



Munich Personal RePEc Archive

IMTS: a programming language to manage the time series data base

Giorgio Calzolari

IBM Scientific Center, Pisa, Italy

1973

Online at <https://mpra.ub.uni-muenchen.de/24439/>

MPRA Paper No. 24439, posted 2. September 2010 16:12 UTC

Centro Scientifico IBM - Pisa

I M T S

U N L I N G U A G G I O P E R L A
G E S T I O N E D E L L ' A R C H I V I O
D E L L E S E R I E S T O R I C H E

NOTA TECNICA

G. Calzolari

INDICE

1. Note introduttive.....	pag. 1
2. Gestione della memoria centrale.....	5
3. Gestione dei files.....	8
4. Accesso all' archivio e elaborazione dei dati.....	17
5. Programmi di utilita'.....	33
6. Organizzazione dei dischi.....	35
7. Esempio di seduta di lavoro al terminale.....	36

1. NOTE INTRODUTTIVE

Per lo studio della struttura dell' economia Italiana e' di importanza vitale la costituzione di un archivio dei dati, con il principale obiettivo di mettere a disposizione dei ricercatori un bagaglio di informazioni, uguali per tutti, che possa arricchirsi col procedere degli studi.

Queste due esigenze fondamentali si sono concretizzate nella filosofia organizzativa del linguaggio interattivo per la gestione dell' archivio dei dati. Infatti un numero elevato di utenti possono attingere interattivamente alle stesse informazioni, senza intralciarsi l' un l' altro, mentre e' demandata ad un solo responsabile la decisione dell' aggiornamento, anche se chiunque puo' preparare i dati da memorizzare.

L' archivio centrale e' unico, dato che le prove effettuate durante le ricerche debbono avere una base comune di riferimento e di confronto.

Tale impostazione obbedisce alla necessita' di avere sotto controllo la situazione dell' archivio in ogni istante della ricerca, eliminando nello stesso tempo possibili conflitti derivanti da aggiornamenti simultanei.

LINGUAGGIO DI GESTIONE DELL' ARCHIVIO

L' archivio centrale e' costituito da un insieme di serie di dati, globalmente gestite da un solo programma denominato IMTS (Interactive Management for Time Series).

Col richiamo di tale programma e' possibile estrarre le informazioni dall' archivio centrale e da eventuali archivi privati dei singoli utenti, utilizzando come riferimento il solo nome della serie dei dati richiesta.

Sulle informazioni contenute nelle serie storiche e' possibile operare le trasformazioni che spesso vengono richieste durante l' analisi economica. Per esempio oltre alle operazioni elementari e alla possibilita' di applicare operatori matematici, e' possibile ottenere il valor medio, la varianza, lo slittamento dei dati e l' estrazione di un dato elementare della serie.

Non e' richiesta all' utente alcuna conoscenza a priori circa il contenuto delle serie. E' richiesto invece che egli conosca i nomi delle serie che intende adoperare, la cui lista e' ottenibile da terminale.

INTERAZIONI CON LA RICERCA

Data la natura sperimentale della ricerca, e' del tutto comprensibile come alcune funzioni possano non essere complete, o manchino. Si pensa che un contributo della ricerca sia quello di migliorare e completare il linguaggio, inteso come strumento flessibile di analisi.

Tenendo presente la finalita' dell' archivio dei dati, due sono gli utilizzi principali:

- 1) Studio di funzioni matematiche, ai fini di individuare le leggi che governano la struttura economica.
- 2) Strutturazione delle serie temporali, ai fini della simulazione di modelli economici.

La prima fase e' di carattere sperimentale, e richiede quindi la presenza di un mezzo flessibile di trattamento delle informazioni, fornendo tutti i possibili strumenti di indagine. Questo e' l' obiettivo principale del linguaggio descritto, pur permettendo collegamenti con programmi gia' esistenti, per perfezionare il dialogo tra ricercatore e archivio dei dati.

La seconda fase, essenzialmente di verifica economica, e' la meno impegnativa come struttura dei dati, in quanto gia' il DMS/2 ha in se' la organizzazione portante per affrontare la simulazione di un modello econometrico. Cio' nondimeno e' necessario coordinare in un unico disegno sia

la fase sperimentale che quella di verifica per rendere il ricercatore il più possibile indipendente da problemi che coinvolgano l'organizzazione dei dati.

2. GESTIONE DELLA MEMORIA CENTRALE

Si possono distinguere quattro zone di memoria.

- 1) Nucleo del sistema operativo CMS, area libera per le funzioni del sistema, e Transient Area (0-12000 hex.).
- 2) Area libera con basso indirizzo, destinata a contenere la matrice di lavoro, o a caricare programmi formato module del sistema o dell'utente
(EDIT, COMBINE, SORT, NEW, CREARCH, APLIUP, ecc.)
(12000-36000 hex.).
- 3) Area protetta, contenente le routines permanentemente residenti in memoria (statiche) del programma, ad esempio le routines di codifica e decodifica, le routines di I/O, ecc.
(36000-4f000 hex. circa).
- 4) Area libera con indirizzo alto, destinata al caricamento dinamico delle routines di plotter o regressione (4f000-fine della memoria virtuale).

Prima di interpretare un comando, il programma agisce sulla tabella NUCON del CMS, modificandone cinque indirizzi nel modo seguente.

LOWEXT (estremo inferiore degli indirizzi) di memoria alta occupata dal sistema operativo tramite le routines FREE e EXTEND): viene salvato il valore, e rimpiazzato da 34000

(hex.); qualunque modulo eseguibile (del sistema o dell'utente), o qualunque modulo TEXT venga caricato, non può essere allocata (e quindi sporcata) memoria al di sopra di quell'indirizzo.

LOCNT (location counter) e HMAIN (estremo superiore degli indirizzi di memoria bassa allocati dalla routine GETMAIN del sistema operativo): viene salvato il contenuto, e rimpiazzato da 12000 (hex.) in entrambe.

LSTADR (indirizzo dell'ultima pagina di memoria della macchina virtuale) e LDRTBL (ultimo indirizzo delle Loader Tables, praticamente uguale all'ultimo indirizzo della memoria virtuale): viene salvato il contenuto, e sostituito da 35000 e 36000 (hex.) rispettivamente.

