

G. MASSIMETTI (*), G. BOZZI (**), P. CARLUCCI (**)
E. PASCULLI (***), F. LOGGINI (****)

SAS: UN SOFTWARE STATISTICO PER L'UTENTE FINALE, ANALIZZATO NELLA SUA UTILIZZAZIONE IN UNA INDAGINE PRELIMINARE SULL'ALCOLISMO IN PROVINCIA DI MASSA

Riassunto — Lo scopo di questo articolo è di illustrare le principali caratteristiche e lo schema funzionale SAS (Statistical Analysis System), un software statistico che, per la sua capacità di coniugare un'estrema semplicità d'uso ad una grande potenzialità di elaborazione, è particolarmente indicato per l'utente finale. Gli esempi esplicativi sono stati tratti da uno studio sull'alcolismo condotto nella zona di Massa e Carrara.

Abstract — *SAS: a statistical software for the final user.* The aim of this article is to describe the main characteristics and the functional scheme of SAS (Statistical Analysis System), a powerful statistical software suitable to the final user on account of its very friendly machine interface. The explanatory examples are taken from a study on alcoholism in the area of Massa and Carrara.

Key words — SAS Statistical Analysis System - alcolismo.

INTRODUZIONE

È noto come nelle indagini socio psichiatriche il ricercatore abbia spesso a che fare con una grande mole di dati dovuta all'alto numero di soggetti analizzati ed all'impiego di strumenti standardizzati ricchi di variabili; risulta quindi indispensabile l'utilizzazione del computer e di sistemi automatici di gestione ed elaborazione statistica delle informazioni.

(*) Istituto di Clinica Psichiatrica dell'Università di Pisa.

(**) Direzione Sanitaria Ospedale S. Chiara Pisa.

(***) U.O. di Psichiatria USL 2 Reg. Toscana.

(****) Facoltà di Farmacia dell'Università di Pisa, Cattedra di Microbiologia ed Igiene.

Il SAS (Statistical Analysis System) è un sistema di uso generale in grado di venire incontro alle necessità di chi ha molti dati da analizzare minimizzando i problemi connessi alla loro elaborazione e consentendo di focalizzare l'attenzione quasi unicamente sul problema specifico e sui risultati piuttosto che sul modo di ottenerli. Infatti, la presenza di programmi di utilità generale di utilizzo più flessibile rispetto a quelli disponibili sui sistemi operativi e la struttura modulare (vedi fig. 1) che permette un facile spostamento dell'utente dall'ambiente principale (SAS Base) a gli altri ambienti consente di evitare un preliminare apprendimento di linguaggi di programmazione, di package statistici e di programmi di utilità (BRACCI e LAFORENZA, 1985).

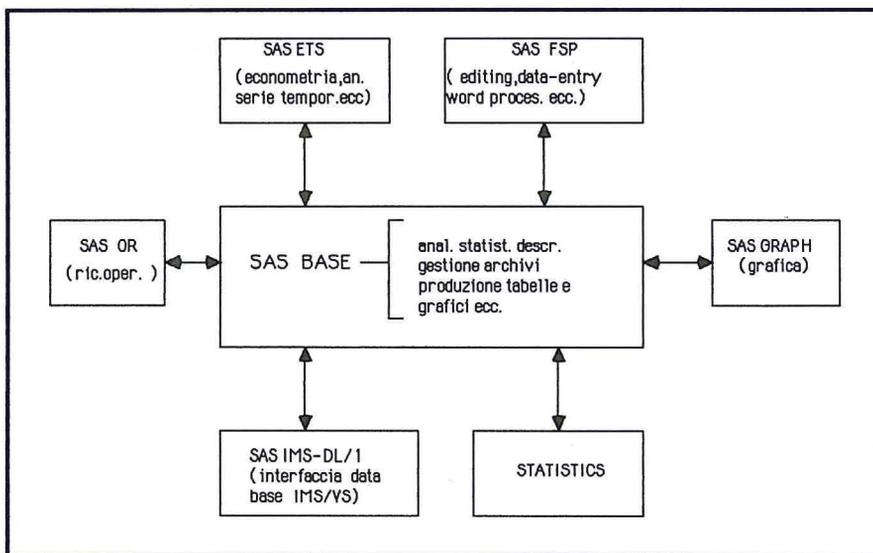


Fig. 1 - Struttura del sistema SAS.

Il sistema è costituito da un potente linguaggio e da una serie di librerie composte da procedure «preconfezionate» in grado di risolvere problemi specifici quali data-entry, gestione di archivi, produzione di output tabellare e grafico, elaborazioni statistico-descrittive, elaborazioni statistiche uni e multi-variate, analisi di serie temporali ecc.

L'estrema semplicità di utilizzazione da un lato e la grande po-

tenzialità di elaborazione dall'altro hanno favorito sia un costante allargamento della comunità degli utilizzatori che una diversificazione delle applicazioni originariamente quasi esclusivamente di tipo statistico (oggi le installazioni SAS nel mondo sono circa 7000 di cui circa 100 in Italia). La grande versatilità di questo software lo rende così adatto alle più svariate applicazioni tanto da poter essere utilizzato da statistici, ricercatori di marketing, biologi, sociologi, medici, informatici ecc. (PISANTI, 1985).

Il sistema SAS funziona: su IBM 370/30xx/43xx e compatibili in modalità batch ed interattiva sotto sistemi operativi OS, VM/CMS, DOS/VSE, TSO, su Digital Vax 11/7xx sotto sistema VMS, su Data General ECLIPSE serie MV sotto sistema AOS/VS ed è assistito in Italia dalla SINTAX (BRACCI e LAFORENZA, 1985).

In questo articolo cercheremo di descrivere il funzionamento ed i concetti base di questo software commentando alcuni programmi SAS utilizzati nell'ambito di uno studio epidemiologico sull'alcolismo attualmente in corso nell'area di Massa e Carrara; ricerca attuata mediante la collaborazione tra un gruppo di operatori della USL 2 e la Clinica Psichiatrica di Pisa.

L'hardware utilizzato per questa applicazione è l'IBM 3081 del CNUCE di Pisa ed il sistema operativo il CMS.

Dal momento che l'indagine è tuttora in corso, abbiamo dedicato all'analisi ed al commento dei risultati preliminari fin qui ottenuti solo una brevissima trattazione privilegiando invece l'aspetto divulgativo delle informazioni relative alle potenzialità ed all'uso del SAS.

