

## PACCHETTO STATISTICO

Insieme di programmi  
+  
Linguaggio

### Origine dei pacchetti

|           |                         |
|-----------|-------------------------|
| BMDP      | Scienze biomediche      |
| GENSTAT   | Scienze agrarie         |
| GLIM      | Statistica              |
| MINITAB   | Insegnamento statistica |
| S-plus, R | ricerca                 |
| SPSS      | Scienze sociali         |
| STATA     | Scienze biomediche      |
| SYSTAT    | Psicologia              |

## CARATTERISTICHE PACCHETTI

### (1) INTERFACCIA UTENTE

- Command driven (SAS)
- Menu driven (STATISTICA)

### (2) STRUTTURE DATI

Solitamente strutture rettangolari.

+

- Dati relazionali (SAS, DATA DESK)
- Matrici (S-PLUS, GAUSS, SPSS, SAS)

## CARATTERISTICHE PACCHETTI

### (3) GRAFICI

Utili per:

- analisi esplorative;
- presentazione analisi (report);
- plot per adattamento modello.

Alcuni pacchetti consentono grafici dinamici (p.e. SAS, S-plus).

### (4) INTERFACCIA STATISTICA

- Descrittiva+test (SPSS, SYSTAT)
- Modelli+test (S-PLUS, GAUSS)

## CARATTERISTICHE PACCHETTI

### (5) PROCEDURE STATISTICHE

Tutti i pacchetti dispongono di procedure per statistiche descrittive.

Differenza principale tra pacchetti: *tipo* di test e statistiche *direttamente* disponibili:

- Regress. tutti
- GLM GLIM, GENSTAT, S-Plus, SAS
- ANOVA tutti, GENSTAT più flessibile
- molte PROC SPSS, BMDP, SYSTAT, SAS, STATA, S-PLUS

*Pacchetti specializzati (selezione)*

- |           |                         |
|-----------|-------------------------|
| Clustan   | cluster analysis        |
| StatExact | test esatti             |
| XlispStat | grafici dinamici        |
| Echip     | esperimenti industriali |

## CARATTERISTICHE PACCHETTI

### (6) FLESSIBILITÀ

*È possibile eseguire calcoli o predisporre algoritmi non previsti dal pacchetto?*

- Impossibile con alcuni (DataDesk, Statgraphics)
- abbastanza facile per pacchetti con linguaggio (SAS, SPSS, S-plus, etc)
- librerie di dominio pubblico in Statlib (p.e. MINITAB, S-plus, STATA, SAS)

## CARATTERISTICHE PACCHETTI

### (7) PIATTAFORMA

- Mac DataDesk
- Windows Statgraphics, S-plus, R
- Unix, Linux S-plus, R

Molti pacchetti disponibili su più piattaforme (p.e. SAS, SPSS, MINITAB, R).

## SOFTWARE STATISTICO DISPONIBILE PRESSO IL DIPARTIMENTO

### *Pacchetti generali (selezione)*

|              |               |
|--------------|---------------|
| SPSS         | Windows       |
| <b>SAS</b>   | Windows, Unix |
| <b>STATA</b> | Windows       |
| STATISTICA   | Mac, Windows  |
| MINITAB      | Mac, Windows  |
| STATGRAPHICS | Dos           |
| JUMP         | MAC           |

### *Linguaggi con molte proc*

|                  |               |
|------------------|---------------|
| S-PLUS, <b>R</b> | Windows, Unix |
| GAUSS            | Dos, Unix     |

### *Pacchetti specializzati (selezione)*

|            |               |
|------------|---------------|
| XLISP-STAT | Windows, Unix |
| MLWIN      | Windows       |
| EGRET      | Windows       |
| LISREL     | Windows       |
| MATLAB     | Windows       |
| MAPLE      | Windows, Mac  |

## VALUTAZIONE DI UN PACCHETTO

- Procedure offerte;
- possibilità import/export files;
- trasformazione dati;
- disponibilità e qualità grafici;
- tipo interfaccia: command vs menu driven;
- documentazione e help in linea;
- macro linguaggio per automatizzare;
- macro linguaggio per aggiungere e/o modificare procedure.

## IL SISTEMA SAS

Sistema integrato di prodotti software

- data entry, manipolazione archivi;
- stesura di report e grafici;
- analisi statistiche e matematiche;
- previsioni e supporto alle decisioni;
- ricerca operativa e project management;
- sviluppo di applicazioni.

Il “nocciolo” del sistema SAS è il modulo  
**SAS Base**

- linguaggio SAS;
- procedure per analisi dei dati e stesura di report;
- macro-linguaggio;
- Display Manager System.

## STRUTTURA DEL SISTEMA SAS

Analisi dei dati → 2 passi fondamentali:

- (1) organizzazione dei dati;
- (2) analisi dei dati.

*Sistema SAS:*

- (1) DATA step: creazione e manipolazione SAS data set;
- (2) PROC step: analisi dei dati attraverso le procedure SAS o procedure utente.

## STRUTTURA PROGRAMMA SAS

### *Data step*

consente di creare o modificare archivi (*data set*) SAS

### *Proc step*

Richiede di richiamare una procedura fra quelle disponibili e di eseguirla su un SAS data set.

### *Risultati procedure*

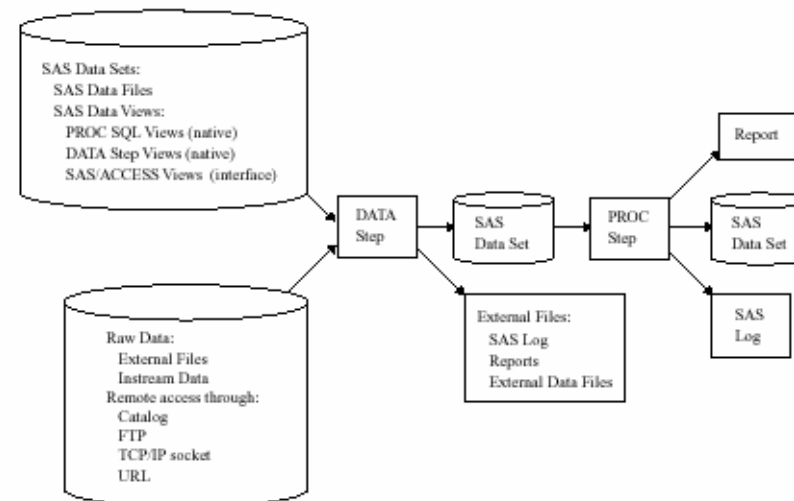
- File ASCII (report);
- SAS data set;
- ODS

*Programma SAS può essere costituito da:*

- un solo Data step;
- un solo Proc step;
- più Data step e/o Proc step.

## SAS PROCESSING

Figure 2.1 SAS Processing



Lettura dei dati (INPUT):

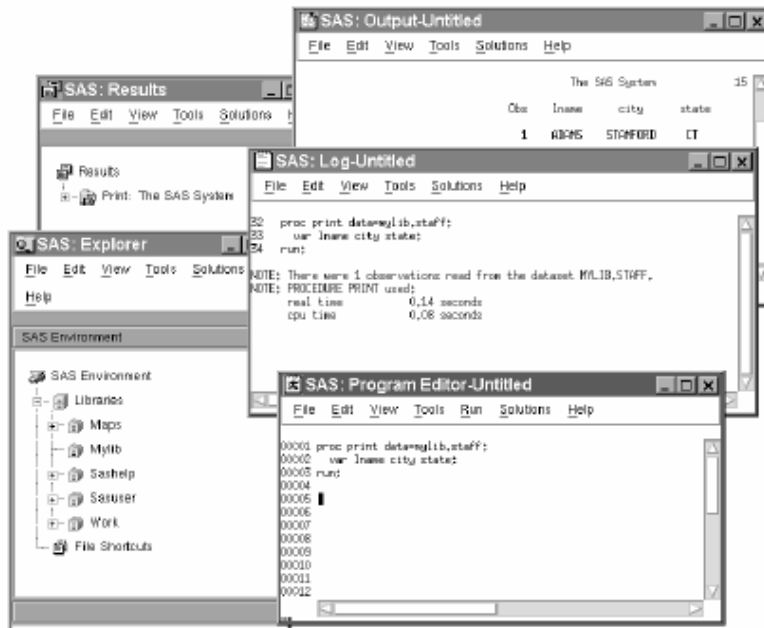
- SAS data set
- Viste
- Dati grezzi

Scrittura dati e risultati (OUTPUT):

- report
- SAS Log
- SAS Data set

## AMBIENTE DI LAVORO A FINESTRE

Figure 1.1 SAS Windowing Environment



## COMPONENTI DEL LINGUAGGIO SAS

**SAS Files:** risiedono in librerie e sono di vario tipo (data sets, catalogs, programmi compilati)

**SAS Data sets:** data file, “viste”

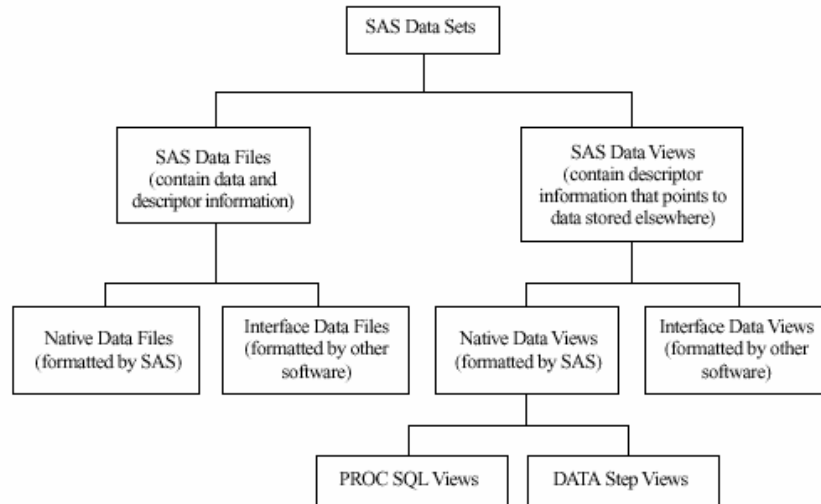
**Files esterni:** dati “grezzi”, istruzioni, output delle procedure

**Linguaggio:** istruzioni, espressioni, opzioni, formati, funzioni

**Macro facility:** estensione programmi SAS, riduzione istruzioni.

# TIPI DI SAS DATA SETS

Figure 28.1 Types of SAS Data Sets



# SAS DATA SET

Files binari generati dal SAS.

Dati organizzati in *forma rettangolare*

|  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

osservazione

variabile

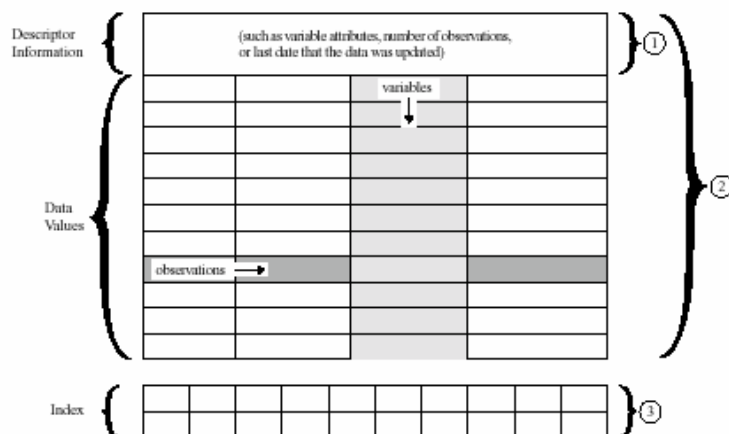
- ogni riga rappresenta una osservazione;
- ogni colonna rappresenta una variabile;
- tutte le osservazioni possiedono le stesse variabili → le variabili non presenti in una data osservazione sono registrate come mancanti (*missing*).