Subito dopo il programma controlla se il comando da eseguire è un comando di CMS. In caso affermativo il CMS prende il controllo, esegue il comando, e ritorna al programma chiamante, che provvede a ripristinare gli indirizzi modificati in NUCON. In caso negativo, invece, vengono subito ripristinati gli indirizzi modificati in NUCON, e viene passato il controllo alle routines di codifica e decodifica del comando (permanentemente residenti in memoria). Tali routines utilizzano l'area libera (12000-36000) organizzata in forma di matrice di lavoro.

Qualora il comando dato richieda l'utilizzo di una routine di regressione o di plotter, non residente in memoria, la routine richiesta viene caricata dinamicamente

dal LOADER del CMS nell' area libera con indirizzo alto, e, tra i vari parametri, le viene passato l' indirizzo della matrice di lavoro.

Qualunque comando di CMS puo' essere eseguito. Non e' possibile, pero', far eseguire comandi concatenati che facciano riferimento ad una comune area di memoria (ad esempio LOAD e START di un programma, pero' LOAD (XEQ) e' possibile, e cosi' pure \$).

In particolare i comandi di CMS di cui e' spesso opportuno servirsi sono i seguenti:

EDIT
ALTER
COMBINE
LISTF
STAT
PRINT e OFFLINE PRINT
SORT
APLIUP

al quali si aggiungono i seguenti

NEW
CREARCH

che sono nomi di programmi di utilita' (contenuti in formato MODULE nella macchina virtuale che contiene l' archivio delle serie storiche) caricati ed eseguiti dal CMS allo stesso modo dei moduli eseguibili del sistema.

3. GESTIONE DEI FILES

Le routines di input/output del programma permettono di gestire quattro classi di files.

La prima e' costituita da un unico file, l'archivio delle serie storiche permanenti e utilizzabili da tutti gli utenti.

La seconda e' costituita anch'essa da un unico file, contenente i nomi abbreviati di tutte le serie storiche permanenti. E' l'indice dell'archivio.

La terza comprende tutti i files che contengono serie storiche private dei singoli utenti, o serie storiche ottenute come risultati intermedi di una elaborazione, da utilizzare in elaborazioni successive.

La quarta classe, infine, e' costituita da un unico file, che memorizza di volta in volta i comandi da eseguire. Viene utilizzato sia per correggere errori formali dei comandi, sia per eseguire una successione laboriosa di comandi, comandi che possono essere memorizzati una volta per tutte e richiamati di volta in volta.

Le routines di I/O richiedono che i files abbiano tutti records di lunghezza fissa.

3-1 L' archivio delle serie storiche

Identificatore CMS: ARCHIVIO SERIESTO P5.

E' un file ad accesso diretto, composto da 2100 records di lunghezza fissa 800.

Ogni record, destinato a contenere una sola serie storica, contiene le seguenti informazioni:

- 1) Il nome completo della serie storica.
- 2) Il nome abbreviato della serie.
- 3) La fonte dei dati.
- 4) L' unita' di misura.
- 5) Un codice per indicare se si tratta di dati annuali, semestrali, trimestrali, o mensili.
- 6) L' anno iniziale.
- 7) L' anno finale.
- 8) Il trimestre (o semestre, o mese) iniziale.
- 9) Il trimestre (o semestre, o mese) finale.
- 10) Il numero di dati della serie.
- 11) I dati veri e propri.
- 12) Un codice per indicare se i dati continuano in un altro record (questa eventualita' si verifica solo nel caso di dati mensili per almeno 14 anni).

Tra l' inizio e la fine i dati devono essere completi (non

sono ammessi buchi).

Il numero massimo di valori di una serie e' 360.

Alla routine che gestisce l' Input/output dall' archivio viene passato, come parametro, il numero d' ordine del record da leggere (o da scrivere). Il record, senza formato, viene letto e allocato in un opportuno buffer (o scritto dal medesimo) di 800 bytes.

3-2 Il directory

Identificatore CMS: TABELLA ABBREV P5.

E' un file di 210 records di lunghezza fissa 80 (formato scheda).

Ogni serie storica e' contraddistinta dal suo nome abbreviato.

Ogni record del directory contiene 10 nomi abbreviati di serie storiche (8 bytes ciascuno). Al nome contenuto nel record x , posizione y , corrisponde, biunivocamente, il record $10(x-1) + y$ dell' archivio.

Gli ultimi 100 records dell' archivio e gli ultimi 10 del directory sono riservati alle serie storiche con continuazione.

Non sono ammesse plu' serie con lo stesso nome abbreviato. Tutti gli altri dati, invece, possono essere comuni a piu' serie.

Il file viene letto all' inizio della seduta di lavoro, e allocato in una opportuna tabella residente in memoria (statica). La ricerca di una serie viene svolta nel modo seguente: il programma fa uno scan della tabella (in memoria) confrontando i nomi abbreviati di serie storiche in essa contenuti con il nome richiesto. Trovato il nome, il programma ne calcola il numero d' ordine, passa questo numero alla routine di I/O dell' archivio, che finalmente legge dall' archivio il record contenente la serie richiesta. Nel record e' contenuto un flag che indica se la serie continua in un altro record. In caso affermativo il programma esegue lo scan delle ultime posizioni della tabella, alla ricerca dell' identificatore 'nnnn****', dove nnnn e' il numero d' ordine della serie, e procede come sopra.

I valori numerici della serie storica vengono allocati in una colonna della matrice di lavoro.

3-3 I files di lavoro e le serie storiche private

Identificatore CMS: FILE nome_abbreviato P1.

Ogni utente puo' utilizzare, oltre all' archivio centrale, anche serie storiche private, contenute nel proprio disco di lavoro. In questo caso ogni serie storica e' un file, (il cui FILETYPE coincide col nome abbreviato della serie stessa.