SCHEMA OPERATIVO DEL SAS

Tutti gli utenti SAS hanno certamente almeno due cose in comune da gestire: un insieme di dati ed un insieme di interrogazioni su di essi. Come ogni altro linguaggio, il SAS ha un suo proprio vocabolario ed una sua propria grammatica con cui è possibile costruire le istruzioni dei programmi necessarie alla definizione dei dati ed alle relative interrogazioni. L'insieme del programma e delle istruzioni di controllo appropriate al sistema operativo in uso costituiscono un job (o sessione) SAS.

I programmi SAS presentano una struttura «a passi» di due diversi tipi: «Data Step» in cui vengono creati e mantenuti gli archivi SAS ed i «Proc Step» in cui vengono elaborati i dati.

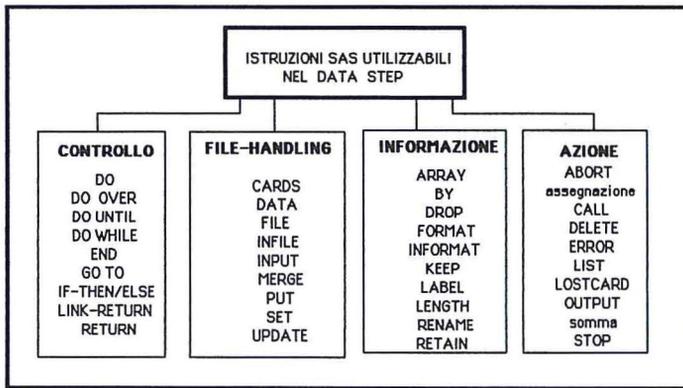
rettangolare in modo che le righe rappresentino le osservazioni e le colonne le variabili; nella fig. 2 è mostrato il dataset ottenuto dalla lettura della SAD (Scala di Autovalutazione per la Depressione) uno degli strumenti utilizzati nell'indagine sugli alcolisti costituito da 31 item esploranti la sintomatologia ansiosa e depressiva; in ogni riga ci sono i punteggi che ogni soggetto analizzato ha riportato nei 31 item della scala oltre alle variabili di identificazione ed a quelle (N e TOT) funzione delle variabili base.

Non ci sono limiti al numero dei dataset che un programma SAS può gestire. La lettura dei valori può avvenire sia da schede perforate che da qualsiasi supporto magnetico. In un Data Step è possibile inoltre combinare i dati di diversi dataset, etichettare i nomi ed i possibili valori delle variabili, costruire nuove variabili funzione di quelle di base (nella fig. 3 è mostrato il Data Step che calcola il totale SAD ottenuto sommando tutti gli item della SAD e dividendo il risultato per il numero totale degli item non missing N); a questo scopo si possono utilizzare istruzioni di controllo del tipo IF-THEN/ELSE e DO-END.

```
DATA A;
  ARRAY SAD (I) SAD1-SAD31;
  INPUT ( SEX COD FORM SAD1-SAD31 ) (formato);
  .
  N = 0.0; TOT = 0.0;
  DO OVER SAD;
    IF SAD NE . THEN DO;
      TOT = TOT + SAD;
      N = N + 1;
    END;
  END;
  TOT = TOT/N;
```

Fig. 3 - Data Step per il calcolo del Tot. SAD.

Nella fig. 4 sono riassunte tutte le istruzioni utilizzabili nel Data Step suddivise in istruzioni di controllo, di azione, di gestione di file (ad es. l'istruzione INPUT indica il formato), di informazione (ad es. l'istruzione non eseguibile ARRAY SAD(I) SAD1-SAD31 di fig. 3 definisce l'insieme di variabili SAD1-SAD31 della SAD).



CONTROLLO: Permettono il trasferimento del controllo da una parte all'altra del programma cambiando l'ordine di esecuzione delle istruzioni altrimenti rigidamente sequenziale (ad es. al verificarsi o meno di una condizione consentono l'esecuzione di una istruzione successiva, consentono il salto ad un'istruzione etichettata ecc.).

FILE-HANDLING: Consentono l'input, l'output e l'aggiornamento dei dataset SAS.

INFORMAZIONE: forniscono al SAS informazioni aggiuntive circa i dataset; ad es. definiscono insiemi di variabili che devono essere elaborate nello stesso modo, identificano variabili da escludere dall'analisi, specificano formati per la stampa di valori ecc.

AZIONE: In fase di creazione (DataStep) di un dataset SAS permettono di apportare modifiche ai dati quali appaiono in fase di input, selezionare solo certe osservazioni per idataset che si va creando, ricercare errori nei dati di input ecc.

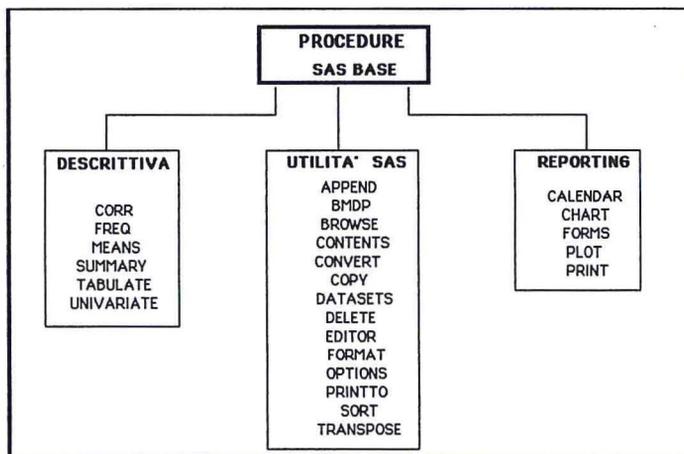
Fig. 4 - Istruzioni utilizzabili nel Data Step.

PROC STEP

Una volta che i dati sono stati letti in un dataset SAS è possibile analizzarli tramite le procedure SAS contenute in opportune librerie richiamabili nel Proc Step.

Le procedure SAS sono programmi che leggono dataset SAS, calcolano statistiche, stampano risultati e creano altri dataset SAS. Mentre nei Data Step è l'utente che crea i propri programmi utilizzando istruzioni SAS per definire e manipolare i dati, nei Proc Step i programmi sono già confezionati; per richiamarli è necessario solo fornire il nome ed eventuali ulteriori informazioni tramite l'utilizzazione di procedure information statement e/o l'aggiunta di opzioni all'istruzione Proc, la prima istruzione dello step. Le opzioni e le information statement servono a specificare analisi più dettagliate ed a personalizzare le procedure e dipendono dalla specifica procedura richiamata.