## SAS DATA SET

Ogni SAS data set è auto-descrittivo.

Figure 27.1 Logical Components of a SAS Data Set



Componenti logiche del SAS data set:

- (1) DIA Descriptor Information Area
- (2) Dati
- (3) Indici

## IL LINGUAGGIO SAS

Data step e Proc step sono costituiti da *istruzioni* elementari.

### *Istruzioni SAS*

- Solitamente iniziano con una parola chiave;
- terminano con il carattere “;” (punto e virgola);
- possono essere scritte indifferentemente con minuscole o maiuscole;
- possono iniziare in una qualsiasi colonna della riga e proseguire su più righe;
- più istruzioni possono essere scritte sulla stessa riga.

### *Istruzioni:*

- solo Data step;
- solo Proc step;
- comuni.

## ESECUZIONE PROGRAMMI SAS

Esecuzione in modo semi-interattivo tramite DMS (Display Manager System)

### *DMS ambiente a finestre:*

- stesura programma;
- visualizzazione risultati;
- segnalazione inviate dal sistema (messaggi di errore, warning e altro).

Il programma SAS viene compilato automaticamente dal sistema prima di essere eseguito:

- compilazione e esecuzione avvengono “a blocchi” (Data step e Proc step);
- blocchi chiusi da successiva parola chiave DATA o PROC o da istruzione RUN.

## INTRODUZIONE AL DATA STEP

Creazione di archivi

- (a) da file “esterni”, per esempio file ASCII;
- (b) da file “interni”, cioè da SAS data set.

### *Struttura Data step*

**DATA** *nome (opzioni)*;

.....

.....

.....

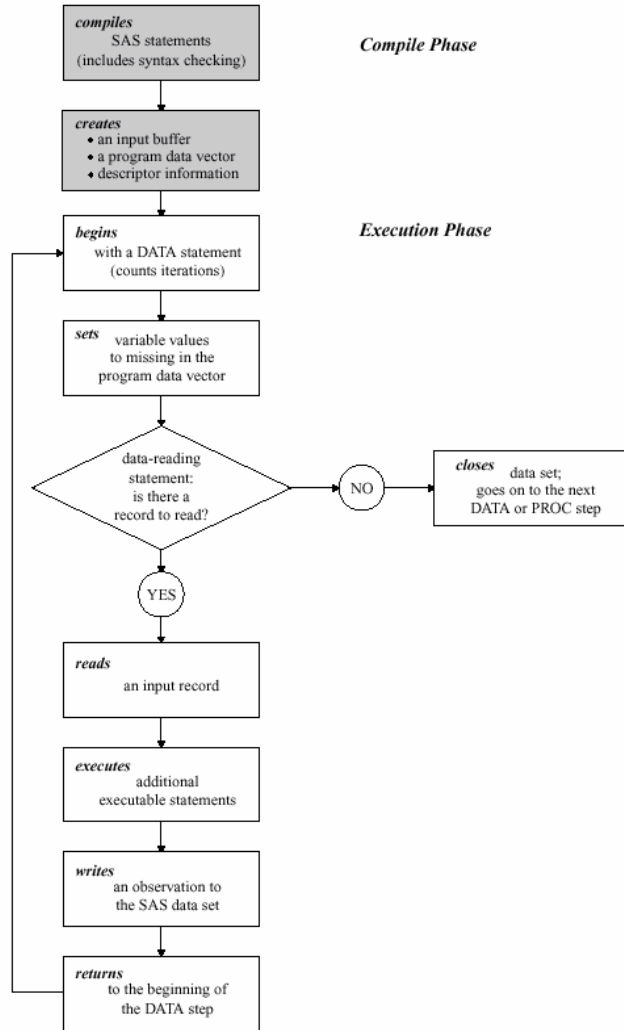
**RUN**;

*Istruzioni fondamentali per la lettura di file ASCII:*

- INPUT
- INFILE

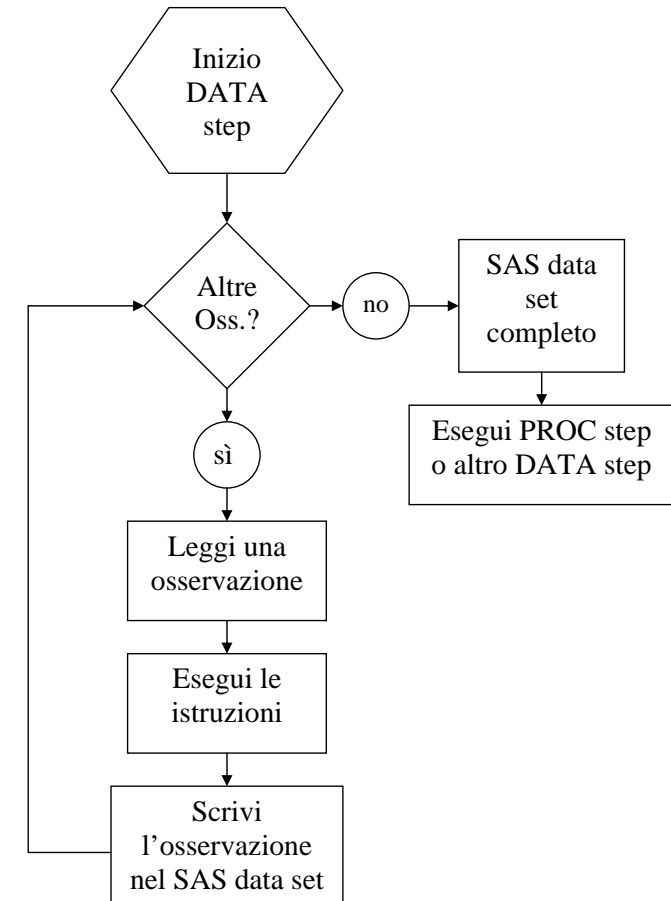
## FLUSSO DI LETTURA NEL DATA STEP

Figure 21.1 Flow of Action in the DATA Step



NB: il flusso di lettura di default può essere alterato in tutte le sue parti attraverso opportune istruzioni.

## FLUSSO DI LETTURA FILE DI TESTO



## ESECUZIONE PROGRAMMA SAS

Dopo la fase di compilazione, se non ci sono errori, il sistema predispone nella RAM per l'esecuzione del programma:

- (a) Input Buffer
- (b) Program Data Vector
- (c) Descriptor information Area (DIA)

### *Input Buffer*

Creato solo per lettura da file esterni (INPUT): area di memoria in cui viene trasferito l'intero record letto.

### *PDV*

Area di memoria per la costruzione obs da scrivere nel SAS data set.

*DIA* (Descriptor Information Area) contiene info gestite dal sistema e memorizzate con SAS data set.

## PROGRAM DATA VECTOR

L'istruzione INPUT legge i dati dal record nell'Input Buffer e li scrive nel PDV dove diventano valori da assegnare alle variabili del data set SAS.

```
Data pippo;  
input id $ x1 x2;  
y=x2-x1;  
cards;  
2477 195 163  
2431 220 198  
2456 173 155  
;
```

### *input buffer*

```
-----+-----1-----+-----2-----+-----3-----+...  
2477 195 163
```

### *PDV*

| ID   | X1  | X2  | Y | _N_ | _ERROR_ |
|------|-----|-----|---|-----|---------|
| 2477 | 195 | 163 | . | 1   | 0       |
| K    | D   | D   | K | D   | D       |

Y viene riempita quando si esegue istr assegnazione.

## INPUT BUFFER E PROGRAM DATA VECTOR

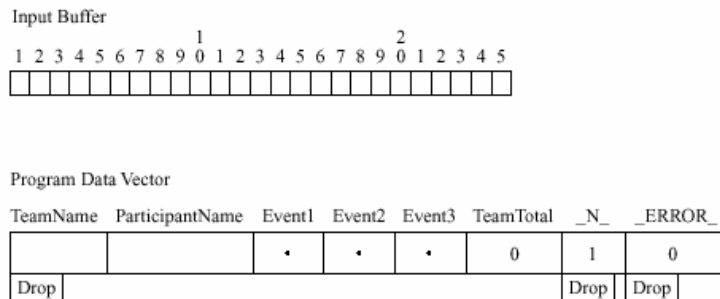
```

data total_points (drop=TeamName); ❶
  input TeamName $ ParticipantName $ Event1 Event2 Event3; ❷
  TeamTotal + (Event1 + Event2 + Event3); ❸
  datalines;
Knights Sue      6 8 8
Cardinals Jane   9 7 8
Knights John     7 7 7
Knights Lisa     8 9 9
Knights Fran     7 6 6
Knights Walter   9 8 10
;

```

- (1) L'opzione DROP= evita che la variabile TeamName sia scritta nell'output SAS data set TOTAL\_POINTS;
- (2) L'istruzione INPUT descrive il record dei dati, per ogni variabile: assegna un nome, identifica il tipo di dato (carattere o numerico), e la sua posizione relativa nel record.
- (3) l'istruzione Sum accumula I punteggi dei tre eventi nella variabile TeamTotal.

Figure 21.2 Input Buffer and Program Data Vector



## ISTRUZIONE DATA

**DATA** [*SASdataset*[(*opzioni*)]] ...;

*SASdataset*

nome archivio SAS creato dal Data step.  
 Nome a due livelli separati da un punto:

*nome1.nome2*

*nome1* indica la libreria in cui memorizzare il data set (default *WORK*, corrispondente alla directory *\sas\saswork*);

*nome2* utilizzato dal sistema operativo come nome del file, con estensione *.sas7bdat*, in cui memorizzare il SAS data set creato (default *DATAN*).

## ISTRUZIONE LIBNAME

Consente di associare il riferimento ad una libreria SAS (*libref*) con il nome fisico di una libreria SAS permanente. Il nome fisico è il *percorso* di una directory.

**LIBNAME** *libref* '*percorso*';

## PAROLE SAS

### *parola SAS*

insieme di caratteri che comunica una informazione a SAS e non è divisibile in unità indipendenti significative.

### *4 tipi fondamentali:*

- **nome:** successione di caratteri, inizia per lettera o \_
- **stringa:** successione di caratteri inclusa tra apici (lunghezza max 32767)
- **numero:** insieme di cifre, punto decimale, segno; ammessi notazione scientifica, formato esadecimale, date e tempi
- **caratteri speciali:** = @ & % ecc.

### *Spaziatura*

- Lo spazio non è considerato un carattere, a meno che non sia incluso tra apici;
- Inizio e fine di una parola sono riconosciuti dal sistema per la presenza di operatori o di spazi bianchi:
  - Spazi non necessari Total = x+y
  - Spazi indispensabili input yy 15 x 20

## NOMI SAS

Due tipi:

- nomi di elementi del linguaggio SAS
- nomi definiti dall'utente

**Table 3.1** Maximum Length of User-Supplied SAS Names

| SAS Language Element                                                                                    | Maximum Length                 |
|---------------------------------------------------------------------------------------------------------|--------------------------------|
| Members of SAS data libraries (SAS data sets, views, catalogs, indexes) except for generation data sets | 32                             |
| Generation data sets                                                                                    | 28                             |
| Catalog entries                                                                                         | 32                             |
| Engines                                                                                                 | 8                              |
| Librefs                                                                                                 | 8                              |
| Filerefs                                                                                                | 8                              |
| Passwords                                                                                               | 8                              |
| DATA step variables                                                                                     | 32                             |
| DATA step variable labels                                                                               | 256                            |
| DATA step statement labels                                                                              | 32                             |
| Arrays                                                                                                  | 32                             |
| DATA step windows                                                                                       | 32                             |
| Functions                                                                                               | 16                             |
| CALL routines                                                                                           | 16                             |
| Formats                                                                                                 | 8                              |
| Informats                                                                                               | 7                              |
| Macros                                                                                                  | 32                             |
| Macro variables                                                                                         | 32                             |
| Macro windows                                                                                           | 32                             |
| SCL variables                                                                                           | 32                             |
| SAS/EIS items                                                                                           | depends on the data dictionary |
| Procedure names (First 8 characters must be unique, and may not begin with "SAS")                       | 16                             |

## REGOLE PER NOMI SAS DEFINITI DALL'UTENTE

- Numero massimo di caratteri dipende dal tipo di nome (32, 8, altro);
- il primo carattere deve essere una lettera o un “\_” (underscore);
- i caratteri successivi al primo possono essere lettere, cifre o “\_”;
- il sas non fa differenza tra lettere maiuscole e minuscole e converte tutti i nomi in maiuscolo;
- non possono esserci spazi bianchi;
- non sono ammessi caratteri speciali (\$, &, \*, ..), a meno di “\_”;
- non si possono utilizzare nomi di variabili automatiche (p.e. `_N_` o `_ERROR_`);
- non si possono usare nomi che il SAS riserva a librerie speciali (p.e. LIBRARY, SASHELP, WORK, USER).