Ogni file di questo genere deve avere records fissi, lunghi 80 (formato scheda), con il seguente tracciato:

Prima scheda.

col 1- 4	Numero d'ordine della serie.
5-12	Nome abbreviato.
13	Suddivisione in periodi (Y=dati annuali) (S=dati semestrali) (Q=dati trimestrali) (M=dati mensili)
14-15	Anno iniziale (ultime due cifre).
16-17	Periodo iniziale (ad es 01 per dati mensili significa gennaio, per dati semestrali significa primo semestre; per dati annuali si deve indicare sempre 01).
18-19	Anno finale.
20-21	Periodo finale.
22-25	Numero di dati.
26-75	Nome completo della serie.
76-77	Sigla per indicare se si tratta di numeri indici, o livelli, o differenze prime, o tassi di variazione (rispettivamente IN,LE,FD,RV).

Seconda scheda.

col 1-32	Fonte dei dati.
----------	-----------------

33-52 Unità' di misura.

Dalla terza in poi.

col 8-22 Valori della serie in sequenza
(formato F15.6 o formato E15.8).

Ultima scheda (tappo).

col 1- 7 ENDFILE

Se i dati vengono preparati su schede, e' necessario rispettare i formati sopra descritti. Per preparare, invece, nuove serie da terminale, o aggiornare serie preesistenti, sono utilizzabili alcuni programmi, descritti nel seguito, che svincolano l'utente da questa necessita'.

I files di lavoro, ottenuti da serie storiche contenute in files o nell'archivio centrale dopo elaborazioni di vario genere (ad esempio la somma di tre serie), hanno lo stesso formato, salvo che le posizioni relative al nome completo, alla fonte dei dati e all'unità' di misura, diventando prive di significato, sono lasciate in bianco. Qualora l'utente decidesse poi di conservare quel file, attribuendo loro particolari significati, potrà provvedere all'aggiornamento ricorrendo al programma EDIT del CMS.

Per inserire una nuova serie nell'archivio centrale, occorre preparare un file di questo genere, il cui identificatore CMS deve essere FILE FT01F001 P1. Più serie

possono essere accodate nello stesso file. Dalla macchina virtuale, cui e' affidata la gestione dell' archivio, viene fatto partire il programma di agglornamento (CREARCH).

Il programma di agglornamento legge il file ed esegue i seguenti controlli:

1) Il numero vero di dati deve coincidere con quello indicato.

2) Il numero di dati deve coincidere con quello ricavabile dall' anno iniziale e finale e dal numero di dati per ogni anno.

Se uno di questi due controlli da' esito negativo, l' agglornamento non viene eseguito.

3) Controlla se esiste gia' una serie con lo stesso nome abbreviato. In caso affermativo chiede l' autorizzazione a sostituirla con la nuova serie.

4) Controlla se il record corrispondente al numero d' ordine indicato e' libero o occupato. Se e' occupato, esegue ugualmente l' agglornamento, inserendo la serie nel primo record libero e segnalandone il numero a console.

5) Se la serie richiede un record continuazione, cerca la prima posizione libera nelle 100 riservate. La corrispondente casella del directory viene contrassegnata da 'nnnn****' dove nnnn e' il numero del record che contiene la parte iniziale della serie.

3-4 Il file per la correzione
degli errori formali

Identificatore CMS: GO IMTS P1.

Nel corso di una seduta di lavoro al terminale puo' capitare di dare comandi contenenti errori formali (parentesi o apici non chiusi, ad esempio). Soprattutto nel caso di comandi complessi e' comodo correggere il comando dato, anziche' riscriverlo per intero. Ogni comando che passa per le routines di codifica e decodifica (esclusi quindi i comandi di CMS) viene memorizzato nel file GO IMTS, a partire dal primo record.

Il file ha records fissi di lunghezza 130, lunghi cioe' come la riga di input del terminale. In caso di errore formale, il file viene automaticamente editato e, con i consueti comandi dell' EDIT del CMS, puo' essere corretto dall' utente. Subito dopo, anziche' dare altri comandi, l' utente puo' utilizzare il comando GO, che fa si' che il programma esegua i comandi contenuti nel file. Ogni comando nuovo scritto al terminale si sovrappone ai comandi precedentemente scritti nel file; piu' esattamente ogni nuovo comando viene memorizzato nel primo record del file, e continua nei records successivi solo se il comando dato, non essendo sufficiente una riga, contiene il segno di continuazione (=).

Questo file viene allocato all' inizio di ogni seduta di lavoro, e automaticamente cancellato alla fine. Se usato per memorizzare ed eseguire una lunga serie di comandi, come ad esempio l' intera serie di regressioni per la stima dei coefficienti delle equazioni di un modello, e' necessario conservare una copia del file con un altro nome, per evitarne la cancellazione alla fine della seduta di lavoro.

4. ACCESSO ALL' ARCHIVIO
E ELABORAZIONE DEI DATI

Per l'utente tutte le funzioni di input dei comandi e di stampa dei risultati sono svolte attraverso il terminale. L'utente deve scrivere al terminale il comando che intende far eseguire al programma, composto da nomi di serie storiche (dell'archivio centrale o private; in caso di omonimia prevalgono le ultime), costanti numeriche, segni di operatori e nomi di funzioni. Un comando deve, normalmente, essere contenuto in una riga, ma e' prevista la possibilita' di continuare in una o piu' righe successive, utilizzando il segno speciale di continuazione (=) al termine di ogni riga, tranne l'ultima. Il programma decodifica ogni riga subito dopo che e' stata battuta, ma aspetta a iniziare l'esecuzione del comando fino a quando non incontra una riga che non termina col carattere di continuazione.

Le costanti numeriche devono essere scritte con il punto decimale, e possono avere fino a 8 cifre complessivamente; non sono ammessi spazi bianchi all'interno della costante.

I nomi delle serie storiche, lunghi fino a 8 caratteri, devono essere scritti tra apici, e cio' per evitare possibili confusioni tra nomi di variabili e nomi di operatori semplici o composti. Non devono essere aggiunti

spazi bianchi entro gli apici in testa al nome o in mezzo, in coda, invece, sono ammessi.

I nomi delle funzioni devono essere scritti senza spazi bianchi intermedi. In tutti gli altri casi gli spazi bianchi possono essere aggiunti senza problemi.

Il segno di continuazione della riga non deve spezzare il nome di una variabile, o di un operatore, o una costante numerica.