Nella fig. 5a sono indicate tutte le procedure del SAS-Base, nella fig. 5b quelle del SAS-Statistics; notare la vasta gamma di programmi per analisi statistiche descrittive ed uni e multi-variate e le numerose utility (sort, copy, ecc.) disponibili.



DESCRITTIVA: producono statistiche descrittive dei dati quali medie, deviazioni standard, frequenze, cross-tabulation ecc.

UTILITA': adibite allo svolgimento di funzioni diverse dall'analisi e la produzione di rapporti. Consentono l'interfacciamento con altri sistemi statistici, editing dei dati, definizione di formati di output, copie di dataset e librerie, ordinamento di dataset in base ai valori di una o più variabili ecc.

REPORTING: adibite alla produzione di rapporti, di liste dei dati, display organizzati di dati, grafici, istogrammi ecc.

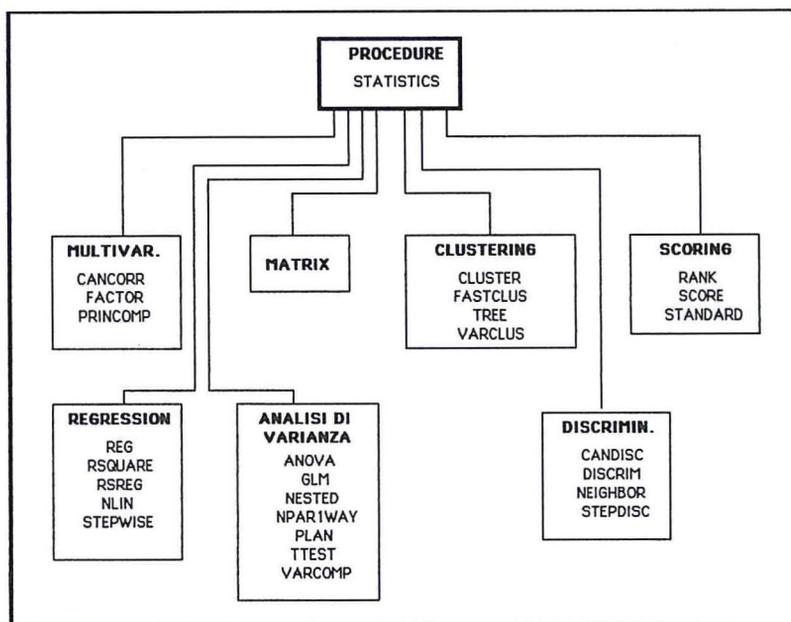
Fig. 5a - Procedure utilizzabili nel SAS - Base.

ESEMPIO ESPLICATIVO

Un esempio esplicativo della struttura di un job SAS può essere quello di fig. 6 in grado di fornire una statistica descrittiva su tutti gli item della SAD del campione utilizzato per l'indagine sugli alcolisti di cui abbiamo accennato sopra.

Il job è costituito da un Data Step e cinque Proc Step, una di tipo reporting (PRINT), due di tipo utilità (SORT, FORMAT) e due di tipo descrittivo (FREQ, UNIVARIATE).

Il Data Step inizia con l'istruzione DATA che chiede al SAS di creare un dataset di nome A, tramite l'istruzione INPUT viene poi



MULTIVAR.: procedure per l'analisi delle componenti principali, analisi fattoriale, analisi delle correlazioni canoniche.

MATRIX: è un vero e proprio linguaggio di programmazione per la definizione di eventuali nuove procedure statistiche non preesistenti nelle librerie SAS.

CLUSTERING: procedure per il cluster delle osservazioni o delle variabili di un dataset SAS. Possono essere ottenuti cluster gerarchici o disgiunti.

SCORING: procedure che permettono di produrre un output dataset trasformando i dati del vecchio dataset. Così ad es. è possibile standardizzare le variabili rispetto ad una data media ed una data deviazione standard, costruire nuove variabili come combinazioni lineari delle vecchie in accordo ad uno scoring dataset ecc.

REGRESSION: procedure per l'analisi di regressione con uno o più predittori.

ANALISI DI VARIANZA: procedure per l'analisi di varianza, covarianza, uni e multi-variate.

DISCRIM.: procedure per l'analisi discriminante.

Fig. 5b - Procedure utilizzabili nel SAS - Statistics.

dichiarato il formato dei dati ed i nomi delle variabili su cui vengono letti, nel caso specifico le variabili che andranno a costituire ogni osservazione sono SEX, COD, FORM ed i 31 item della SAD definiti su un insieme dimensionato (ARRAY) per facilitare la definizione e la generazione della variabile TOT nelle istruzioni di controllo successive. Il nome del file di input viene definito tramite l'istruzione INFILE; da notare che, utilizzando il sistema operativo VM/CMS è

```

CMS FILEDEF SAD DISK USL 2 B;
DATA A;
ARRAY SAD (1) SAD1-SAD31;
INFILE SAD;
INPUT ( SEX COD FORM SAD1-SAD31) (format)
( *8 *5 $1. $4. $3. *21 31*1.);
N = 0.0; TOT = 0.0;
DO OVER SAD;
  IF SAD NE . THEN DO;
    TOT = TOT + SAD;
    N = N + 1;
  END;
END;
TOT = TOT/N;
PROC SORT; BY COD;
PROC FORMAT;
VALUE $SES 1 = MASCHIO
          2 = FEMMINA;
PROC PRINT;
TITLE SAD LISTA DEI DATI;
VAR COD SEX SAD1-SAD31 N TOT;
FORMAT SEX $SES. ;
PROC FREQ;
TABLE SEX;
FORMAT SEX $SES. ;
TITLE FREQUENZE SUL SESSO;
PROC UNIVARIATE PLOT FREQ;
VAR SAD1-SAD31 TOT;
TITLE SAD DESCRITTIVA VARIABILI NUMERICHE;
RUN;

```

Fig. 6 - Esempio di struttura di job SAS.

necessario associare la DDname SAD all'effettivo file su disco su cui sono memorizzati i dati (il file di nome USL 2 B) tramite il comando CMS FILEDEF. Il resto delle istruzioni del Data Step serve a definire, per ogni osservazione letta, la variabile TOT ottenuta sommando i valori degli item SAD e dividendo il risultato per il numero degli item non mancanti (variabile N); il carattere «.» sta ad indicare dato mancante. Le due variabili generate, TOT ed N, vanno ad aggiungersi al set di variabili costituenti la singola osservazione.

