si segue una o l'altra delle direzioni che portano ad attività o a funzioni tra di loro diverse.

Essendo il *flow chart* uno strumento di routine per i calcolatori elettronici, la pratica suggerisce di utilizzare 'la losanga' secondo una logica binaria; non escludendo che si possa utilizzare anche una logica ternaria (terza risposta: l'incertezza). Incidentalmente, qualche anno fa c'era stato un tentativo, per la verità senza grande successo, di introdurre calcolatori elettronici fondati su base di logica ternaria.

APPENDICE B

Matrice

Si dice matrice una tabella che contiene n successioni di m elementi ciascuna; n rappresenta il numero delle righe della matrice ed m quello delle colonne.

Ogni elemento della matrice è rappresentato dal simbolo $c_{i,j}$, dove i e j indicano che l'elemento c si trova, rispettivamente all'orizzontale i -esima riga ed alla verticale j -esima colonna. Ciascuno dei due indici varia: i da 1 ad n ; j da 1 a m .

La dimensione della matrice (M) è data da $(n \times m)$. Se $n = m$, la matrice si dice quadrata; se n è diverso da m , la matrice è rettangolare. Se $n = 1$ ed $m > 1$; oppure, se $n > 1$ e $m = 1$, si parla di vettori. La dimensione di un vettore (V) è $(n \times 1)$ oppure $(1 \times m)$. Nel primo caso, in pratica, si ha una M ad una colonna; nel secondo M ad una riga.

Da M di qualunque dimensione, purché sia n sia m non siano contemporaneamente uguali ad uno, si possono estrarre sub matrici, fra le quali V , ovvero M uni-dimensionali. Cioè:

$$M (\underline{m} \times \underline{n})$$

dove $\underline{m} < m$ e/o $\underline{n} < n$.

Esempio

Matrice rettangolare $M (n \times m)$, con $n=3$ ed $m=5$

$$\begin{array}{ccccc} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} & c_{1,5} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} & c_{2,5} \\ c_{3,1} & c_{3,2} & c_{3,3} & c_{3,4} & c_{3,5} \end{array}$$

Matrice quadrata $M (m \times n)$, con $n=3$ ed $m=3$

$$\begin{array}{ccc} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,1} & c_{3,2} & c_{3,3} \end{array}$$

Vettore $V (1 \times m)$, con $n=1$ ed $m=3$

$$c_1 \quad c_2 \quad c_3$$

Vettore $V (m \times 1)$, con $n=3$ ed $m=1$

$$\begin{array}{c} c_1 \\ c_2 \\ c_3 \end{array}$$

Scambiando le righe con le colonne di $M (m \times n)$, si ottiene la matrice T , che è la sua trasposta. Se $m=5$ ed $n=3 \rightarrow T (n \times m)$ è:

$$\begin{array}{ccc} c_{1,1} & c_{2,1} & c_{3,1} \\ c_{1,2} & c_{2,2} & c_{3,2} \\ c_{1,3} & c_{2,3} & c_{3,3} \\ c_{1,4} & c_{2,4} & c_{3,4} \\ c_{1,5} & c_{2,5} & c_{3,5} \end{array}$$

Elementi della matrice sono i numeri. Come detto nel paragrafo 1.1.2. questi possono essere unità statistiche quali, per esempio, valori di concentrazioni di inquinanti, le loro medie orarie, giornaliere, mensili, ecc., ma anche correlazioni tra due o più inquinanti, varianze, ecc. Notevole significato assume la matrice di covarianza, che riveste un importante significato statistico, per l'approfondimento del quale si rimanda a testi specializzati.

Per l'operatore che effettua il controllo della qualità dell'aria, la matrice diventa uno strumento prezioso nelle fasi di tabulazione e di presentazione dei risultati di elaborazioni specifiche anche ai fini dell'organizzazione dei dati in computer, in quanto ogni elemento m è individuato da un puntatore (i, j) e viceversa.

APPENDICE C

Percentile

Dato un insieme di n valori di concentrazioni di un contaminante, si voglia calcolare il valore di concentrazione che cade nel 95° percentile. Supposto $n=110$ il numero dei valori di concentrazione, il valore di concentrazione cercato si trova nella posizione:

$$C_{95} = 95 \times 110/100 = 104,5$$

Per trovare il valore di concentrazione in questo percentile, l'insieme è preventivamente distribuito in classi di frequenze in ordinamento naturale crescente e riportato in una tabella che include: ordine, modalità (valore di concentrazione), frequenza della modalità, frequenza cumulata. Nella tabella esemplificativa, l'insieme di n valori di concentrazione si distribuisce in 17 classi di frequenza. La frequenza totale cumulata è, ovviamente, uguale ad n .

Nell'esempio il valore 104,5 è compreso tra le frequenze cumulate, rispettivamente 103 e 106, corrispondenti al 13° e 14° posto nell'ordinamento naturale. Al numero di ordine 13 si trovano le concentrazioni di valore 150, mentre al numero d'ordine 14 quelle di 170. Frequenze relative: 5 e 3.

Verifichiamo le 2 condizioni della definizione (paragrafo 2.2):

- a) la Σ delle frequenze associate a concentrazione di posto 13° è 103;
 - b) la Σ delle frequenze associate a concentrazione di posto 14° è 106;
- risulta, pertanto, $103 < 104,5 < 106$.

Le condizioni sono verificate, allora il valore del 95° percentile è la concentrazione di valore 170.

Esempio

Ordine naturale N	Modalità c_i	Frequenza f_i	Frequenza cumulata Σf_i
1	10	1	1
2	20	3	4
3	40	3	7
4	50	7	14
5	60	9	23
6	70	10	33
7	80	13	46
8	90	16	62
9	100	13	75
10	110	9	84
11	120	7	91
12	140	7	98
13	150	5	103
14	170	3	106
15	200	2	108
16	240	1	109
17	250	1	110

*Direttore dell'Istituto Superiore di Sanità
e Direttore responsabile: Giuseppe Benagiano*

*Coordinamento redazionale:
Paola De Castro e Sandra Salinetti*

*Stampato dal Servizio per le attività editoriali
dell'Istituto Superiore di Sanità, Viale Regina Elena, 299 - 00161 ROMA*

*La riproduzione parziale o totale dei Rapporti e Congressi ISTISAN
deve essere preventivamente autorizzata.*

Reg. Stampa - Tribunale di Roma n. 131/88 del 1° marzo 1988

Roma, marzo 2001 (n. 1) 1° Suppl.

*La responsabilità dei dati scientifici e tecnici
pubblicati nei Rapporti e Congressi ISTISAN è dei singoli autori*