Finlto di eseguire un comando, il programma ne stampa i risultati al terminale, e questi risultati non vengono memorizzati. Qualora si voglia conservare i risultati (e questo e' possibile solo quando i risultati siano una serie storica o un singolo risultato numerico, e' escluso ad esempio l' output del plotter) occorre scrivere in fondo alla stringa comando la parola chiave FILE seguita, senza spazi bianchi intermedi, dal nome del file che si vuole creare: ad esempioFILEPRODNAZ I risultati del comando vanno nel FILE PRODNAZ P1, e possono essere utilizzati in tutte le operazioni seguenti semplicemente facendo riferimento alla serie storica 'PRODNAZ'.

Sulle serie storiche, e su eventuali costanti, e' possibile eseguire le seguenti operazioni:

1) Operazioni e funzioni eseguite dalle routines residenti in memoria (statiche).

+	Somma
-	Sottrazione
*	Moltiplicazione
/	Divisione
**	Elevazione a potenza
LOG	Logaritmo naturale
LOG10	Logaritmo decimale
EXP	Esponenziale (base e)
SIN	Seno (radianti)
COS	Coseno (radianti)
ATAN	Arcotangente (radianti)
SINH	Seno iperbolico
COSH	Coseno iperbolico
ABS	Valore assoluto
MAX	Valor massimo
MIN	Valor minimo
MEAN	Valor medio
VAR	Varianza
INT	Parte Intera
RV	Tasso di variazione
LAG(nn)	Slittamento dei dati
ONE(nn/mm)	Estrazione di un solo dato da una serie
SEL(nn/mm, ll/kk)	Estrazione di una parte di una serie
CPR(nn)	Compressione di una serie

2) Operazioni e funzioni eseguite da routines a caricamento dinamico.

PLOT	Plotter in linea
OLS	Regressione minimi quadrati ordinari
TSLs	Regressione minimi quadrati a due stadi
LISE	Limited information single equation
DMS	Collegamento col DMS/2

4-1 Priorita' degli operatori

Le funzioni (LOG, LOG10, EXP, SIN, COS, SINH, COSH, ATAN, ABS, MAX, MIN, MEAN, VAR, INT, RV, LAG, ONE, SEL, CPR) hanno priorita' massima. Se piu' funzioni vengono concatenate, la priorita' e' crescente da sinistra a destra (viene eseguita prima la funzione interna).

Segue, nell' ordine delle priorita', l' elevazione a potenza.

Seguono moltiplicazione e divisione, in ordine da sinistra a destra.

Seguono somma e sottrazione in ordine da sinistra a destra.

Per cambiare le priorita' possono essere inserite parentesi fino a tre livelli.

4-2 Operatori diadici

+,-,*,/,**

L'operatore deve essere scritto in mezzo ai due operandi cui si riferisce. Non sono necessari particolari segni separatori.

L'esecuzione di una di queste operazioni è soggetta alle seguenti regole. Tra due costanti, il risultato è una nuova costante. Tra una costante e una serie, il risultato è una nuova serie con le stesse caratteristiche (data d'inizio e di fine, suddivisione in periodi) della serie di partenza. Tra due serie, l'operazione è ammessa solo se le serie hanno uguale suddivisione in periodi (ad esempio entrambe dati mensili, o entrambe dati annuali); in tal caso il risultato è una serie limitata al periodo comune alle due serie di partenza (ad esempio la somma di due serie di dati annuali, rispettivamente dal 1953 al 1972 e dal 1951 al 1970, è una serie di dati annuali dal 1953 al 1970).

4-3 Operatori monadici

1)EXP,LOG,LOG10,SIN,COS,ATAN,SINH,COSH,ABS,INT

Il simbolo dell'operatore deve precedere il nome della variabile (o la costante) cui deve essere applicato. Più operatori possono essere concatenati, senza bisogno di segni

separatori tra l' uno e l' altro.

Queste funzioni, applicate ad una costante, la trasformano in una nuova costante. Applicate ad una serie, la trasformano in una nuova serie, con le stesse caratteristiche (data iniziale e finale, suddivisione in periodi) della serie data.

2)MAX,MIN,MEAN,VAR

Restituiscono un solo valore numerico, che ha tutte le caratteristiche di una costante (viene liberato da date ecc.). Se però sono associate all' operatore CPR (compressione di una serie secondo un altro operatore), restituiscono una serie (v. CPR).

3)ONE(nn/mm)

Estrae dalla serie il valore numerico corrispondente all' anno e mese (o trimestre, o semestre) indicato. Ad esempio ONE(67/11)'SERIE1' restituisce il valore del mese di novembre dell' anno 1967 ('SERIE1' deve avere dati mensili); ONE(67)'SERIE1' restituisce il primo (o unico) dato dell'anno 1967. Il valore numerico restituito ha tutte le caratteristiche di una costante.

Le parentesi scritte sopra fanno parte integrante dell' operatore; vanno pertanto scritte senza spazi bianchi intermedi, e non intervengono nel conto dei livelli di parentesi.

4)LAG(nn)

Fa scorrere (in un verso solo) i dati di una serie storica

di tante posizioni quante indicate da nn(da 0 a 99). Ad esempio: LAG(4) 'SERIE1', dove SERIE1 comincia nel 1961 e finisce nel 1969, con dati annuali, restituisce una serie che comincia nel 1965 e finisce nel 1969, con il vecchio valore del 61 assegnato al 65 ecc.

Le parentesi scritte sopra fanno parte integrante dell'operatore; vanno pertanto scritte senza spazi bianchi intermedi, e non intervengono nel conto del livello di parentesi.

5)SEL(nn/mm, ll/kk)

Estrae dalla serie i valori compresi fra l'anno e mese(o trimestre) indicato per primo, e quello indicato per secondo. Ad esempio SEL(55/10,66/2)'SERIE1' restituisce una serie con dati mensili (come SERIE1) compresi fra l'ottobre 1955 e il febbraio 1966. Se applicata ad una costante, questa funzione la trasforma in una serie con valori tutti uguali tra loro, con periodi iniziale e finale pari a quelli indicati, e suddivisione in periodi come dal seguente esempio: SEL(55,57)(888.) restituisce una serie di dati annuali dal 1955 al 1957 (tre dati) tutti uguali a 888.; SEL(55/2,61/3)(777.) restituisce una serie di dati trimestrali (massimo tra 2 e 3 arrotondato per eccesso a 4) dal secondo trim. 1955 al terzo del 1961 tutti uguali a 777. ecc.