## LETTURA DATI DI TIPO ASCII

- istruzione INFILE;
- istruzioni DATALINES e CARDS;
- istruzione INPUT.

### *Istruzione INFILE*

Indica al sistema dove leggere i dati, definendo anche le caratteristiche del file contenente i dati.

### *Istruzione INPUT*

Definisce le variabili da leggere per ogni record del file di dati.

### *Istruzione DATALINES (CARDS)*

Indica che i dati sono inseriti direttamente nel programma e seguono immediatamente questa istruzione.

## ISTRUZIONE INFILE

INFILE '*nome file*' [*opzioni*];

*'nome file'*

nome, con eventuale percorso, del file ASCII da leggere o parola chiave CARDS se i dati sono inseriti da programma.

*opzioni* (selezione)

LRECL=*m*

lunghezza max record (utile se record >132 caratteri);

FIRSTOBS= *m*

numero d'ordine del record da cui iniziare la lettura;

OBS= *m*

numero di record da leggere.



## ISTRUZIONE INPUT

Definisce nome, tipo e modo di lettura delle variabili.

- Variabili numeriche in virgola mobile (8 byte), variabili carattere identificate da \$, numero di byte varia a seconda del tipo di lettura;
- Modo di lettura: descrizione della forma in cui sono registrati i dati e posizionamento del record che contiene i dati.

### *Descrizione forma di registrazione*

- Input a lista;
- input a colonna;
- input con formato.

### *Posizionamento nel record dati*

/, @, #

## LETTURA A LISTA

Possibile quando i dati sono registrati in formato libero, con almeno uno spazio bianco tra un campo e il successivo.

**INPUT** *var1 var2 ...;*

**INPUT** *x1-x10 ...;*

**INPUT** *var1 \$ var2 ...;*

```
data pippo;
input x1-x3;
cards;
1 5 7
9 3
2
6 9 8 10 12
13 5 8
;
run;
```

|  | x1 | x2 | x3 |
|--|----|----|----|
|  | 1  | 5  | 7  |
|  | 9  | 3  | 2  |
|  | 6  | 9  | 8  |
|  | 13 | 5  | 8  |

## INPUT A COLONNA

```
INPUT var [$] coll[-col2][.dec][@|@@];
```

```
INPUT nome $ 1-12 peso 13-15 eta 16-17
```

- Indicazione colonna iniziale e finale del campo occupato dalla variabile.
  - Le variabili devono essere memorizzate nelle stesse posizioni per tutte le obs;

### *Vantaggi*

- non occorre che i campi corrispondenti alle variabili siano separati da *blanks*;
- non è necessario indicare esplicitamente i *missing*: un campo vuoto è interpretato come *missing*;
- non si è vincolati a leggere le variabili sequenzialmente:

```
INPUT nome $ 1-12 eta 16-17 peso 13-15
```

## INPUT A COLONNA

- È possibile specificare il punto decimale:

| dati in input      | Istruzione SAS     | output |
|--------------------|--------------------|--------|
| -----+-----1-----+ | input num 5-10 .2; |        |
| 2314               |                    | 23.14  |
| 2                  |                    | .02    |
| -140               |                    | -1.40  |
| 12.2               |                    | 12.2   |

Prevale il punto decimale che compare nei dati in lettura.

- Quando le variabili occupano più record è possibile controllare lo spostamento nei record successivi:

/ passa a col. 1 del record successivo;  
#n passa a col. 1 del record *n*-esimo.

## LETTURA CON FORMATO

**INPUT** *var informat.;*

**INPUT** (*varlist*) (*informat list*);

**INPUT** (*varlist*) (*[n\*]informat.*);

Consente di indicare il formato (tipo di dato e ampiezza del campo) di lettura per ogni variabile del file di input.

Sintassi generale *informat*:

- w.* legge numeri interi o numeri decimali con punto decimale codificato nel campo;
- w.d* legge numeri con cifre decimali, con punto decimale non codificato nel campo;
- \$w.* legge stringhe di caratteri ASCII.

## ESEMPI INPUT CON FORMATO

Es. 1 - Input a formato: controllo puntatore

```
data pippo;
infile 'a:\prova.dat';
input nome $10. +5 eta 2. +2 voto 4.;
run;
```

```
-----1-----2-----+
giuseppe          24  30
maria              22  26
laura              23  18
franco             21  22
```

Es. 2 - Input a formato: uso lista di formati

```
data pippo;
input (nome test1-test5)($10. 5*4.);
cards;
tizio          121 114 137 156 142
caio           111  97 122 143 127
;
```

oppure

```
input nome $10. (test1-test5)(4.);
```

## Es. 3 - Input misto colonna/formato

```
data pippo;
infile 'a:\dati.txt'
input nome $ 1-10 @16 eta 2. @20 voto 4.;
run;
```

```
-----+-----1-----+-----2-----+
giuseppe          24    30
maria              22    26
laura              23    18
franco             21    22
```

### *input a formato*

- Valori mancanti:
  - . per variabili numeriche
  - blank per variabili carattere;
- le variabili devono essere elencate nello stesso ordine in cui si trovano nel file di lettura.
- Indispensabile per lettura dati non standard, per date e tempi

Table 5.4 Categories and Descriptions of Informats

| Category                 | Informat                                                     | Description                                                                                        |
|--------------------------|--------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| Character                | "\$ASCIIw." on page 644                                      | Converts ASCII character data to native format                                                     |
|                          | "\$BINARYw." on page 645                                     | Converts binary data to character data                                                             |
|                          | "\$CHARw." on page 647                                       | Reads character data with blanks                                                                   |
|                          | "\$CHARZBW." on page 648                                     | Converts binary 0s to blanks                                                                       |
|                          | "\$EBCDICw." on page 649                                     | Converts EBCDIC character data to native format                                                    |
|                          | "\$HEXw." on page 650                                        | Converts hexadecimal data to character data                                                        |
|                          | "\$OCTALw." on page 652                                      | Converts octal data to character data                                                              |
|                          | "\$PHEXw." on page 653                                       | Converts packed hexadecimal data to character data                                                 |
|                          | "\$QUOTEw." on page 654                                      | Removes matching quotation marks from character data                                               |
|                          | "\$REVERJw." on page 654                                     | Reads character data from right to left and preserves blanks                                       |
|                          | "\$REVERS w." on page 655                                    | Reads character data from right to left and left aligns                                            |
|                          | "\$UPCASEw." on page 656                                     | Converts character data to uppercase                                                               |
|                          | "\$VARYINGw." on page 657                                    | Reads character data of varying length                                                             |
| Column Binary            | "\$w." on page 659                                           | Reads standard character data                                                                      |
|                          | "\$CBw." on page 646                                         | Reads standard character data from column-binary files                                             |
|                          | "\$CBw.d" on page 663                                        | Reads standard numeric values from column-binary files                                             |
|                          | "\$PUNCH.d" on page 705                                      | Reads whether a row of column-binary data is punched                                               |
| DBCS                     | "\$ROWw.d" on page 710                                       | Reads a column-binary field down a card column                                                     |
|                          | "\$KANJIw." on page 651                                      | Removes shift code data from DBCS data                                                             |
| Date and Time            | "\$KANJIXw." on page 651                                     | Adds shift code data to DBCS data                                                                  |
|                          | "\$DATEw." on page 666                                       | Reads date values in the form <i>ddmmyy</i> or <i>ddmmmyyy</i>                                     |
|                          | "\$DATETIMEw." on page 667                                   | Reads datetime values in the form <i>ddmmyy hh:mm:ss</i> or <i>ddmmmyyy hh:mm:ss</i>               |
|                          | "\$DDMMYYw." on page 669                                     | Reads date values in the form <i>ddmmyy</i> or <i>ddmmmyyy</i>                                     |
|                          | "\$EURDFDEw." on page 671                                    | Reads international date values                                                                    |
|                          | "\$EURDFDTw." on page 673                                    | Reads international datetime values in the form <i>ddmmyy hh:mm:ss</i> or <i>ddmmmyyy hh:mm:ss</i> |
|                          | "\$EURDFMYw." on page 675                                    | Reads month and year date values in the form <i>mmyy</i> or <i>mmmyyy</i>                          |
|                          | "\$JDATEYMDw." on page 683                                   | Reads Japanese kanji date values in the format <i>yyymmdd</i> or <i>yyyymmdd</i>                   |
| "\$JNENGOw." on page 684 | Reads Japanese Kanji date values in the form <i>yyymmdd</i>  |                                                                                                    |
| "\$JULIANw." on page 685 | Reads Julian dates in the form <i>yyddd</i> or <i>yyyddd</i> |                                                                                                    |
| "\$MINGUOw." on page 686 | Reads dates in Taiwanese form                                |                                                                                                    |

"MMDDYYw." on page 687 Reads date values in the form *mmddy* or *mmddyyyy*

"MONYYw." on page 689 Reads month and year date values in the form *mmmyy* or *mmmyyyy*

"MSECw." on page 690 Reads TIME MIC values

"NENGOw." on page 691 Reads Japanese date values in the form *eymmdd*

"PDJULGw." on page 696 Reads packed Julian date values in the hexadecimal form *yyyydddF* for IBM

"PDJULw." on page 697 Reads packed Julian dates in the hexadecimal format *ccyyddd F* for IBM

"PDTIMEw." on page 699 Reads packed decimal time of SMF and RMF records

"RMFDURw." on page 707 Reads duration intervals of RMF records

"RMFSTAMPw." on page 709 Reads time and date fields of RMF records

"SHRSTAMPw." on page 711 Reads date and time values of SHR records

"SMFSTAMPw." on page 713 Reads time and date values of SMF records

"TIMEw." on page 727 Reads hours, minutes, and seconds in the form *hh:mm:ss*

"TODSTAMPw." on page 729 Reads an eight-byte time-of-day stamp

"TUw." on page 730 Reads timer units

"YYMDDw." on page 733 Reads date values in the form *yyymmdd* or *yyyymmdd*

"YYMMNw." on page 735 Reads date values in the form *yyyymm* or *yyym*

"YYQw." on page 736 Reads quarters of the year

"BINARYw.d" on page 660 Converts positive binary values to integers

"BITSw.d" on page 661 Extracts bits

"BZw.d" on page 662 Converts blanks to 0s

"COMMAw.d" on page 664 Removes embedded characters

"COMMAXw.d" on page 665 Removes embedded characters

"Ew.d" on page 670 Reads numeric values that are stored in scientific notation and double-precision scientific notation

"FLOATw.d" on page 677 Reads a native single-precision, floating-point value and divides it by 10 raised to the *d*th power

"HEXw." on page 678 Converts hexadecimal positive binary values to either integer (fixed-point) or real (floating-point) binary values

"IBw.d" on page 679 Reads native integer binary (fixed-point) values, including negative values

"IBRw.d" on page 680 Reads integer binary (fixed-point) values in Intel and DEC formats