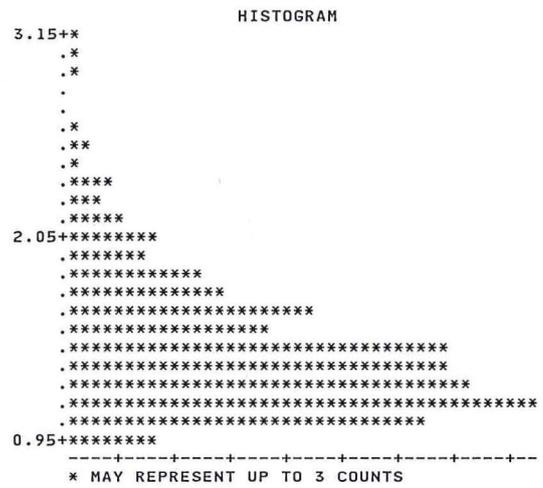
Il Data Step è seguito dal Proc Step relativo alla esecuzione della procedura SORT che organizza le osservazioni del dataset A in ordine crescente secondo il codice soggetto (vedi istruzione BY COD). La PROC FORMAT serve ad associare i nomi (MASCHIO e FEMMINA) ai valori della variabile SEX in tutti i Data Step e Proc Step che presentano l'istruzione FORMAT, così accade ad es. nella PROC PRINT successiva che visualizza la lista dei dati (fig. 2), ovvero tutte le osservazioni da cui è composto il campione. Le procedure FREQ ed UNIVARIATE producono semplici statistiche descrittive (fig. 7 e fig. 8) che includono: dettagli sui valori estremi delle variabili, quantili, tre plot che caratterizzano la distribuzione e tabelle di fre-

SAD DESCRITTIVA VARIABILI NUMERICHE

UNIVARIATE

VARIABLE=TOT

MOMENTS				QUANTILES(DEF=4)				EXTREMES	
N	833	SUM WGTS	833	100% MAX	3.12903	99%	2.51613	LOWEST	HIGHEST
MEAN	1.44113	SUM	1200.46	75% Q3	1.6129	95%	2.12903	1	2.6129
STD DEV	0.34373	VARIANCE	0.11815	50% MED	1.35484	90%	1.90322	1	2.64516
SKEWNESS	1.26877	KURTOSIS	1.95643	25% Q1	1.19355	10%	1.07871	1	3
USS	1828.33	CSS	98.3011	0% MIN	1	5%	1.03226	1	3.03226
CV	23.8514	STD MEAN	0.0119095			1%	1	1	3.12903
T:MEAN=0	121.006	PROB> T	0.0001	RANGE	2.12903				
SGN RANK	173681	PROB> S	0.0001	Q3-Q1	0.419354				
NUM ^= 0	833			MODE	1.12903				



#	BOXPLOT
1	*
1	*
1	
2	0
5	0
1	0
10	0
9	0
13	
24	
19	
35	
40	
65	+-----+
54	
102	+
101	*-----*
108	
124	+-----+
94	
24	

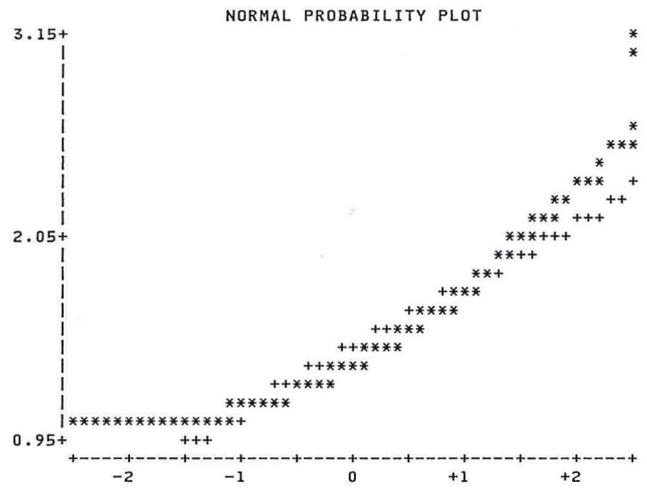


Fig. 7

SAD DESCRITTIVA VARIABILI NUMERICHE

UNIVARIATE

VARIABLE=TOT

FREQUENCY TABLE

PERCENTS															
VALUE	COUNT	CELL	CUM												
1	24	2.9	2.9	1.35484	35	4.2	51.0	1.64516	20	2.4	77.9	2.12903	2	0.2	95.2
1.03226	25	3.0	5.9	1.35714	1	0.1	51.1	1.67742	23	2.8	80.7	2.16129	2	0.2	95.4
1.03333	1	0.1	6.0	1.36667	1	0.1	51.3	1.70968	14	1.7	82.4	2.16667	1	0.1	95.6
1.06452	32	3.8	9.8	1.37037	1	0.1	51.4	1.73333	1	0.1	82.5	2.19355	7	0.8	96.4
1.06667	1	0.1	10.0	1.3871	22	2.6	54.0	1.74193	13	1.6	84.0	2.22581	1	0.1	96.5
1.09677	34	4.1	14.0	1.4	1	0.1	54.1	1.76667	2	0.2	84.3	2.25806	3	0.4	96.9
1.1	1	0.1	14.2	1.41379	1	0.1	54.3	1.77419	10	1.2	85.5	2.26667	1	0.1	97.0
1.12903	47	5.6	19.8	1.41935	42	5.0	59.3	1.80645	15	1.8	87.3	2.29032	3	0.4	97.4
1.13333	2	0.2	20.0	1.43333	1	0.1	59.4	1.83871	7	0.8	88.1	2.3	1	0.1	97.5
1.16129	32	3.8	23.9	1.44444	1	0.1	59.5	1.87097	11	1.3	89.4	2.32258	5	0.6	98.1
1.16667	2	0.2	24.1	1.45161	22	2.6	62.2	1.9	2	0.2	89.7	2.35484	4	0.5	98.6
1.19355	39	4.7	28.8	1.46667	1	0.1	62.3	1.90322	7	0.8	90.5	2.4	1	0.1	98.7
1.2	2	0.2	29.1	1.48387	32	3.8	66.1	1.93548	2	0.2	90.8	2.48387	1	0.1	98.8
1.22581	38	4.6	33.6	1.5	2	0.2	66.4	1.96774	5	0.6	91.4	2.51613	4	0.5	99.3
1.23333	2	0.2	33.9	1.51613	23	2.8	69.1	2	5	0.6	92.0	2.54839	1	0.1	99.4
1.25806	35	4.2	38.1	1.51852	1	0.1	69.3	2.03226	7	0.8	92.8	2.6129	1	0.1	99.5
1.26667	3	0.4	38.4	1.54839	12	1.4	70.7	2.06452	5	0.6	93.4	2.64516	1	0.1	99.6
1.29032	28	3.4	41.8	1.56667	2	0.2	70.9	2.06667	3	0.4	93.8	3	1	0.1	99.8
1.3	2	0.2	42.0	1.58064	16	1.9	72.9	2.09677	8	1.0	94.7	3.03226	1	0.1	99.9
1.32258	38	4.6	46.6	1.6129	21	2.5	75.4	2.1	1	0.1	94.8	3.12903	1	0.1	100.0
1.33333	2	0.2	46.8	1.63333	1	0.1	75.5	2.11111	1	0.1	95.0				