Le parentesi scritte sopra fanno parte integrante dell'operatore; vanno pertanto scritte senza spazi bianchi

intermedi, e non intervengono nel conto dei livelli di parentesi.

6)RV

Calcola i tassi di variazione di una serie, considerando le differenze tra un dato e il precedente, qualunque sia la suddivisione in periodi. I risultati sono in valore percentuale.

4-4 Operatori composti

CPR(nn)

La funzione CPR(nn) (compressione di una serie secondo un altro operatore) deve essere usata assieme a uno degli operatori MAX,MIN,MEAN,VAR,+,-, mai da sola. Restituisce una serie le cui caratteristiche sono illustrate dagli esempi che seguono: MAX CPR(1) 'SERIE1', dove SERIE1 ha dati mensili, restituisce una serie di valori annuali (1 per anno) uguali al valor massimo tra i valori mensili di ogni anno; MEAN CPR(4) 'SERIE1' restituisce una serie di valori trimestrali (4 per anno) uguali alla media tra i valori di ogni trimestre della serie; 0.+CPR(2)'SERIE1' restituisce una serie di valori semestrali (2 all' anno) uguali alla somma di tutti i valori di ogni semestre della serie data. In questi casi gli operatori MAX,MIN,MEAN,VAR restituiscono una serie invece di una costante.

Le parentesi scritte sopra fanno parte integrante dell'

operatore; vanno pertanto scritte senza spazi bianchi intermedi, e non intervengono nel conto dei livelli di parentesi.

4-5 Funzioni speciali

Rientrano in questo gruppo tutte le routines a caricamento dinamico nell'area libera con alto indirizzo di memoria, e cioè le routines di plotter in linea, le routines di stima (OLS, TSLS, LISE), e la routine di collegamento col DMS/2.

1) OLS (ordinary least squares)

La regressione lineare multipla col metodo dei minimi quadrati ordinari viene realizzata per mezzo delle routines di libreria dell' SSP (Scientific Subroutine Package). Basandosi sul metodo dei minimi quadrati ordinari, tali routines consentono di calcolare medie e deviazioni standard sia della variabile dipendente che di quelle indipendenti, i coefficienti di correlazione semplice e multipla, e di calcolare i coefficienti della regressione. Viene effettuata anche l'analisi di varianza, e, a richiesta, viene stampata la tabella dei residui (differenza tra valori osservati ed interpolati). Sempre a richiesta, e' possibile specificare l'unita' di uscita (6=terminale telescrivente, 8=stampante veloce o, piu' propriamente, FILE FT08F001) dove e' desiderata la stampa dei risultati. Ad esempio:

OLS('SERIE1',RV('SERIE2'+ 'SERIE3'),'SERIE4'); SERIE1 e' la variabile dipendente (viene scritta per prima nella lista delle variabili separate da virgole), mentre le altre due, e cioe' rispettivamente la serie risultante dall'operazione RV('SERIE2'+ 'SERIE3') e la 'SERIE4', sono le Independenti.

Tra i metodi di stima 'single equation' di sistemi di equazioni simultanee sono qui proposti il TSLS e il LISE.

2) TSLS (two stages least squares)

Il comando TSLS ('SERIE1','SERIE2',.....) permette di selezionare tutte le variabili endogene e predeterminate (esogene e endogene ritardate) coinvolte nel sistema da stimare: prima si definiscono tutte le endogene, quindi le predeterminate. Dopo un primo controllo sul numero totale di variabili selezionate, controllo ottenuto mediante la specificazione del numero delle variabili endogene e di quello delle variabili predeterminate, il programma esegue la stima OLS della 'forma ridotta' (primo stadio). Quindi (secondo stadio), per ogni equazione strutturale (al massimo tante quante sono le variabili endogene) l'utente deve introdurre quelle specifiche (numero di variabili coinvolte e indice di selezione) che permettono di selezionare il sottoinsieme di variabili coinvolte in quella equazione. Per ogni equazione e' fornita la stima dei coefficienti e la stima dello standard error dei coefficienti stessi.

3) LISE (limited information single equation)

Il comando LISE ha le stesse caratteristiche operative del TSLS, tuttavia, e anche se cio' non e' avvertito dall'utente, e' bene notare che la stima della 'forma ridotta' e' funzionale soltanto alla stima dello 'standard error', mentre per la stima dei coefficienti delle equazioni strutturali, nel processo iterativo che porta alla massimizzazione del 'variance ratio', quale punto di partenza e' stata scelta la stima OLS dell'equazione stessa.

4) PLOT

Il comando PLOT permette di stampare al terminale un grafico per punti di alcune variabili (dipendenti) rispetto ad un'altra variabile (indipendente). La scala viene aggiustata automaticamente in base ai valori minimo e massimo delle variabili. Ad esempio:

```
PLOT('SERIE1','SERIE2',RV'SERIE3'*MAX'SERIE3'/100.);
```

'SERIE1' e' la variabile indipendente (ascissa), mentre 'SERIE2' e' la serie risultante dall'operazione $RV \text{ 'SERIE3' } * \text{ MAX 'SERIE3' } / 100.$ sono le variabili dipendenti (ordinate).

Viene richiesto in forma conversazionale quale dei due tipi di plotter in linea si voglia, e cioe' se si vuole il plotter normale, con scale e valori numerici, che richiede un discreto tempo di stampa, oppure il plotter veloce, qualitativo, senza scale ne' valori numerici, che,

soprattutto nel caso di un' unica variabile dipendente, permette un considerevole risparmio di tempo.

5)DMS

Il comando DMS ('SERIE1','SERIE2',.....) permette di selezionare alcune serie storiche, scrivendone nomi e valori numerici sul FILE DMS P1; In tale file vengono scritti con il formato richiesto dall' input del DMS/2; I dati sono però ordinati serie per serie, mentre il DMS/2 li richiede ordinati anno per anno; occorre allora utilizzare il SORT del CMS che permette di riordinare il file in ordine di date crescenti. In questo caso serie di lunghezza diversa non vengono troncate, essendo questa operazione demandata al DMS/2.