"IEEEw.d" on page 682 Reads an IEEE floating-point value and divides it by 10 raised to the *d*th power

"NUMXw.d" on page 693 Reads numeric values with a comma in place of the decimal point

"OCTALw.d" on page 694 Converts positive octal values to integers

"PDw.d" on page 695 Reads data that are stored in IBM packed decimal format

"PERCENTw.d" on page 700 Reads percentages as numeric values

"PIBw.d" on page 701 Reads positive integer binary (fixed-point) values

"PIBRw.d" on page 702 Reads positive integer binary (fixed-point) values in Intel and DEC formats

"PKw.d" on page 704 Reads unsigned packed decimal data

"RBw.d" on page 706 Reads numeric data that are stored in real binary (floating-point) notation

"S370FFw.d" on page 714 Reads EBCDIC numeric data

"S370FIBw.d" on page 715 Reads integer binary (fixed-point) values, including negative values, in IBM mainframe format

"S370FIBUw.d" on page 716 Reads unsigned integer binary (fixed-point) values in IBM mainframe format

"S370FPDw.d" on page 718 Reads packed data in IBM mainframe format

"S370FPDUw.d" on page 719 Reads unsigned packed decimal data in IBM mainframe format

"S370FPPIBw.d" on page 720 Reads positive integer binary (fixed-point) values in IBM mainframe format

"S370FRBw.d" on page 721 Reads real binary (floating-point) data in IBM mainframe format

"S370FZDw.d" on page 722 Reads zoned decimal data in IBM mainframe format

"S370FZDLw.d" on page 723 Reads zoned decimal leading-sign data in IBM mainframe format

"S370FZDSw.d" on page 724 Reads zoned decimal separate leading-sign data in IBM mainframe format

"S370FZDTw.d" on page 725 Reads zoned decimal separate trailing-sign data in IBM mainframe format

"S370FZDUw.d" on page 726 Reads unsigned zoned decimal data in IBM mainframe format

"VAXRBw.d" on page 731 Reads real binary (floating-point) data in VMS format

"w.d" on page 731 Reads standard numeric data

"YENw.d" on page 732 Removes embedded yen signs, commas, and decimal points

"ZDw.d" on page 738 Reads zoned decimal data

"ZDBw.d" on page 739 Reads zoned decimal data in which zeros have been left blank

"ZDVw.d" on page 740 Reads and validates zoned decimal data

## LETTURA A FORMATO

INDISPENSABILE NEL CASO IN CUI  
CI SIANO DA LEGGERE CAMPI  
DATA

### Esempio (punti.sas)

Punteggi assegnati da 5 degustatori ai biscotti della ditta X. Il file contiene: nome prodotto anno di inizio produzione, tipo (cioccolato si/no), data del giudizio, punteggi.

```
*punti.sas;
*****;
***lettura di dati con input a formato***;
*****;

title 'Lettura di dati con input a formato';
data punti;
infile cards;
input nome $16. anno 3. +1 tipo $2. +1 data ddmmyy10.
(punt1-punt5)(4.1);
*calcola punteggio medio per biscotto;
media=sum(punt1,punt2,punt3,punt4,punt5)/5;
cards;
frollini miei      93 si 28-10-1999 7.8 6.5 7.2 8.0 7.9
buon mattino      90 no 30-10-1999 6.5 5.9 6.8 6.0 8.1
belli freschi     92 no 29-10-1999 8.9 7.9 8.5 9.0 8.8
dolci momenti     88 si 30-10-1999 6.7 5.6 4.9 5.2 6.1
;
*stampa i dati letti per controllo;
title2 'NB: le date sono stampate come n. di giorni dal
1/1/1960.';
proc print;
RUN;
```

Lettura di dati con input a formato  
NB: le date sono stampate come n. di giorni dal 1/1/1960.

| nome          | anno | tipo | data  |
|---------------|------|------|-------|
| frollini miei | 93   | si   | 14545 |
| buon mattino  | 90   | no   | 14547 |
| belli freschi | 92   | no   | 14546 |
| dolci momenti | 88   | si   | 14547 |

```
title2 'Usare istruzione Format per vedere le date
in altro formato';
proc print data=punti(drop=punt1-punt5 media);
id nome;
format data date8.;
run;
```

Lettura di dati con input a formato  
Usare istruzione Format per vedere le date in altro  
formato

| nome          | anno | tipo | data    |
|---------------|------|------|---------|
| frollini miei | 93   | si   | 28OCT99 |
| buon mattino  | 90   | no   | 30OCT99 |
| belli freschi | 92   | no   | 29OCT99 |
| dolci momenti | 88   | si   | 30OCT99 |

## INPUT MISTO

Posizione del **puntatore** nell'*input buffer*:

- *input a lista*: puntatore sulla 2.a colonna dopo il valore letto;
- *input a formato/colonna*: puntatore sulla 1.a colonna dopo il valore letto

```
INPUT A 4-5 B;
```

A → letta da colonne 4-5

B → letta da colonna 6, fino al primo blank

## Controllo puntatore

```
Input @4 A / @5 (B1-B3)(6.2);
```

@4 → A letta da colonna 4 fino al blank;

/ → caricamento del record successivo;

@5 → lettura B1,B2 e B3 da colonna 5 con formato 6.2.

## CONTROLLO DEL PUNTATORE

```
*input2.sas;
*controllo puntatore;
/*****;

----+----1----+----2----+
giuseppe      24  30

*/;

title 'Input2.sas: controllo del puntatore';
data pippo;
infile datalines;
input nome $10. +5 eta 2. +2 voto 4.;
datalines;
giuseppe      24  30
maria         22  26
laura         23  18
franco        21  22
;
proc print;
RUN;
```

Input2.sas: controllo del puntatore

| Obs | nome     | eta | voto |
|-----|----------|-----|------|
| 1   | giuseppe | 24  | 30   |
| 2   | maria    | 22  | 26   |
| 3   | laura    | 23  | 18   |
| 4   | franco   | 21  | 22   |

## LISTA DI FORMATI

```
*input3.sas;
*uso lista formati;
title 'input3.sas: uso lista formati';
data pippo;
input (nome test1-test5)($10. 5*4.);
cards;
tizio      121 114 137 156 142
caio       111  97 122 143 127
;
proc print; run;

data pippo;
input nome $10. (test1-test5)(4.);
cards;
tizio      121 114 137 156 142
caio       111  97 122 143 127
;
proc print; run;
```

input3.sas: uso lista formati

| Obs | nome  | test1 | test2 | test3 | test4 | test5 |
|-----|-------|-------|-------|-------|-------|-------|
| 1   | tizio | 121   | 114   | 137   | 156   | 142   |
| 2   | caio  | 111   | 97    | 122   | 143   | 127   |

## INPUT MISTO COLONNA/FORMATO

```
*input4.sas;
*input misto colonna/formato;
/*
-----1-----2-----+
giuseppe      24  30
*/
title 'input4.sas: input misto colonna/formato';
data pippo;
infile datalines;
input nome $ 1-10 @16 eta 2. @20 voto 4.;
datalines;
giuseppe      24  30
maria          22  26
laura         23  18
franco        21  22
;
run;
proc print;
run;
```

input4.sas: input misto colonna/formato

| Obs | nome     | eta | voto |
|-----|----------|-----|------|
| 1   | giuseppe | 24  | 30   |
| 2   | maria    | 22  | 26   |
| 3   | laura    | 23  | 18   |
| 4   | franco   | 21  | 22   |



## STRUMENTI AVANZATI PER LETTURA DATI

| <b>Tipo di lettura</b>                               | <b>Strumenti SAS</b>                                         |
|------------------------------------------------------|--------------------------------------------------------------|
| Più record per obs                                   | #n o / nella INPUT                                           |
| Un solo record per più obs                           | @ @ nella INPUT<br>@ con +INPUT e OUTPUT                     |
| Record a lunghezza variabile                         | Opzione LENGHT= nella INPUT                                  |
| Record diversi a seconda dell'obs letta              | IF-THEN con più INPUT e @                                    |
| Lettura a lista con separatore diverso da blank      | Opzione DELIMITER= nella INFILE                              |
| Controllo lunghezza record                           | Opzione LINESIZE= nella INFILE                               |
| Specificazione prima e ultima obs da leggere         | Opzioni FIRSTOBS= e OBS= nel nome del SAS Data set           |
| Specificazione primo e ultimo record da leggere      | Opzioni FIRSTOBS= e OBS= nella INFILE                        |
| Controllo INPUT in caso di missing o errori nei dati | Opzioni FLOWOVER, MISCOVER, STOPOVER, TRUNCOVER nella INFILE |

## ISTRUZIONI DI LETTURA DALLO STESSO RECORD

@ al termine dell'istruzione input non viene caricato un nuovo record nell'Input Buffer.

|                                                                                                                |                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| <pre>Data uno; input eta peso; input var2; cards; 18 66 24 23 19 71 26 21 20 68 23 29 22 70 25 24 ; run;</pre> | <pre>Data uno; input eta peso @; input var2; cards; 18 66 24 23 19 71 26 21 20 68 23 29 22 70 25 24 ; run;</pre> |
|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|

| ETA | PESO | VAR2 | ETA | PESO | VAR2 |
|-----|------|------|-----|------|------|
| 18  | 66   | 19   | 18  | 66   | 24   |
| 20  | 68   | 22   | 19  | 71   | 26   |
|     |      |      | 20  | 68   | 23   |
|     |      |      | 22  | 70   | 25   |

## LETTURA DI PIÙ OSSERVAZIONI DA UNO STESSO RECORD

@@ trattiene il record nell'Input Buffer per successive esecuzioni del Data step, fino a fine record fisico.

| Data uno;               | SESSO | PESO |
|-------------------------|-------|------|
| input sesso \$ peso @@; | M     | 70   |
| cards;                  | F     | 65   |
| M 70 F 65 M 74 M 60     | M     | 74   |
| M 82 F 59 M 77 F 66     | M     | 60   |
| ;                       | M     | 82   |
| run;                    | F     | 59   |
|                         | M     | 77   |
|                         | F     | 66   |

## LETTURA DI UNA OSSERVAZIONE SU PIÙ RECORDS

Un file contiene i dati relativi alle temperature per il mese di luglio in 3 Paesi. L'osservazione relativa ad ogni Paese è registrata su 3 record:

- 1° record contiene la città e lo stato;
- 2° la temperatura media min e max;
- 3° la temperatura "record" min e max.

```
*piurec.sas;
*****;
*** \ va a record successivo, #n va a record n ***;
*** NB: osserva con attenzione la finestra LOG ***;
*****;
title 'Letture di una osservazione su più records';
data tempera;
infile cards;
input citta $ stato $ / maxnorm minnorm #3 maxrec
minrec;
*trasforma temperature da F° a C°;
array f(*) maxnorm minnorm maxrec minrec;
do i=1 to dim(f);
    f(i)=(5/9)*(f(i)-32);
end;
cards;
Nome AK
55 44
88 29
Miami FL
90 75
97 65
Raleigh NC
88 68
105 50
;
```

```
proc print data=tempera;  
run;
```

Lettura di una osservazione su più records

| Obs | citta   | stato | maxnorm | minnorm | maxrec  | minrec  |
|-----|---------|-------|---------|---------|---------|---------|
| 1   | Nome    | AK    | 12.7778 | 6.6667  | 31.1111 | -1.6667 |
| 2   | Miami   | FL    | 32.2222 | 23.8889 | 36.1111 | 18.3333 |
| 3   | Raleigh | Nc    | 31.1111 | 20.0000 | 40.5556 | 10.0000 |

log

NOTE: The data set WORK.TEMPERA has 3 observations and 7 variables.