Fig. 8

quenza (vedi opzioni PLOT e FREQ nella istruzione PROC UNIVARIATE).

Riepilogando: una volta definito l'ARRAY SAD, SAS comincia a leggere i valori dal file su disco identificato dal comando CMS FILEDEF, i valori in lettura vengono attribuiti alle variabili secondo l'ordine indicato nell'istruzione INPUT. Terminato di leggere le variabili associate ad un soggetto, SAS crea nuove variabili secondo le istruzioni contenute nel Data Step (nel caso specifico TOT ed N); fatto ciò si ricomincia con il soggetto successivo finché ci sono dati nel file su disco. Quando i dati sono stati tutti letti ed organizzati nel dataset SAS di nome A vengono eseguiti i Proc Step.

RILEVAMENTI CON IL SAS SULL'ALCOLISMO IN PROVINCIA DI MASSA E CARRARA

Lo scopo dello studio sull'alcolismo è quello di chiarire i rapporti fra abuso alcoolico ed altre componenti psicologiche e psicopatologiche al fine di impostare mirati interventi di ordine preventivo, terapeutico e rieducativo.

Materiali e metodi

L'indagine, iniziata nel 1985 e tuttora in corso, è fondata sull'impiego di strumenti standardizzati quali: SAD (Scala di Autovalutazione della Depressione), VARA-TEST (Questionario per la Valutazione di Abuso di Alcoolici e di Rischio di Alcolismo), QTA (Questionario per Tipizzazione dell'Aggressività); SCL-90, strumento che indaga i vari aspetti psicopatologici della personalità; infine una scheda dello Stato di Salute stilata dal curante in cui vengono riportate eventuali diagnosi, durata di malattia, gravità e prognosi, presenza di inabilità.

Nella prima fase è stato esaminato un campione significativo della frazione montana di Forno (Massa) e di Gragnana (Carrara); la seconda verrà effettuata su un campione randomizzato della città di Massa.

Primi risultati

Il campione studiato è composto da 833 soggetti 440 dei quali

```

DATA A;
ARRAY V(J) V1-V40;
ARRAY SAD (J) SAD1-SAD31;
ARRAY VFT1 (I) V1 V2 V6 V13 V15 V16 V17 V22 V23;
ARRAY VFT2 (I) V24-V29 V39;
ARRAY VFT3 (I) V19-V21;
ARRAY VFT4 (I) V4-V12;
ARRAY VFT5 (I) V30 V31 V33;
ARRAY VFT6 (I) V35 V36;
ARRAY SFT1 (I) SAD1 SAD6 SAD10-SAD14 SAD16-SAD18
             SAD20 SAD29 SAD31;
ARRAY SFT2 (I) SAD1 SAD3 SAD9-SAD11 SAD13 SAD15
             SAD19 SAD22 SAD24 SAD26 SAD28 SAD31;
ARRAY SFT3 (I) SAD13 SAD16 SAD23 SAD25 SAD27 SAD30;
ARRAY SFT4 (I) SAD4 SAD5 SAD7;
ARRAY SFT5 (I) SAD2 SAD29;
ARRAY SFT6 (I) SAD14 SAD17 SAD20 SAD21 SAD30;
ARRAY NVAR (L) NVAR1-NVAR6;
ARRAY NSAD (L) NSAD1-NSAD6;
ARRAY NELVAR (L) NELVAR1-NELVAR6;
ARRAY NELSAD (L) NELSAD1-NELSAD6;
ARRAY ALLVFT (L) VFT1-VFT6;
ARRAY ALLSFT (L) SFT1-SFT6;
ARRAY FATTSAD (L) FATTSAD1-FATTSAD6;
ARRAY FATTVAR (L) FATTVAR1-FATTVAR6;

RETAIN NELSAD1 13 NELSAD2 13 NELSAD3 6
        NELSAD4 3 NELSAD5 2 NELSAD6 5;
RETAIN NELVAR1 9 NELVAR2 7 NELVAR3 3
        NELVAR4 9 NELVAR5 3 NELVAR6 2;
DO L = 1 TO 6;
  NSAD = 0; FATTSAD=0;
  DO I = 1 TO NELSAD;
    IF ALLSFT NE . THEN DO;
      FATTSAD=FATTSAD+ALLSFT;
      NSAD=NSAD+1;
    END;
  END;
  IF NSAD NE 0 THEN FATTSAD=FATTSAD/NSAD;
  ELSE FATTSAD= . ;
END;
PROC CORR;
VAR TV; WITH FATTSAD1-FATTSAD6;
TITLE CORRELAZIONI TOTVARA VS FATTORI SAD (TUTTO IL CAMPIONE);
PROC CORR;
VAR TV; WITH FATTSAD1-FATTSAD6; BY GR;
TITLE CORRELAZIONI TOTVARA VS FATTORI SAD (BY GR);
RUN;

```

Fig. 9 - Programma per il calcolo dei fattori SAD e VARA e delle correlazioni fra il totale VARA ed i fattori SAD.

di sesso maschile e 393 di sesso femminile, tutti di età compresa tra i 18 ed i 65 anni.

Gli alcolisti sono risultati in prevalenza uomini con 104 casi rilevati su un totale di 440 soggetti esaminati (23,63%); nelle donne si sono riscontrati 13 casi su 393 soggetti analizzati (3,3%).