4-5 Formato dei risultati

Escludendo le funzioni speciali di plotter e regressione, i cui outputs hanno caratteristiche particolari, i risultati di una elaborazione di qualsiasi altro genere sono soggetti alle seguenti regole di formato.

La opzione FILEaaaaaaa scritta alla fine di un comando permette di conservare i risultati, memorizzandoli nel FILE aaaaaaaa (nome qualsiasi) P1, file che può essere utilizzato come input in un comando successivo.

Omettendo questo comando, i risultati vengono stampati al terminale, e non sono riutilizzabili.

Se il risultato finale ha le caratteristiche di una costante, viene stampato (o messo su un file di lavoro) il solo dato numerico; se invece ha le caratteristiche di una serie, vengono stampati (o messi su file) i dati numerici con le indicazioni degli anni e dei mesi (o semestri ecc.) a cui si riferiscono.

Nel caso di serie storiche stampate al terminale, sono forniti due metodi di stampa: il primo viene richiamato scrivendo al terminale soltanto il nome di un' unica serie storica, contenuta nell' archivio centrale o in un file, scritto come sempre tra apici, senza aggiunta di alcun operatore; in questo caso vengono stampati oltre ai dati numerici anche le altre indicazioni (nome completo, unita' di misura, ecc.): ad esempio 'SERIE1'. Il secondo viene richiamato in tutti gli altri casi, anche quando viene compiuta su una sola serie una operazione fittizia (ad esempio 'SERIE1'+0.); in questo caso vengono stampati solo i risultati numerici.

In entrambi i casi i dati numerici vengono stampati uno per riga, in formato F15.6 (o E15.8 se si tratta di numeri troppo grandi) affiancati ciascuno dalla indicazione dell' anno e mese (o sem. o trim) di riferimento.

Nel caso di serie storiche da memorizzare in un file, il formato del file e' quello solito, (descritto in 3-3) con nome completo, unita' di misura e fonte dei dati lasciate in bianco, in quanto, a seguito dell' elaborazione, divenute

prive di significato.

4-7 Errori formali

Nel corso delle operazioni di codifica e decodifica dei comandi, il programma può incontrare i seguenti tipi di errori formali:

- parentesi aperte in numero diverso dalle parentesi chiuse;
- apici aperti per contenere nomi di serie e non richiusi;
- nomi di serie più lunghi di 8 caratteri;
- costanti numeriche senza il punto decimale, o con più di 8 cifre;

- simboli sbagliati, usati al posto degli operatori diadici;
- nomi di operatori monadici sbagliati;

- simboli consentiti, ma usati impropriamente (ad esempio un numero con due punti decimali);

comandi di CMS che, contenendo un carattere sbagliato, vengono rifiutati dal CMS e passati alle routines di decodifica del programma.

Il programma, dopo aver segnalato il tipo di errore, e il nome della routine in cui tale errore è stato trovato (indicazione che per l'utente è normalmente inessenziale, ma può essere di utilità per il sistemista in fase di messa a punto di nuove funzioni), richiama il programma di EDIT del CMS, applicandolo al file GO IMTS, a quel file, cioè, dove di volta in volta viene memorizzato ogni comando

prima della decodifica. Utilizzando i consueti comandi dell'EDIT, l'utente corregge il comando dato, cancella dal file eventuali altri records che non interessano, utilizza il comando FILE, che salva il contenuto del file e ripassa il controllo al programma principale, poi, se vuole, col comando GO rimanda in esecuzione il comando corretto. Se questo contiene ancora errori formali, la procedura ricomincia. Se anzichè correggere il comando memorizzato, l'utente preferisce riscriverlo per intero, basta che esca dall'EDIT col comando QUIT, che lascia il file come sta, e non dia il comando GO.

Sono previsti altri tipi di errori:

serie storiche che non esistono ne' nell'archivio centrale, ne' nel P-DISK dell'utente;

operazioni che non portano a nessun risultato, come ad esempio operazioni aritmetiche su due serie incompatibili, una di dati mensili e una di dati annuali, oppure su due serie compatibili, ma prive di periodi comuni;

divisioni per zero, logaritmi di numeri non positivi, ecc.;

plotter di variabili in cui coincidano minimo e massimo, per cui è impossibile calcolare la scala;

regressioni in cui sia richiesta l'inversione di una matrice che risulti essere singolare.

Per tutti questi tipi di errori, che non rientrano nella categoria degli errori formali, la procedura di correzione

non e' chiamata automaticamente. L'utente, pero', se lo ritiene opportuno, puo' editare il file GO IMTS e procedere come per gli errori formali.

5. PROGRAMMI DI UTILITA'

Rientrano nella categoria dei programmi di utilita' tutte le routines del sistema operativo (residenti nel nucleo del CMS, transienti, o residenti nel disco sistema) richiamabili dall' utente (EDIT, COMBINE, LISTF, ecc.), e alcuni programmi speciali, esterni al programma principale, ma richiamabili dall' interno (CREARCH, NEW). Circa l' allocazione in memoria di questi programmi, si veda il capitolo 2.

Le routines del CMS vengono richiamate dall' interno del programma con gli usuali comandi di CMS. I programmi speciali, essendo contenuti nel medesimo disco in cui e' contenuto l' archivio centrale in formato MODULE, vengono richiamati col solo nome.

1)CREARCH

Questo programma e' stato utilizzato all' inizio per creare l' archivio centrale, e viene utilizzato ogni volta per inserire nuove serie storiche nell' archivio stesso. Puo' essere utilizzato soltanto dalla macchina virtuale che gestisce l' archivio, archivio che, come richiede la filosofia del programma, puo' essere letto da piu' macchine virtuali contemporaneamente, ma modificato da una sola.

2)NEW

E' un programma che permette di creare nel P-DISK dell'utente un file contenente una nuova serie storica, senza problemi di formato dei dati. Legge i dati che vengono scritti di volta in volta, e li mette nel file voluto con il formato corretto. Il file, una volta creato e completato, puo' essere conservato nel P-DISK, oppure inviato alla macchina virtuale che ha accesso in READ/WRITE all'archivio centrale, e da questa la nuova serie puo' essere aggiunta, diventando cosi' utilizzabile da tutti gli utenti.