NOTE: DATA statement used:

real time 0.16 seconds

## DATI NEI RECORD E VARIABILI IN INPUT

- (a) le variabili specificate esauriscono completamente i dati presenti nel record;
  - (b) le variabili specificate richiedono un numero di dati minore rispetto a quelli esistenti;
  - (c) le variabili specificate richiedono un numero di dati superiore rispetto a quello esistente.
- (c) → SAS va a capo a leggere nel record successivo, partendo dalla prima colonna e dà il seguente messaggio:

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.

## Esempio

```
Data uno;  
input var1-var6;  
cards;  
18 66 24 23  
19 71 26 21  
20 68 23 29  
21 65 28 24  
;  
run;  
proc print;
```

| VAR1 | VAR2 | VAR3 | VAR4 | VAR5 | VAR6 |
|------|------|------|------|------|------|
| 18   | 66   | 24   | 23   | 19   | 71   |
| 20   | 68   | 23   | 29   | 21   | 65   |

se cancelliamo ultimo record dei dati:

```
NOTE: LOST CARD.  
RULE:-----1-----2-----3-----4-----  
+----5-----6-----7-----8-----+---  
258 ;  
VAR1=20 VAR2=68 VAR3=23 VAR4=29 VAR5=. VAR6=.  
_ERROR_=1 _N_=2  
NOTE: SAS went to a new line when INPUT statement  
reached past the end of a line.  
NOTE: The data set WORK.UNO has 1 observations and  
6 variables.
```

## CONTROLLO ESAURIMENTO DATI

**INFILE** *nome\_file* **STOPOVER**;

tratta la situazione di esaurimento dei dati come errore e arresta esecuzione del Data step.

```
Data due;  
input var1-var6;  
infile cards stopover;  
cards;  
18 66 24 23  
19 71 26 21  
20 68 23 29  
;  
run;
```

```
ERROR: INPUT statement exceeded record length.  
INFILE CARDS OPTION STOPOVER specified.  
RULE:-----1-----2-----3-----4-----  
+----5-----6-----7-----8-----+---  
277 18 66 24 23  
VAR1=18 VAR2=66 VAR3=24 VAR4=23 VAR5=. VAR6=.  
_ERROR_=1 _N_=1  
NOTE: SAS System ha interrotto l'elaborazione del  
passo a causa errori.
```

## CONTROLLO ESAURIMENTO DATI

### INFILE *nome\_file* MISCOVER;

Fa sì che la lettura non vada oltre la fine del record → assegna valore mancante a quelle variabili per cui non vi sono dati.

```
Data tre;  
input var1-var6;  
infile cards miscover;  
cards;  
18 66 24 23  
19 71 26 21  
20 68 23 29  
;  
run;  
proc print data=tre;  
run;
```

NOTE: The data set WORK.TRE has 3 observations and 6 variables.

NOTE: The DATA statement used 0.0 seconds.

| OBS | VAR1 | VAR2 | VAR3 | VAR4 | VAR5 | VAR6 |
|-----|------|------|------|------|------|------|
| 1   | 18   | 66   | 24   | 23   | .    | .    |
| 2   | 19   | 71   | 26   | 21   | .    | .    |
| 3   | 20   | 68   | 23   | 29   | .    | .    |

## CONTROLLO ESAURIMENTO DATI

### INFILE *nome\_file* TRUNCOVER;

Fa sì che la lettura non vada oltre la fine del record → assegna valore mancante a quelle variabili per cui non vi sono dati. TRUNCOVER consente di leggere records a lunghezza variabile quando alcuni records sono più corti di quanto dichiarato nell'istruzione INPUT.

```
Data tre;  
input var1-var6;  
infile cards truncover;  
cards;  
18 66 24 23  
19 71 26 21  
20 68 23 29  
;  
run;  
proc print data=tre;  
run;
```

NOTE: The data set WORK.TRE has 3 observations and 6 variables.

NOTE: The DATA statement used 0.0 seconds.

| OBS | VAR1 | VAR2 | VAR3 | VAR4 | VAR5 | VAR6 |
|-----|------|------|------|------|------|------|
| 1   | 18   | 66   | 24   | 23   | .    | .    |
| 2   | 19   | 71   | 26   | 21   | .    | .    |
| 3   | 20   | 68   | 23   | 29   | .    | .    |

## OPZIONI DEL NOME DEL SAS DATA SET

**DATA** *nome*(*opzione1* *opzione2* ...);

**PROC** *proc* data=*nome*(*opz1* *opz2* ...);

*Opzioni (selezione)*

- (a) Selezione delle variabili: **KEEP**, **DROP**;
- (b) Selezione delle osservazioni rispetto al loro numero d'ordine: **OBS**, **FIRSTOBS**;
- (c) Selezione delle osservazioni rispetto alle modalità delle variabili: **WHERE**;
- (d) Assegnazione di un nuovo nome alle variabili: **RENAME**;
- (e) Assegnazione di un nome esteso al data set: **LABEL**.

## ESEMPIO DI LETTURA FILE ASCII

V:\CARLA\SAS\REGIONI.PRN

|     |                       |          |         |         |
|-----|-----------------------|----------|---------|---------|
| 1   | PIEMONTE              | 25398.94 | 4377229 | 2118931 |
| 1   | VALLE D'AOSTA         | 3262.26  | 114325  | 56630   |
| 1   | LOMBARDIA             | 23858.55 | 8886402 | 4287700 |
| 2   | TRENTINO-ALTO ADIGE   | 13618.31 | 881986  | 430461  |
| 2   | VENETO                | 18364.56 | 4374911 | 2120827 |
| 2   | FRIULI-VENEZIA GIULIA | 7844.13  | 1210242 | 577195  |
| ... |                       |          |         |         |
| 4   | CALABRIA              | 15080.32 | 2146724 | 1060391 |
| 5   | SICILIA               | 25707.23 | 5141343 | 2512105 |
| 5   | SARDEGNA              | 24089.89 | 1651218 | 814856  |

20 osservazioni e 5 variabili: ripartizione geografica (codici da 1 a 5), regione (carattere, 21 colonne), superficie, popolazione residente totale e maschile al 1988.

- creare un archivio SAS temporaneo di nome REGIONI, assegnando alle variabili i nomi x1-x5;
- assegnare alle variabili una descrizione (LABEL);
- assegnare alla variabile DENSITA la densità di popolazione;
- creare una variabile che contenga il rapporto tra popolazione maschile e popolazione totale;
- eliminare la variabile relativa alla popolazione maschile;
- rinominare la variabile x4 pop88;
- stampare il contenuto dell'archivio e della DIA, assegnando all'output il titolo: Popolazione residente e superficie al 1988 per regione.

## OPZIONI DEL NOME DEL SAS DATA SET

### *(a) Selezione delle variabili: KEEP, DROP*

```
*dataset1.sas;  
title 'Selezione variabili';  
data uno (keep=regione superf) due (drop=maschi);  
infile 'v:\carla\sas\regioni.prn';  
input ripartiz regione $21. superf pop88 maschi;  
run;  
proc contents data=uno;  
run;  
proc contents data=due;  
run;  
proc print data=due(keep=regione pop88);  
run;
```

## OPZIONI DEL NOME DEL SAS DATA SET

### *(a) Selezione delle osservazioni rispetto al loro numero d'ordine: OBS, FIRSTOBS*

```
*dataset2.sas;  
title 'Selezione osservazioni per numero d'ordine';  
data regioni;  
infile 'v:\ carla\sas\regioni.prn';  
input ripartiz regione $21. superf pop88 maschi;  
run;  
*selezione nelle procedure;  
proc contents;  
proc print data=regioni(firstobs=5 obs=10 keep=regione  
superf);  
run;  
*selezione nel data step;  
data uno(keep=regione superf);  
infile'v:\carla\sas\regioni.prn' firstobs=5 obs=10;  
input ripartiz regione $21. superf pop88 maschi;  
run;  
proc contents;  
proc print;  
run;
```

FIRSTOBS= e OBS= operano solo su sas data set già esistenti. La selezione di osservazioni da file esterno deve essere fatta nell'istruzione INFILE.

## OPZIONI DEL NOME DEL SAS DATA SET

(b) *Selezione delle osservazioni rispetto alle modalità delle variabili: WHERE*

```
*dataset3.sas;
```

```
title 'Selezione osservazioni per modalita' delle  
variabili';  
*selezione nelle procedure;  
title2 'regioni centro';  
proc print data=regioni(where=(ripartiz=3));  
run;
```

```
*selezione nel data step;
```

```
data uno;  
infile 'v:\didat\carla\sas\regioni.prn';  
input ripartiz regione $21. superf pop88 maschi;  
if ripartiz=3;  
run;  
proc contents;  
proc print;  
run;
```

```
*selezione nel DATA step:opzione where;
```

```
data pippo;  
set regioni(where=(ripartiz=3));  
run;  
proc print data=pippo;  
run;
```

L'opzione WHERE può essere utilizzata per selezionare osservazioni da data set SAS già esistenti.

## OPZIONI DEL NOME DEL SAS DATA SET

(c) Assegnazione di un nuovo nome alle variabili: **RENAME**;  
(d) Assegnazione di un nome esteso al data set: **LABEL**.

```
*dataset4.sas;  
title 'rinomina le variabili';  
data regioni(rename=(b=regione) label='archivio  
regionale');  
infile 'v:\carla\sas\regioni.prn';  
input a b $21. superf pop88 maschi;  
c=b;  
label a=ripartizione;  
run;  
proc contents position;  
run;  
proc print data=regioni(obs=2);  
run;
```



## OPZIONI DEL NOME DEL SAS DATA SET: CONTROLLO DEL DATA SET

| Category         | Data Set Option | Description                                                                                                 |
|------------------|-----------------|-------------------------------------------------------------------------------------------------------------|
| Data Set Control | ALTER=          | Assigns an alter password to a SAS file and enables access to a password-protected SAS file                 |
|                  | BUFNO=          | Specifies the number of buffers for processing a SAS data set                                               |
|                  | BUFSIZE=        | Specifies a permanent buffer size for output SAS data sets                                                  |
|                  | CNTLLEV=        | Specifies the level of shared access to SAS data sets                                                       |
|                  | COMPRESS=       | Compresses observations in an output SAS data set                                                           |
|                  | DLDMGACTION=    | Specifies what type of action to take when a SAS data set in a SAS data library is detected as damaged      |
|                  | ENCRYPT=        | Encrypts SAS data files                                                                                     |
|                  | GENMAX=         | Requests generations for a data set and specifies the maximum number of versions                            |
|                  | GENNUM=         | References a specific generation of a data set                                                              |
|                  | INDEX=          | Defines indexes when a SAS data set is created                                                              |
|                  | LABEL=          | Specifies a label for the data set                                                                          |
|                  | OUTREP=         | Specifies an operating environment's requirements vector for an output file                                 |
|                  | PW=             | Assigns a read, write, or alter password to a SAS file and enables access to a password-protected SAS file  |
|                  | PWREQ=          | Controls the pop up of a requester window for a data set password                                           |
|                  | READ=           | Assigns a read password to a SAS file and enables access to a read-protected SAS file                       |
|                  | REPEMPTY=       | Controls replacement of like-named temporary or permanent SAS data sets when the new one is empty.          |
|                  | REPLACE=        | Controls replacement of like-named temporary or permanent SAS data sets                                     |
|                  | REUSE=          | Specifies whether new observations are written to free space in compressed SAS data sets                    |
|                  | SORTEDBY=       | Specifies how the data set is currently sorted                                                              |
|                  | TOBSNO=         | Specifies the number of observations to be transmitted in each multi-observation exchange with a SAS server |
|                  | TRANTAB=        | Specifies a translation table for character conversions                                                     |
|                  | TYPE=           | Specifies the data set type for a specially structured SAS data set                                         |
|                  | WRITE=          | Assigns a write password to a SAS file and enables access to a write-protected SAS file                     |