La percentuale di alcolismo negli uomini risulta bassa nei soggetti della fascia d'età compresa tra i 18 ed i 34 anni, mentre appare elevata nelle fasce 35-49 (25,71%) e 50-65 (40,13%) anni (vedi Tab. 1).

L'indagine ha inoltre confermato la presenza di stretti rapporti fra alcolismo e manifestazioni patologiche (più dell'80% degli etilisti, sia uomini che donne, presentano componenti di rilevanza clinica rilevati dagli strumenti SAD ed SCL-90) (vedi Tab. 2). La SAD ha evidenziato negli alcolisti un significativo disagio emotivo a colorito nevrotico e depressivo. Queste componenti psicopatologiche non sono correlate né all'età né al sesso.

TAB. 1 - *Frequenze alcolisti per sesso e classi di età.*

Età *	Maschi			Femmine		
	N. sogg. esam.	Alcolisti	%	N. sogg. esam.	Alcolisti	%
18-34	150	9	6	169	—	—
35-49	140	36	25,71	117	4	3
50-65	144	59	40,13	106	9	8
66-70	3	—	—	1	—	—

* Limitatamente a 3 rilevazioni è mancante il dato relativo all'età.

TAB. 2 - *Alcolisti: componenti psicopatologiche evidenziate dai questionari.*

	Maschi		Femmine	
	n.	%	n.	%
Depressi (SAD)	46	44,23	10	76,92
Altre componenti (SCL-90)	40	38,46	1	7,69
Assenza	18	17,30	2	15,30

Negli alcolisti il punteggio totale medio del VARA Test si correla con la sintomatologia psicopatologica esplorata dalla SAD nei suoi vari fattori (vedi fig. 10). L'entità della compromissione affettiva espressa dal punteggio totale medio SAD è risultata correlare positivamente con le motivazioni e giustificazioni al bere, con le conoscenze di persone con problemi legati all'alcool e con le conseguenze sociali e familiari, fattori questi esplorati dal VARA Test. L'indagine sembra quindi confermare la presenza di un'alta prevalenza di ma-

CORRELAZIONI TOTVARA VS FATTORI SAD (BY GR)

GR=ALCOOLISTI

VARIABLE	N	MEAN	STD DEV	SUM	MINIMUM	MAXIMUM
TV	117	2.002139	0.4047135	234.2502	1.075000	2.950000
FATTSAD1	117	1.650175	0.4229206	193.0705	1.000000	2.846154
FATTSAD2	117	1.754767	0.4746084	205.3077	1.000000	3.153846
FATTSAD3	117	1.447293	0.4319882	169.3333	1.000000	2.666667
FATTSAD4	117	1.441595	0.4394987	168.6667	1.000000	2.666667
FATTSAD5	117	1.547009	0.7070547	181.0000	1.000000	3.500000
FATTSAD6	117	1.311111	0.3754946	153.4000	1.000000	2.400000

PEARSON CORRELATION COEFFICIENTS / PROB > |R| UNDER H0:RHO=0 / N = 117

TV

FATTSAD1	0.41928 0.0001
FATTSAD2	0.35794 0.0001
FATTSAD3	0.30229 0.0009
FATTSAD4	0.34591 0.0001
FATTSAD5	0.30116 0.0010
FATTSAD6	0.19299 0.0371

Fig. 10

nifestazioni psicopatologiche, tra cui spiccano i disagi emotivi a sfondo depressivo negli alcolisti non ospedalizzati e studiati nel proprio ambiente di vita, indica inoltre nel sesso maschile una condizione a rischio per l'etilismo. Da sottolineare che la presenza di sintomatologia depressiva incide specialmente nel fornire all'alcolista le motivazioni e le giustificazioni al bere più che correlarsi alla quantità (CASTROGIOVANNI *et al.*, 1980).

Senza addentrarci oltre nella descrizione della metodologia statistica utilizzata e del commento dei risultati (peraltro ancora parziali) ottenuti, cercheremo ora di illustrare alcune parti dei programmi SAS a nostro avviso più adatte a dare al lettore una prima sommaria idea delle problematiche incontrate e delle principali caratteristiche del sistema.

ALCUNI ESEMPI TRATTI DALL'APPLICAZIONE

Il problema si è presentato particolarmente complesso per quel che riguarda la definizione dei dati. Gli strumenti utilizzati sono infatti costituiti da numerose variabili da raggruppare in fattori sulla base di analisi fattoriali esploranti diversi settori. Così, ad esempio, il VARA-TEST, composto da 40 item, esplora 6 settori (fattori): a) modalità e quantità nel bere, familiari e personali, b) conseguenze sociali, c) coscienza dell'esistenza nel proprio ambiente di individui con problemi causati dall'alcool, d) motivazioni e giustificazioni che accompagnano l'uso abituale di alcoolici, e) sintomi somatici, f) modificazioni dell'attività e degli interessi sessuali. Così, ad esempio, la SAD, composta da 31 item, esplora 6 settori (fattori): a) depressione, b) ansietà (psichica e somatica), c) insicurezza, d) sintomi somatici, e) variazioni diurne, f) distacco dalla vita (CASTROGIOVANNI *et al.*, 1980).

Si è sfruttata la possibilità di definire dei vettori (ARRAY) e di generare nuove variabili (i fattori)⁽¹⁾ con l'istruzione DO/END.

La fig. 9 riporta le istruzioni più significative del programma SAS utilizzato per verificare l'esistenza di eventuali correlazioni esistenti fra il totale VARA ed i fattori SAD e fra il totale SAD ed i fattori VARA sia sull'intero campione che sui singoli gruppi diagnostici.

Come si può vedere nel Data Step sono stati definiti e dimensionati numerosi vettori: il vettore SAD per contenere i valori delle variabili della SAD, i 6 vettori SFT che definiscono gli item della SAD da utilizzare per ricavare i 6 punteggi fattoriali SAD, il vettore NELSAD che, inizializzato tramite la successiva istruzione RETAIN, serve ad indicare da quanti elementi è costituito ogni singolo fatto-

(1) I punteggi fattoriali sono stati ottenuti sommando i valori degli item costituenti i fattori e dividendo per il numero degli item.

re, il vettore ALLSFT che ha come suoi elementi altri vettori ⁽²⁾ (i vettori SFT) ed il vettore FATTSD che alla fine del Data Step conterrà i punteggi fattoriali SAD. Interessanti i cicli di DO che calcolano i fattori SAD ed in particolare l'utilizzazione del vettore ALLSFT che permette, per ogni subscripto L, di cumulare sui 6 elementi del vettore FATTSD le appropriate variabili definite negli ARRAY con subscripto I; così ad es. per L=2 saranno cumulati (vedi istruzione FATTSD=FATTSD+ALLSFT) su FATTSD2 i valori delle variabili indicati nell'ARRAY SFT2 col subscripto I che assumerà tutti i valori compresi fra 1 e 13 (per L=2 NELSAD=13).