6. ORGANIZZAZIONE DEI DISCHI

L'archivio, il programma per l'elaborazione dei dati, e i programmi di utilita' sono contenuti in un unico minidisco CMS. Una sola macchina virtuale puo' accedere a questo disco in READ / WRITE. Cio' implica che l'aggiornamento dell'archivio puo' essere fatto solo da quella macchina virtuale.

Tutte le altre macchine virtuali autorizzate possono accedere a questo disco in READ - ONLY mediante il seguente comando CMS: LOGIN 193 A,P.

I files di lavoro di ogni macchina virtuale vengono creati nel disco di lavoro della macchina stessa, e possono, quindi, essere manipolati dall'utente.

Il programma di elaborazione dei dati e' in formato MODULE; puo' quindi essere richiamato scrivendo a terminale soltanto il comando IMTS.

Appena il programma ha preso il controllo, stampa a terminale una freccia e la tastiera si sblocca, in attesa delle operazione da eseguire. Finita una elaborazione, il programma ricomincia dall'inizio scrivendo di nuovo la freccia.

Per chiudere, basta scrivere FINE a colonna 1-4.

7. ESEMPIO DI SEDUTA DI LAVORO AL TERMINALE

```
CP
ipl cms
CMS..VERSION 3.0 (27 ago 73)
```

```
R; T=0.02/0.09 12.08.52
```

```
logln 193 a,p
** A (193) READ-ONLY **
R; T=0.13/0.21 12.09.54
```

```
lmts
```

```
--->
1 file *
FILE NOT FOUND
BAD: 2
```

```
--->
new
```

```
...NUOVA SERIE...
```

```
NOME ABBREVIATO: prova1
SUDDIVISIONE IN PERIODI (Y,S,Q,M): y
ANNO INIZIALE (2 CIFRE): 55
NUME FILE: 'sefile' dimostrazione del 6 dicembre 1973
TIPO DI DATI (LE,FD,IN,RV): le
FONTE DEI DATI: fittizia
UNITA' DI MISURA: fittizia
1955/ 1 1955.
1956/ 1 1956.
1957/ 1 1957.
1958/ 1 1958.
1959/ 1 1959.
1960/ 1 1965.
1961/ 1 1970.
1962/ 1 1980.
1963/ 1 1995.
1964/ 1 2015.
1965/ 1 2040.
```

```
E' STATO CREATO IL FILE PROVA1 P1.
```

```
--->
```

```

--->
max'proval'
      2040.000000

```

```

--->
l file *
FILENAME FILETYPE MODE NO.REC. DATE
FILE      PROVA1   P1      2   12/06

```

```

--->
mean'proval'
      1977.272727

```

```

--->
max'proval'-min'proval'
      85.000000

```

```

--->
rv'proval'
56/ 1      0.051151
57/ 1      0.051125
58/ 1      0.051099
59/ 1      0.051073
60/ 1      0.306279
61/ 1      0.254453
62/ 1      0.507614
63/ 1      0.757576
64/ 1      1.002506
65/ 1      1.240695

```

```

--->
edit tabella abbrev
DEFAULT TABS SET.
EDIT;
print 4
CAPA   CPA     CPR     DI      DIS     HF      HPREFD  HTIN    HTRE    HVA
HVAD   HVAI    HXMO    HY      HYD     HYDW    HYW     HYWI    HYWPA  HYWR
HYWTA  I       IAB     IES     II      IT      K       KMA     KMAX   LE
LI     LIH     LPA     LT      MM      MS      O       PA     PAB    TVPCV
quit

```

```

--->
max sel(60,64) rv'proval'
      1.002506

```

```

--->

```


--->

'capa'

NUMERO DELLA SERIE: 1
 NOME ABBREVIATO: CAPA
 ANNO INIZIALE: 1951
 ANNO FINALE: 1970
 NOME INTERO: ACQUISTI P.A. DAL SETTORE PRIVATO
 FONTE DEI DATI: (CPA-VAPA)
 UNITA' DI MISURA: MILIARDI LIRE 1963
 I DATI SONO ANNUALI.
 LIVELLI.

1951/ 1	429.000000
1952/ 1	479.000000
1953/ 1	482.000000
1954/ 1	567.000000
1955/ 1	524.000000
1956/ 1	567.000000
1957/ 1	532.000000
1958/ 1	608.000000
1959/ 1	635.000000
1960/ 1	680.000000
1961/ 1	740.000000
1962/ 1	863.000000
1963/ 1	942.000000
1964/ 1	974.000000
1965/ 1	1039.000000
1966/ 1	1045.000000
1967/ 1	1159.000000
1968/ 1	1251.000000
1969/ 1	1341.000000
1970/ 1	1391.000000

--->

max rv'capa'

17.634855

--->

min rv'capa'

-7.583774

--->

one(63)'capa'

942.000000

--->

--->

lag(1)rv'capa'

53/ 1	11.655012
54/ 1	0.626305
55/ 1	17.634855
56/ 1	-7.583774
57/ 1	8.206107
58/ 1	-6.172840
59/ 1	14.285714
60/ 1	4.440789
61/ 1	7.086614
62/ 1	8.823529
63/ 1	16.621622
64/ 1	9.154114
65/ 1	3.397028
66/ 1	6.673511
67/ 1	0.577478
68/ 1	10.909091
69/ 1	7.937877
70/ 1	7.194245

--->

rv log'capa' fileprova2

--->

p file prova2

1PROVA2 Y52 170 1 19

52/ 1	1.818765
53/ 1	0.101164
54/ 1	2.628952
55/ 1	-1.243898
56/ 1	1.259566
57/ 1	-1.004924
58/ 1	2.127433
59/ 1	0.677830
60/ 1	1.060920
61/ 1	1.296476
62/ 1	2.327420
63/ 1	1.295639
64/ 1	0.487821
65/ 1	0.938800
66/ 1	0.082899
67/ 1	1.489414
68/ 1	1.082669
69/ 1	0.974135
70/ 1	0.508352

ENDFILE

--->

```
--->
ols ('capa', 'vapa', 'cpa', 'vapa')
```

SONO DISPONIBILI 20 DATI PER OGNI SERIE, DAL 1951/ 1 AL 1970/ 1 : 1 PER ANNO.
EXECUTION BEGINS...