## OPZIONI DEL NOME DEL SAS DATA SET: ALTRE OPZIONI

| Category                        | Data Set Option | Description                                                                                                                           |
|---------------------------------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------|
| Miscellaneous                   | FILECLOSE=      | Specifies how a tape is positioned when a SAS file on the tape is closed                                                              |
| Observation Control             | FIRSTOBS=       | Causes processing to begin at a specified observation                                                                                 |
|                                 | IN=             | Creates a variable that indicates whether the data set contributed data to the current observation                                    |
|                                 | OBS=            | Causes processing to end with the <i>n</i> th observation                                                                             |
|                                 | POINTOBS=       | Controls whether a compressed data set may be processed with random access (by observation number) rather than sequential access only |
|                                 | WHERE=          | Selects observations that meet the specified condition                                                                                |
|                                 | WHEREUP=        | Specifies whether to evaluate added observations and modified observations against a WHERE expression                                 |
| User Control of SAS Index Usage | IDXNAME=        | Directs SAS to use a specific index to satisfy the conditions of a WHERE expression                                                   |
|                                 | IDXWHERE=       | Overrides the SAS System decision about whether to use an index to satisfy the conditions of a WHERE expression                       |
| Variable Control                | DROP=           | Excludes variables from processing or from output SAS data sets                                                                       |
|                                 | KEEP=           | Specifies variables for processing or for writing to output SAS data sets                                                             |
|                                 | RENAME=         | Changes the name of a variable                                                                                                        |

Quasi tutte le opzioni del nome del SAS data set possono essere utilizzate sia in input che in output



## ESEMPIO DI LETTURA E CONTROLLO DATI

```
*valuta1.sas;
*****;
*** input dati questionario valutazione ***;
***      solo le prime 8 domande      ***;
***      lettura dati da file ASCII    ***;
*****;

*opzioni di sistema per stampa output: formato pagina e info aggiuntive;
*(guardare finestra OPTIONS);
options nocenter nodate nonumber;

*titolo output (guardare finestra TITLE);
title1 'Scheda valutazione didattica SCPOL 97/98';

*definizione libref;
libname dati 'd:\';

data dati.valuta(label='dati valutazione SCPOL 97/98');
infile 'v:\didat\carla\sas\valu97.txt';
*input a colonna;
input id 1-8 corso 9-12 docente 13-16 dom1 17-18 dom2 19-22 dom3 23-26
      dom4 27-28 dom5 29-30 dom6 31-33 dom7 34 dom8 35-36;
*nome esteso alle variabili;
label dom1='Sesso' dom2='Anno di nascita' dom3='Prima immatricolazione'
      dom4='Scuola di provenienza' dom5='Voto riportato al Diploma'
      dom6='Anno di corso' dom7='Posizione' dom8='Corsi frequentati
nell''anno';
run;

*stampa contenuto della DIA;
proc contents position;
run;

title2 'controllo dei dati';
*valori mancanti, minimo e massimo;
proc means n nmiss min max maxdec=0;
run;

*distribuzione semplice di frequenza;
proc freq data=dati.valuta(drop=id);
tables corso docente dom1-dom8;
run;
```

## CREAZIONE DI UN SAS DATA SET

### *Da file ASCII (formato testo)*

- **DATA:** segna inizio passo di DATA, dà un nome al SAS data set creato;
- **INFILE:** specifica il file da leggere e le sue caratteristiche;
- **INPUT:** descrive le variabili da leggere (nome, posizione nel record, tipo);
- **CARDS o DATALINES:** inizio dei dati da leggere da programma.

### *Da SAS data set*

- **SET, MERGE, UPDATE:** identificano il/i SAS data set da cui leggere;
- **BY:** specifica variabili da usare con le istruzioni precedenti.

### *Da programma*

DO loops, assegnazioni

## ISTRUZIONI SAS

- Istruzioni del DATA step;
- Istruzioni specifiche delle procedure;
- Istruzioni utilizzabili ovunque.

### Istruzioni del DATA step

- *Eseguibili*: corrispondono ad azioni durante l'esecuzione;
- *Dichiarative*: danno informazione al sistema, hanno effetto in fase di compilazione.

### *Funzione delle istruzioni*

- Istruzioni di *azione* (p.e: creazione e modifica di variabili, selezione delle osservazioni);
- Istruzioni di *controllo*;
- Istruzioni manipolazione *files esterni*;
- Istruzioni di *informazione*.

## ISTRUZIONI GLOBALI

Possono essere utilizzate in un punto qualsiasi del programma e non sono eseguibili, ma vengono utilizzati in fase di compilazione.

- Istruzioni di *identificazione* dei dati: associano nomi di riferimento o valori a librerie, files o dati in input;
- Istruzioni di controllo del LOG;
- Istruzioni di “scambio” con il sistema operativo;
- Istruzioni di controllo dell'OUTPUT;
- Istruzioni di controllo della programmazione.

Table 6.6 Categories and Descriptions of Global Statements

| Category              | Statement                       | Description                                                                                                                                                                                                                        |
|-----------------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Access           | CATNAME                         | Logically combines two or more catalogs into one by associating them with a catref (a shortcut name); clears one or all catrefs; lists the concatenated catalogs in one concatenation or in all concatenations                     |
|                       | FILENAME                        | Associates a SAS fileref with an external file or an output device; disassociates a fileref and external file; lists attributes of external files                                                                                  |
|                       | FILENAME, CATALOG Access Method | References a SAS catalog as an external file                                                                                                                                                                                       |
|                       | FILENAME, FTP Access Method     | Allows you to access remote files using the FTP protocol                                                                                                                                                                           |
|                       | FILENAME, SOCKET Access Method  | Allows you to read from or write to a TCP/IP socket                                                                                                                                                                                |
|                       | FILENAME, URL Access Method     | Allows you to access remote files using the URL access method                                                                                                                                                                      |
|                       | LIBNAME                         | Associates or disassociates a SAS data library with a libref (a shortcut name); clears one or all librefs; lists the characteristics of a SAS data library; concatenates SAS data libraries; implicitly concatenates SAS catalogs. |
|                       | LIBNAME, SAS/ACCESS             | Associates a SAS libref with a database management system (DBMS) database, schema, server, or group of tables or views                                                                                                             |
| Log Control           | Comment                         | Documents the purpose of the statement or program                                                                                                                                                                                  |
|                       | PAGE                            | Skips to a new page in the SAS log                                                                                                                                                                                                 |
|                       | SKIP                            | Creates a blank line in the SAS log                                                                                                                                                                                                |
| Operating Environment | X                               | Issues an operating-environment command from within a SAS session                                                                                                                                                                  |
| Output Control        | FOOTNOTE                        | Prints up to ten lines of text at the bottom of the procedure or DATA step output                                                                                                                                                  |
|                       | ODS EXCLUDE                     | Specifies output objects to exclude from ODS destinations                                                                                                                                                                          |
|                       | ODS HTML                        | Opens, manages, or closes the HTML destination. If the destination is open, you can create HTML output (output that is written in Hypertext Markup Language).                                                                      |
|                       | ODS LISTING                     | Opens, manages, or closes the Listing destination                                                                                                                                                                                  |
|                       | ODS OUTPUT                      | Creates a SAS data set from an output object and manages the selection and exclusion lists for the Output destination                                                                                                              |
|                       | ODS PATH                        | Specifies which locations to search for definitions that were created by PROC EMPLATE, as well as the order in which to search for them                                                                                            |
|                       | ODS PRINTER                     | Opens, manages, or closes the Printer destination. If the destination is open, you can create Printer output (output that is formatted for a high-resolution printer)                                                              |
|                       | ODS SELECT                      | Specifies output objects for ODS destinations                                                                                                                                                                                      |
|                       | ODS SHOW                        | Writes to the SAS log the specified selection or exclusion list                                                                                                                                                                    |
|                       | ODS TRACE                       | Writes to the SAS log a record of each output object that is created, or suppresses the writing of this record                                                                                                                     |
|                       | ODS VERIFY                      | Prints or suppresses a warning that a style definition or a table definition that is used is not supplied by SAS Institute                                                                                                         |
|                       | Program Control                 | TITLE                                                                                                                                                                                                                              |
| DM                    |                                 | Submits SAS Program Editor, Log, Procedure Output or text editor commands as SAS statements                                                                                                                                        |
| ENDSAS                |                                 | Terminates a SAS job or session after the current DATA or PROC step executes                                                                                                                                                       |
| %INCLUDE              |                                 | Brings a SAS programming statement, data lines, or both, into a current SAS program                                                                                                                                                |
| %LIST                 |                                 | Displays lines that are entered in the current session                                                                                                                                                                             |
| OPTIONS               |                                 | Changes the value of one or more SAS system options                                                                                                                                                                                |
| RUN                   |                                 | Executes the previously entered SAS statements                                                                                                                                                                                     |
| Window Display        | DISPLAY                         | Displays a window that is created with the WINDOW statement                                                                                                                                                                        |
|                       | WINDOW                          | Creates customized windows for your applications                                                                                                                                                                                   |

## ISTRUZIONI DEL DATA STEP PER CATEGORIA

### Categoria istr. AZIONE

Consentono di ...  
*creare e modificare variabili* (sum, assegnazione,...);  
*Selezionare* le osservazioni (DELETE, IF, OUTPUT, REMOVE, WHERE,...);  
 Cercare *errori* nella fase di input (ERROR, LOSTCARD).

### CONTROLLO

modificare l'ordine di esecuzione delle istruzioni (LEAVE);  
 trasferire il controllo da una parte del programma a un'altra (GOTO);  
 saltare istruzioni per certe osservazioni (IF-THEN-ELSE).

### GESTIONE FILE

*lavorare* con file in input (CARDS, BY, INFILE, INPUT, MERGE, SET...);  
*gestire* file in output (BY, DATA, FILE, PUT, ... ).

INFORMAZIONE *informazioni* aggiuntive al sistema sul PDV (ARRAY, RETAIN...) o sul data set che si sta creando (FORMAT, LABEL, RENAME...).

## ISTRUZIONI ESEGUIBILI DATA STEP

### Executable Statements

|               |                  |                |
|---------------|------------------|----------------|
| ABORT         | IF, Subsetting   | PUT            |
| Assignment    | IF-THEN/ELSE     | PUT, Column    |
| CALL          | INFILE           | PUT, Formatted |
| CONTINUE      | INPUT            | PUT, List      |
| DELETE        | INPUT, Column    | PUT, Named     |
| DESCRIBE      | INPUT, Formatted | PUT, _ODS_     |
| DISPLAY       | INPUT, List      | REDIRECT       |
| DO            | INPUT, Named     | REMOVE         |
| DO, Iterative | LEAVE            | REPLACE        |
| DO Until      | LINK             | RETURN         |
| DO While      | LIST             | SELECT         |
| ERROR         | LOSTCARD         | SET            |
| EXECUTE       | MERGE            | STOP           |
| FILE          | MODIFY           | Sum            |
| FILE, ODS     | Null             | UPDATE         |
| GO TO         | OUTPUT           |                |

## ISTRUZIONI DI DICHIARAZIONE DEL DATA STEP

### Declarative Statements

|                 |            |                   |
|-----------------|------------|-------------------|
| ARRAY           | DATALINES  | LABEL             |
| Array Reference | DATALINES4 | Labels, Statement |
| ATTRIB          | DROP       | LENGTH            |
| BY              | END        | RENAME            |
| CARDS           | FORMAT     | RETAIN            |
| CARDS4          | INFORMAT   | WHERE             |
| DATA            | KEEP       | WINDOW            |

## ISTRUZIONI KEEP, DROP, RENAME E LABEL

Consentono di:

- (a) assegnare nomi estesi alle variabili;
- (b) modificare il nome delle variabili;
- (c) escludere delle variabili dal Data set che si sta creando.

*(a) Assegnazione di nomi estesi (LABEL) alle variabili*

**LABEL** *variabile*='nome esteso' ...;

- *nome esteso* non può superare i 40 caratteri;
- LABEL è una istruzione globale;
- se usata nel DATA step → label registrata nella DIA.