Nella fig. 9 sono inoltre mostrati due Proc Step relativi al calcolo delle correlazioni fra il totale VARA ed i fattori SAD sia su tutto il campione che sui singoli gruppi diagnostici (l'istruzione BY GR permette di ripetere l'analisi per ognuno dei gruppi diagnostici).

La fig. 10 mostra la parte dell'output relativa al calcolo delle correlazioni fra il totale VARA (TV) ed i 6 fattori della SAD per il gruppo diagnostico degli alcolisti.

Negli esempi illustrati finora abbiamo sempre avuto un Data Step iniziale seguito da un certo numero di Proc Step. In un programma SAS i Data Step ed Proc Step possono però apparire in qualsiasi numero ed in qualsiasi ordine ed un Data Step può addirittura utilizzare al suo interno dei valori prodotti da un Proc Step precedente; è il caso del programma in parte mostrato in fig. 11 dove la media e lo standard error del totale VARA e del totale SAD vengono utilizzati per suddividere il campione in quattro gruppi che esprimono un diverso livello di gravità per quanto riguarda la compromissione affettiva (totale SAD) ed il grado di etilismo (totale VARA). I quattro gruppi sono rappresentati da:

a) soggetti con $TS > TSMEAN + TSSTDE$ e $TV > TVMEAN + TVSTDE$

b) soggetti con $TS > TSMEAN + TSSTDE$ e $TV < TVMEAN + TVSTDE$

c) soggetti con $TS < TSMEAN + TSSTDE$ e $TV > TVMEAN + TVSTDE$

d) soggetti con $TS < TSMEAN + TSSTDE$ e $TV < TVMEAN + TVSTDE$

⁽²⁾ Il SAS prevede per gli insiemi dimensionati non più di una dimensione (un solo subscripto), ma è possibile aggirare questa apparente limitazione definendo un vettore come elemento di un altro vettore che a sua volta può essere definito elemento di un altro vettore ancora e così via fino a raggiungere il livello di subscripti desiderato.

```

DATA A;
ARRAY .....
.
.
INPUT .....
.
.
PROC MEANS;
VAR TV TS;
OUTPUT OUT = B MEAN=TVMEAN TSMEAN STDERR=TVSTDE TSSTDE;
DATA C;
RETAIN TVMEAN TVSTDE TSMEAN TSSTDE;
IF _N_=1 THEN SET B;
SET A;
IF TS GT (TSMEAN+TSSTDE) AND TV GT (TVMEAN+TVSTDE) THEN GRSV=1;
IF TS GT (TSMEAN+TSSTDE) AND TV LT (TVMEAN-TVSTDE) THEN GRSV=2;
IF TS LT (TSMEAN-TSSTDE) AND TV GT (TVMEAN+TVSTDE) THEN GRSV=3;
IF TS LT (TSMEAN-TSSTDE) AND TV LT (TVMEAN-TVSTDE) THEN GRSV=4;
RUN;
.
.
.

```

Fig. 11 - ...in un programma SAS i Proc Steps ed i Data Steps possono apparire in qualsiasi numero ed ordine.

dove: TSMEAN è il punteggio totale medio SAD, TVMEAN è il punteggio totale medio VARA, TSSTDE è l'errore standard del totale SAD, TVSTDE è l'errore standard del totale VARA. La PROC MEANS calcola le medie e gli errori standard del totale VARA (TV) e del totale SAD (TS) ed archivia i relativi valori nel dataset B (vedi l'istruzione OUTPUT OUT=B...). Le medie e gli errori standard vengono poi utilizzati nel Data Step C per definire, tramite le istruzioni di controllo IF la nuova variabile GRSV che suddivide il campione nei quattro gruppi specificati sopra. Da notare che il dataset C viene costruito tramite la lettura delle osservazioni dal dataset A e B (istruzioni SET A e SET B). Poiché il dataset B presenta un'unica osservazione (costituita come abbiamo visto dalle variabili TVMEAN, TSMEAN, TVSTDE, TSSTDE), essa viene letta una sola volta all'inizio dell'esecuzione del Data Step C; ciò è stato possibile tramite l'utilizzazione della variabile `__N__` il cui valore, generato

automaticamente dal SAS, indica quante volte è stato eseguito il Data Step fino a quel momento ⁽³⁾.

Una procedura particolarmente utile e potente è la TABULATE che produce delle statistiche descrittive in tabelle gerarchiche.

```

PROC FORMAT;
VALUE $D 1 = 'MENO DI 6M'
      2 = '6M-1A'
      3 = 'PIU di 1A';
.
.
.
DATA A;
INPUT (. . . . DURMP GRMP PRMP CGEN SDIS.....) (formato);
.
.
.
PROC TABULATE;
TITLE RILEVAZIONE STATO MALATTIA;
TITLE2 GRUPPO ALCOOLISTI;
CLASS GRMP DURMP PRMP CGEN SDIS;
FORMAT DURMP $D. ;
FORMAT GRMP $G. ;
FORMAT PRMP $P. ;
FORMAT CGEN $C. ;
FORMAT SDIS $S. ;
LABEL DURMP = 'DUR.MAL.'
      GRMP = 'GRA.MAL.'
      PRMP = 'PRO.MAL.'
      CGEN = 'CON.GEN.'
      SDIS = 'STA.DIS.';
TABLE (GRMP*DURMP*PRMP*CGEN*SDIS ALL),
(N PCTN);
RUN;

```

Fig. 12 - La procedura TABULATE.