1 SYMBOLIC UNIT FOR THE RESULTS. (6-TERMINAL,8-OFFLINE)
2-9 NAME
10-11 (-1,00,01) RESIDUALS.

```
-.....--
6regress1-1
MULTIPLE REGRESSION.....REGRESS1
```

THE MATRIX IS SINGULAR

```
--->
ols('capa', 'cpa', lag1@(1)'capa', lag(1)'cpa', rv'cpa')
```

SONO DISPONIBILI 19 DATI PER OGNI SERIE, DAL 1952/ 1 AL 1970/ 1 : 1 PER ANNO.
1 SYMBOLIC UNIT FOR THE RESULTS. (6-TERMINAL,8-OFFLINE)

2-9 NAME
10-11 (-1,00,01) RESIDUALS.

```
-.....--
6regress2-1
MULTIPLE REGRESSION.....REGRESS2
```

VARIABLE NO.	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD. ERROR OF REG. COEF.	COMPUTED T VALUE
2	0.99264	1.71679	1.08847	1.57726
3	0.99024	-1.35126	0.23121	5.84436
4	0.99243	-1.88413	1.16257	-1.62066
5	0.13090	-14.21501	33.60046	-0.42306
DEPENDENT 1				

```
INTERCEPT                204.42244
MULTIPLE CORRELATION        0.99795
STD. ERROR OF ESTIMATE      22.08254
STD. DETERMINANT            0.00000
```

SOURCE	D.F.	A N O V A		F VALUE
		SUM OF SQUARES	MEAN SQUARES	
REGRESSION	4	1661636.4724	415409.1181	851.87907
RESIDUAL	14	6826.9404	487.6386	
TOTAL	18	1668463.4128		

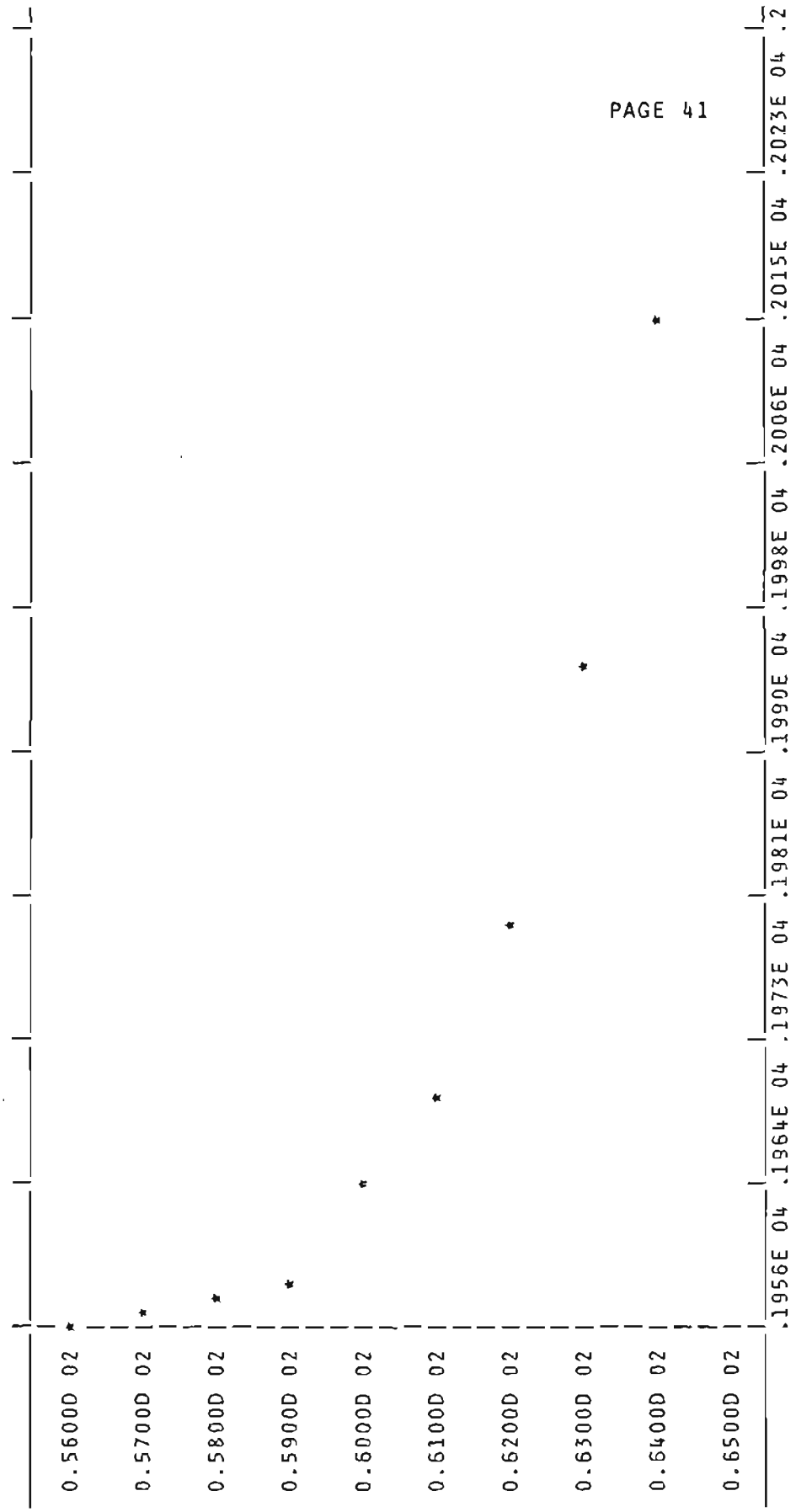
DURBIN WATSON STATISTIC = 2.4442

```
--->
```

--->
plot(sel(56,65)'ytime', 'proval')

SONO DISPONIBILI 10 DATI PER OGNI SERIE, DAL 1956/ I AL 1965/ I : I PER ANNO.
EXECUTION BEGINS...
PLOTTER NORMALE O VELOCE ? (0/1)

0
I SIMBOLI PER LE VARIABILI DIPENDENTI SONO, NELL' ORDINE: *, +, o, x ; & SIGNIFICA SOVRAPPOSIZIONE



--->