## ISTRUZIONE RENAME

Consente di *modificare* il *nome* delle variabili.

**RENAME** *vecchio\_nome=nuovo\_nome ...;*

```
rename var1=eta var2=titstud;
```

Modifica il nome della variabile nel PDV, subito prima della scrittura nel SAS data set → nel Data step la variabile è disponibile con il vecchio nome

```
*rename sas;
data prova;
input x @@;
y=log(x);
rename x=num y=lognum;
z=exp(y);
cards;
1 2 3 4 5 6 7 8 9 10
;
run;
proc print; run;
```

|  | OBS | NUM | LOGNUM  | Z  |
|--|-----|-----|---------|----|
|  | 1   | 1   | 0.00000 | 1  |
|  | 2   | 2   | 0.69315 | 2  |
|  | 3   | 3   | 1.09861 | 3  |
|  | 4   | 4   | 1.38629 | 4  |
|  | 5   | 5   | 1.60944 | 5  |
|  | 6   | 6   | 1.79176 | 6  |
|  | 7   | 7   | 1.94591 | 7  |
|  | 8   | 8   | 2.07944 | 8  |
|  | 9   | 9   | 2.19722 | 9  |
|  | 10  | 10  | 2.30259 | 10 |

## SELEZIONE DELLE VARIABILI

**DROP** *variabili;*

**KEEP** *variabili;*

- l'istruzione **DROP** elenca le variabili da eliminare nel SAS Data set;
- l'istruzione **KEEP** elenca le variabili da scrivere nel SAS Data set.

## LE OPZIONI DEL SISTEMA

Istruzioni che controllano la sessione di lavoro SAS e rimangono attive per tutta la sessione, salvo che non vengano di nuovo specificate. Possono essere controllate:

(a) mediante l'istruzione  
**OPTIONS** *opzioni*;

(b) mediante la finestra **OPTIONS** del DMS.

*Opzioni (selezione)*

(a) controllo delle dimensioni fisiche della pagina di stampa;

PAGESIZE=*n* default=93

LINESIZE=*n* default=58

## LE OPZIONI DEL SISTEMA

(b) controllo delle informazioni aggiuntive stampate

DATE|NODATE

NUMBER|NONUMBER

(c) controllo della forma dell'output

CENTER|NOCENTER

MISSING=*'carattere'* default .

(d) controllo delle osservazioni elaborate

FIRSTOBS=*n* default 1

OBS=*m* default pari a ultima obs

(e) altre

DEVICE=*nome* per i grafici

ERRORS=*n* default 20

REPLACE|NOREPLACE



## ISTRUZIONI GLOBALI PER CATEGORIA (1)

**Table 8.6** Categories and Descriptions of Global Statements

| Category    | Statement                       | Description                                                                                                                                                                                                                        |
|-------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Access | CATNAME                         | Logically combines two or more catalogs into one by associating them with a catref (a shortcut name); clears one or all catrefs; lists the concatenated catalogs in one concatenation or in all concatenations                     |
|             | FILENAME                        | Associates a SAS fileref with an external file or an output device; disassociates a fileref and external file; lists attributes of external files                                                                                  |
|             | FILENAME, CATALOG Access Method | References a SAS catalog as an external file                                                                                                                                                                                       |
|             | FILENAME, FTP Access Method     | Allows you to access remote files using the FTP protocol                                                                                                                                                                           |
|             | FILENAME, SOCKET Access Method  | Allows you to read from or write to a TCP/IP socket                                                                                                                                                                                |
|             | FILENAME, URL Access Method     | Allows you to access remote files using the URL access method                                                                                                                                                                      |
|             | LIBNAME                         | Associates or disassociates a SAS data library with a libref (a shortcut name); clears one or all librefs; lists the characteristics of a SAS data library; concatenates SAS data libraries; implicitly concatenates SAS catalogs. |
|             | LIBNAME, SAS/ACCESS             | Associates a SAS libref with a database management system (DBMS) database, schema, server, or group of tables or views                                                                                                             |
| Log Control | Comment                         | Documents the purpose of the statement or program                                                                                                                                                                                  |
|             | PAGE                            | Skips to a new page in the SAS log                                                                                                                                                                                                 |
|             | SKIP                            | Creates a blank line in the SAS log                                                                                                                                                                                                |

## ISTRUZIONI GLOBALI PER CATEGORIA (2)

|                       |                |                                                                                                                                                                       |
|-----------------------|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Operating Environment | X              | Issues an operating-environment command from within a SAS session                                                                                                     |
| Output Control        | FOOTNOTE       | Prints up to ten lines of text at the bottom of the procedure or DATA step output                                                                                     |
|                       | ODS EXCLUDE    | Specifies output objects to exclude from ODS destinations                                                                                                             |
|                       | ODS HTML       | Opens, manages, or closes the HTML destination. If the destination is open, you can create HTML output (output that is written in Hypertext Markup Language).         |
|                       | ODS LISTING    | Opens, manages, or closes the Listing destination                                                                                                                     |
|                       | ODS OUTPUT     | Creates a SAS data set from an output object and manages the selection and exclusion lists for the Output destination                                                 |
|                       | ODS PATH       | Specifies which locations to search for definitions that were created by PROC EMPLATE, as well as the order in which to search for them                               |
|                       | ODS PRINTER    | Opens, manages, or closes the Printer destination. If the destination is open, you can create Printer output (output that is formatted for a high-resolution printer) |
|                       | ODS SELECT     | Specifies output objects for ODS destinations                                                                                                                         |
|                       | ODS SHOW       | Writes to the SAS log the specified selection or exclusion list                                                                                                       |
|                       | ODS TRACE      | Writes to the SAS log a record of each output object that is created, or suppresses the writing of this record                                                        |
| Program Control       | ODS VERIFY     | Prints or suppresses a warning that a style definition or a table definition that is used is not supplied by SAS Institute                                            |
|                       | TITLE          | Specifies title lines for SAS output                                                                                                                                  |
|                       | DM             | Submits SAS Program Editor, Log, Procedure Output or text editor commands as SAS statements                                                                           |
|                       | ENDSAS         | Terminates a SAS job or session after the current DATA or PROC step executes                                                                                          |
|                       | %INCLUDE       | Brings a SAS programming statement, data lines, or both, into a current SAS program                                                                                   |
|                       | %LIST          | Displays lines that are entered in the current session                                                                                                                |
|                       | OPTIONS        | Changes the value of one or more SAS system options                                                                                                                   |
|                       | RUN            | Executes the previously entered SAS statements                                                                                                                        |
|                       | %RUN           | Ends source statements following a %INCLUDE * statement                                                                                                               |
|                       | Window Display | DISPLAY                                                                                                                                                               |
| WINDOW                |                | Creates customized windows for your applications                                                                                                                      |

## ISTRUZIONI DI ASSEGNAZIONE

*variabile=espressione;*

L'*espressione* è una sequenza di:

- *operandi* (variabili, costanti);
- *operatori* (caratteri speciali, funzioni, parentesi).

*Tipo di operatori*

- aritmetici;
- di comparazione;
- logici;
- carattere.

*Regole ordine di esecuzione*

1. espressioni entro parentesi;
2. 7 livelli di priorità (1=max)
3. se stessa priorità: da dx a sin livello1  
da sin a dx gli altri

## OPERATORI SAS

*Osservare:*

- precedenza tra operatori (*priority*);
- ordine di esecuzione a parità di operatori (*order of evaluation*);
- simbolo (*symbol, mnemonic equivalent*)
- categorie di operatori.

## PRIORITÀ E ORDINE DI ESECUZIONE DEGLI OPERATORI

| Priority | Order of Evaluation | Symbols | Mnemonic Equivalent | Definition                   | Example              |
|----------|---------------------|---------|---------------------|------------------------------|----------------------|
| Group I  | right to left       | **      |                     | exponentiation <sup>1</sup>  | y=a**2;              |
|          |                     | +       |                     | positive prefix <sup>2</sup> | y=+(a*b);            |
|          |                     | -       |                     | negative prefix <sup>3</sup> | z=-(a*b);            |
|          |                     | ~       | NOT                 | logical not <sup>4</sup>     | if not z then put x; |
|          |                     | <       | MIN                 | minimum <sup>5</sup>         | x=(a<b);             |
|          |                     | >       | MAX                 | maximum                      | x=(a>b);             |

| Priority             | Order of Evaluation        | Symbols | Mnemonic Equivalent | Definition                                | Example                                          |
|----------------------|----------------------------|---------|---------------------|-------------------------------------------|--------------------------------------------------|
| Group II             | left to right              | *       |                     | multiplication                            | c=a*b;                                           |
|                      |                            | /       |                     | division                                  | f=g/h;                                           |
| Group III            | left to right              | +       |                     | addition                                  | c=a+b;                                           |
|                      |                            | -       |                     | subtraction                               | f=g-h;                                           |
| Group IV             | left to right              | !!      |                     | concatenate character values <sup>6</sup> | name='J'    'SMITH';                             |
| Group V <sup>7</sup> | left to right <sup>8</sup> | <       | LT                  | less than                                 | if x<y then c=5;                                 |
|                      |                            | <=      | LE                  | less than or equal to                     | if x le y then a=0;                              |
|                      |                            | =       | EQ                  | equal to                                  | if y eq (x+a) then output;                       |
|                      |                            | ≠       | NE                  | not equal to                              | if x ne z then output;                           |
|                      |                            | >=      | GE                  | greater than or equal to                  | if y>=a then output;                             |
|                      |                            | >       | GT                  | greater than                              | if z>a then output;                              |
|                      |                            |         | IN                  | equal to one of a list                    | if state in ('NY', 'NJ', 'PA') then region='NE'; |
| Group VI             | left to right              | &       | AND                 | logical and                               | if a=b & c=d then x=1;                           |
| Group VII            | left to right              | !       | OR                  | logical or <sup>9</sup>                   | if y=2 or x=3 then a=d;                          |

## OPERATORI ARITMETICI

| Priorità | Simbolo | Descrizione          |
|----------|---------|----------------------|
| 1        | **      | Elevamento a potenza |
| 2        | *       | Moltiplicazione      |
| 2        | /       | Divisione            |
| 3        | +       | Addizione            |
| 3        | -       | Sottrazione          |

- agiscono su variabili di tipo numerico;
- conversione automatica carattere/num, conversione non valida → *missing*;
- operatori con uguale priorità vengono eseguiti da sin a dx; elevamento a potenza da dx a sin;
- l'uso delle parentesi può modificare la priorità degli operatori;
- tutte le operazioni aritmetiche vengono eseguite in doppia precisione (memorizzazione in 8 byte).

## OPERATORI DI COMPARAZIONE

|    |    |                                      |
|----|----|--------------------------------------|
| <  | LT | minore di                            |
| <= | LE | minore o uguale a                    |
| =  | EQ | uguale a                             |
| >  | GT | maggiore di                          |
| >= | GE | maggiore o uguale a                  |
| ^= | NE | non uguale                           |
|    | IN | uguale a uno degli elem. della lista |

- Effettuano un confronto tra due operandi.  
Tale confronto genera un valore numerico:
  - 1 confronto *vero*
  - 0 confronto *falso*
- Hanno tutti lo stesso livello gerarchico;
- Possono operare sia su variabili/costanti numeriche che carattere;
- carattere trasformata in numerica se confronto tra numerica e carattere;
- variabili carattere confrontate elemento per elemento;
- Il valore missing è considerato sempre il valore più piccolo.

## Esempio

VAR1 variabile carattere (sì/no)

RISP variabile dicotomica (0/1) da creare per  
VAR1=sì

```
RISP=var1='sì' ;
```

## ESEMPIO CONVERSIONE

```
1 data dati;
2 a='a';
3 x=1;
4 b=a gt x;
5 c=x gt a;
6 run;
```

NOTE: Character values have been converted to numeric values at the places given by:

```
(Line):(Column).
4:3 5:8
```

NOTE: Invalid numeric data, a='a' , at line 4 column 3.