Nel programma di fig. 12 sono richieste, per gli alcolisti, le frequenze del gruppo suddiviso in base alle variabili di classificazione gravità, durata e prognosi della malattia, condizioni generali e stato di disabilità (vedi variabili specificate nell'istruzione CLASS). L'istruzione TABLE specifica la gerarchia fra le variabili di classificazione e le statistiche desiderate. Il programma mostrato fa anche largo

⁽³⁾ È importante ricordare che il SAS esegue, per ogni osservazione letta, tutte le istruzioni del Data Step, il controllo esce dal Data Step solo quando in uno dei dataset da cui si legge terminano le osservazioni. Nel caso specifico nel dataset B è presente una sola osservazione, ma l'esecuzione del Data Step C prosegue fino all'esaurimento delle osservazioni del dataset A per la presenza dell'istruzione RETAIN che rende disponibili i valori dell'osservazione in B per tutta la durata del Data Step.

uso delle istruzioni FORMAT e LABEL per etichettare nomi e valori di variabili. L'output del programma è mostrato in fig. 13.

MODALITÀ DI LAVORO

Per concludere un accenno alle modalità di lavoro della versione VM del SAS; in tale ambiente si può lavorare in tre modi diversi:

RILEVAZIONE STATO MALATTIA
GRUPPO ALCOOLISTI

GRUP=ALCOLISTI

					N	PCTN
DUR.- MAL.	GRA.- MAL.	PRO.- MAL.	CON.- GEN.	STA.DI- S.		
MENO DI 6M	LIEVE	FAUS- TA	BUONE	NON INV.	4.00	3.42
6M-1A	MEDIA	INFA- USTA	MEDI- OCRI	NON INV.	1.00	0.85
			GRAVE	INFA- USTA	1.00	0.85
PIU DI 1A	LIEVE	FAUS- TA	BUONE	NON INV.	8.00	6.84
				INVALI- DO	2.00	1.71
				MEDI- OCRI	4.00	3.42
	MEDIA	FAUS- TA	BUONE	NON INV.	35.00	29.91
				INVALI- DO	5.00	4.27
				INABILE	1.00	0.85
				MEDI- OCRI	7.00	5.98
				INVALI- DO	12.00	10.26
				SCAD- UTE	1.00	0.85
		INFA- USTA	BUONE	NON INV.	3.00	2.56
				INVALI- DO	2.00	1.71
				MEDI- OCRI	4.00	3.42

(CONTINUED)

RILEVAZIONE STATO MALATTIA
GRUPPO ALCOOLISTI

GRUP=ALCOLISTI

					N	PCTN
DUR.- MAL.	GRA.- MAL.	PRO.- MAL.	CON.- GEN.	STA.DI- S.		
PIU DI 1A	MEDIA	INFA- USTA	MEDI- OCRI	INVALI- DO	6.00	5.13
	GRAVE	FAUS- TA	BUONE	NON INV.	1.00	0.85
				INVALI- DO	3.00	2.56
			MEDI- OCRI	NON INV.	3.00	2.56
				INVALI- DO	1.00	0.85
		INFA- USTA	MEDI- OCRI	NON INV.	2.00	1.71
				INVALI- DO	5.00	4.27
				INABILE	2.00	1.71
			SCAD- UTE	INVALI- DO	3.00	2.56
				INABILE	1.00	0.85
ALL					117.00	100.00

Fig. 13

- 1) interattivo
- 2) batch
3. ibrido (interattivo+batch)

Nel modo interattivo il SAS permette di digitare i comandi a terminale; dopo che si è digitato un Data o Proc Step il SAS esegue le istruzioni in quello step; in funzione dei risultati ottenuti l'utente può decidere immediatamente che cosa fare successivamente. Questo metodo è molto comodo se si devono elaborare dei dataset SAS precedentemente costruiti, ma ha lo svantaggio che tutti i comandi SAS digitati vengono perduti quando termina la sessione SAS.

Nel modo batch i comandi SAS vengono memorizzati in files CMS e vengono eseguiti specificando al SAS i nomi dei file. Questo modo di lavorare ha il vantaggio che i programmi SAS creati vengono mantenuti, ma è scomodo se si vogliono fare elaborazioni veloci su dataset SAS già esistenti.

Il terzo modo, cosiddetto ibrido, unisce i vantaggi dei due precedenti metodi; è difatti possibile operare in modalità interattiva ed eseguire contemporaneamente, tramite il comando GETSAS, programmi SAS memorizzati in files CMS; quando tutti i comandi del programma SAS sono stati eseguiti, si resta in ambiente SAS e si possono eseguire interattivamente ulteriori Proc Step e Data Step.

CONCLUSIONI

Nella presente esposizione abbiamo cercato di illustrare le caratteristiche principali dello schema funzionale del SAS cercando al contempo di mettere in risalto quelle che a nostro avviso sono le migliori doti del sistema, ovvero: semplicità di utilizzazione che rende lo strumento facilmente agibile anche in ambienti non prettamente informatici e grande potenzialità di elaborazione dovuta alla ricchezza delle procedure esistenti ed al continuo, periodico aggiornamento delle librerie. Questi motivi hanno spinto un numero sempre maggiore di Utenti a credere in questo prodotto; recentemente, come già in altri paesi, si è costituito anche in Italia il «SAS User's Group Italiano» (SUGI) al fine di permettere il più ampio ed agevole scambio di esperienze fra gli utenti SAS italiani (LAFORENZA, 1985) e favorire un'opera di divulgazione ed aggiornamento del prodotto sottoponendo al SAS Institute, attraverso la casa distributrice per l'Italia, precise richieste di miglioramento.

BIBLIOGRAFIA

- BRACCI E., LAFORENZA D. (1985) - Il SAS al CNUCE. *Copyright-Primo trimestre 1985 by-CNUCE-Pisa, Istituto del Consiglio Nazionale delle Ricerche*, 58-64.
- CASTROGIOVANNI P., PASCULLI E., MANFREDI C., CONTI L., BIANCHI B., CASSANO G.B. (1980) - Prospettive di studio su alcoolismo e depressione. *Pubblicazione a cura del Dipartimento Sicurezza Sociale e del Servizio Editoriale della Giunta Regionale*, 253-264.
- LAFORENZA D. (1985) - Costituzione del «SAS User's Group Italiano (SUGITalia '85)». *Copyright-Secondo trimestre 1985 by-CNUCE-Pisa. Istituto del Consiglio Nazionale delle Ricerche*, 34-40.
- PISANTI P. (1985) - SAS. Un software statistico conversazionale. *Informatica Oggi*, n. 9-'85, Gruppo Editoriale Jackson, 60-64.

(ms. pres. il 14 settembre 1987; ult. bozze il 21 marzo 1988)