NOTE: Invalid numeric data, a='a' , at line 5 column 8.

a=a x=1 b=0 c=1 \_ERROR\_=1 \_N\_=1

NOTE: The data set WORK.DATI has 1 observations and 4 variables.

NOTE: DATA statement used:

```
real time      1.00 seconds
cpu time       0.03 seconds
```

| Obs | a | x | b | c |
|-----|---|---|---|---|
| 1   | a | 1 | 0 | 1 |

## OPERATORI LOGICI

| <i>Priorità</i> |   |     |
|-----------------|---|-----|
| 1               | ^ | NOT |
| 2               | & | AND |
| 3               |   | OR  |

Consentono di legare:

- più confronti;
- due espressioni operando sul loro risultato visto nell'aspetto logico:
  - 0, *missing* → risultato *falso*
  - altro → risultato *vero*

### Esempi

NOT(A&B) equivalente a NOT A | NOT B

NOT(A=2) equivalente a A NE 2

If A=2 or A=5 equivalente a if A in(2,5)

If a then nota='pronto'  
equivalente a

If a ne . and a ne 0 then  
nota='pronto'

## ALTRI OPERATORI

(1) Concatenazione di stringhe: ||

STRINGA='Gianni' || 'Rossi'

conversione in best12. per i numeri

(2) MIN e MAX

>< MIN valore minimo tra due operandi

<> MAX valore massimo tra due operandi

```
data dati;
x1=5;
x2=10;
z=x1 min x2;
y=x1 max x2;
run;
proc print;
run;
```

| Obs | x1 | x2 | z | y  |
|-----|----|----|---|----|
| 1   | 5  | 10 | 5 | 10 |

## REGOLE DI CONVERSIONE AUTOMATICA DA VARIABILE NUMERICA A VARIABILE CARATTERE

A variabile carattere, X variabile numerica

•Se si usa A con un operatore che richiede operandi numerici  $Y=A+X$ , SAS converte A in numerica;

•Se si usa un operatore di comparazione,  $Y=A<X$ , la variabile carattere è convertita in numerica;

## ...SEGUE CONVERSIONE AUTOMATICA (1)

Se si usa X con un operatore che richiede operandi carattere,  $B=A||X$ , X è convertita in carattere, formato BEST12. con allineamento a dx (si può ottenere allineamento a sin con la funzione LEFT);

Se si usa X alla sinistra di una istr. di assegnazione e A a dx,  $X=A$ , A è convertita in numerica;

Viceversa, A alla sinistra e X a dx,  $A=X$ , X è convertita carattere, BESTn., n lunghezza di A

## ...SEGUE CONVERSIONE AUTOMATICA (2)

SAS segnala sempre nel LOG la conversione fatta;

```
1 data pippo;
2 input nome $ sesso $ voto;
3 x=sesso+voto;
4 cards;
```

NOTE: Character values have been converted to numeric values at the places given by: (Line):(Column).  
3:3

Se la conversione non è valida mette missing

NOTE: Character values have been converted to numeric values at the places given by:  
(Line):(Column).  
13:3

NOTE: Invalid numeric data, nome='Bianchi' , at line 13 column 3.

```
RULE:-----1-----2-----3-----4-----5--
---+---6---+---7---+---8---+---9
15 Bianchi 1 25
nome=Bianchi sesso=1 voto=25 x=. _ERROR_=1 _N_=1
```

## Esempi di assegnazione

```
*assegna.sas;
*****;
*** esempi di assegnazione ***;
*** osservare messaggi LOG ***;
*****;
*esami sostenuti al primo anno da un gruppo di
studenti;
data dati;
infile cards;
input matric sesso $ stat1 matemat economia
diritto;
CORSO='ECONOMIA'; *COSTANTE CARATTERE;
CODFAC=11; *COSTANTE NUMERICA;
*CALCOLA num esami superato e voto medio;
*UTILIZZO DI UNA FUNZIONE;
nesa=n(STAT1,MATEMAT,ECONOMIA,DIRITTO);
MEDIA=MEAN(STAT1,MATEMAT,ECONOMIA,DIRITTO);
*esprime il voto medio in centesimi;
MEDIA100=100*MEDIA/30; *ESPRESSIONE ARITMETICA;
*dummy per studenti che superano tutti gli esami;
supera=nesa=4;
femmina=(sesso='F');
*concatenazione;
chiave1=corso||sesso;
chiave2=codfac||matric;
*conversione variabile numerica in carattere;
codice=' '||codfac;
*conversione variabile carattere in numerica;
x=codice*1;
cards;
187645 M 25 . . .
250561 F 22 24 . 26
188472 M 23 25 24 .
224514 M 26 28 26 27
213457 F 30 29 27 28
;
run;
```

## ESEMPIO: FINESTRA LOG

```

22 chiave2=codfac||matric;
23 *conversione variabile numerica in carattere;
24 codice=' '|codfac;
25 *conversione variabile carattere in numerica;
26 x=codice*1;
27 cards;

```

NOTE: Numeric values have been converted to character values at the places given by:

(Line):(Column).

22:9 22:17 24:14

NOTE: Character values have been converted to numeric values at the places given by:

(Line):(Column).

26:3

NOTE: The data set WORK.DATI has 5 observations and 17 variables.

NOTE: DATA statement used:

real time 1.59 seconds

## ESEMPIO: FINESTRA OUTPUT

```
proc contents; run;
```

---Alphabetic List of Variables and Attributes----

| #  | Variable | Type | Len | Pos |
|----|----------|------|-----|-----|
| 8  | CODFAC   | Num  | 8   | 40  |
| 7  | CORSO    | Char | 8   | 104 |
| 10 | MEDIA    | Num  | 8   | 56  |
| 11 | MEDIA100 | Num  | 8   | 64  |
| 14 | chiave1  | Char | 16  | 112 |
| 15 | chiave2  | Char | 24  | 128 |
| 16 | codice   | Char | 14  | 152 |
| 6  | diritto  | Num  | 8   | 32  |
| 5  | economia | Num  | 8   | 24  |
| 13 | femmina  | Num  | 8   | 80  |
| 4  | matemat  | Num  | 8   | 16  |
| 1  | matric   | Num  | 8   | 0   |
| 9  | nesa     | Num  | 8   | 48  |
| 2  | sesso    | Char | 8   | 96  |
| 3  | stat1    | Num  | 8   | 8   |
| 12 | supera   | Num  | 8   | 72  |
| 17 | x        | Num  | 8   | 88  |

```
proc print;run;
```

```

          e          M
          m c d      E f c c
m        a o i      D s e h h c
a s s t n r C O M I u m i i o
t e t e o i O D n E A p m a a d
O r s a m m t R F e D 1 e i v v i
b i s t a i t S A s I 0 r n e e c
s c o 1 t a o O C a A 0 a a 1 2 e x

1 187645 M 25 . . . ECONOMIA 11 1 25.00 83.3333 0 0 ECONOMIAM 11 187645 11 11
2 250561 F 22 24 . 26 ECONOMIA 11 3 24.00 80.0000 0 1 ECONOMIAF 11 250561 11 11
3 188472 M 23 25 24 . ECONOMIA 11 3 24.00 80.0000 0 0 ECONOMIAM 11 188472 11 11
4 224514 M 26 28 26 27 ECONOMIA 11 4 26.75 89.1667 1 0 ECONOMIAM 11 224514 11 11
5 213457 F 30 29 27 28 ECONOMIA 11 4 28.50 95.0000 1 1 ECONOMIAF 11 213457 11 11

```



## DO ITERATIVO

```
*input7.sas;
title 'Lettura record a lunghezza variabile';
data dati;
infile datalines;
input nome $ 1-10 cognome $ 12-26 nesa @;
*loop su dati esami: crea una obs per ogni esame;
do i=1 to nesa;
  input cod_esa voto_esa giorno mese anno @;
  dataesa=mdy(mese,giorno,anno);
  output;
end;
drop i;
datalines;
Giuseppe Bianchi 2 55555 25 10 6 99 44444 27 12 9 99
Maria Rossi 3 54352 28 10 11 98 33333 26 18 2 99 22222 26 25 6 99
Franco Verdi 2 22222 22 6 10 99 33333 23 12 12 99
;
```

```
proc print data=dati;
format dataesa date7.;
run;
```

Lettura record a lunghezza variabile 1

| Obs | nome     | cognome | nesa | cod_esa | voto_esa | giorno | mese | anno | dataesa |
|-----|----------|---------|------|---------|----------|--------|------|------|---------|
| 1   | Giuseppe | Bianchi | 2    | 55555   | 25       | 10     | 6    | 99   | 10JUN99 |
| 2   | Giuseppe | Bianchi | 2    | 44444   | 27       | 12     | 9    | 99   | 12SEP99 |
| 3   | Maria    | Rossi   | 3    | 54352   | 28       | 10     | 11   | 98   | 10NOV98 |
| 4   | Maria    | Rossi   | 3    | 33333   | 26       | 18     | 2    | 99   | 18FEB99 |
| 5   | Maria    | Rossi   | 3    | 22222   | 26       | 25     | 6    | 99   | 25JUN99 |
| 6   | Franco   | Verdi   | 2    | 22222   | 22       | 6      | 10   | 99   | 06OCT99 |
| 7   | Franco   | Verdi   | 2    | 33333   | 23       | 12     | 12   | 99   | 12DEC99 |

## DO WHILE

```
*dowhile.sas;

options nocenter nodate nonumber;
title 'Istruzione DO WHILE';

*****;
*** LEGGE DATI CENSIMENTO POPOLAZIONE 1981 ***;
*** SELEZIONA FAMIGLIE CON CAPOFAMIGLIA OCCUPATO ***;
*****;

*selezione osservazioni con capofamiglia occupato;
data capofam;
infile 'e:\corso_marzo_02\cens81t.txt' missover
obs=30;
input tr 1. @;
*seleziona record residenti famiglia;
if tr ne 2 then delete;
else input idfam 11-13 cf 17 cond 46 @;
*selezione record capofam occupato e scrive record
suo e suoi componenti;
if cf eq 1 and cond ne 1 then delete;
else if cf eq 1 and cond eq 1 then no=idfam;
do while(idfam=no);
  input cf 17 sesso 18 stciv 29 cond 46;
  output;
  input no 11-13 @;
end;
drop no;
label idfam='codice famiglia' cond='condizione
professionale' cf='relazione con c.f.';
run;
```

NOTE: The infile 'e:\corso\_marzo\_02\cens81t.txt' is:  
 File Name=e:\corso\_marzo\_02\cens81t.txt,  
 RECFM=V,LRECL=256

NOTE: 30 records were read from the infile  
 'e:\corso\_marzo\_02\cens81t.txt'.

The minimum record length was 34.

The maximum record length was 71.

NOTE: The data set WORK.CAPOFAM has 17 observations and 6  
 variables.

NOTE: DATA statement used:

real time 0.22 seconds

```
proc print;
id idfam;
run;
```

Istruzione DO WHILE

| idfam | tr | cf | cond | sesso | stciv |
|-------|----|----|------|-------|-------|
| 1     | 2  | 1  | 1    | 1     | 2     |
| 1     | 2  | 2  | 4    | 2     | 2     |
| 1     | 2  | 3  | 5    | 2     | 1     |
| 1     | 2  | 3  | .    | 2     | 1     |
| 1     | 2  | 3  | .    | 1     | 1     |
| 5     | 2  | 1  | 1    | 1     | 2     |
| 5     | 2  | 2  | 4    | 2     | 2     |
| 6     | 2  | 1  | 1    | 1     | 2     |
| 6     | 2  | 2  | 4    | 2     | 2     |
| 6     | 2  | 3  | .    | 1     | 1     |
| 8     | 2  | 1  | 1    | 1     | 2     |
| 8     | 2  | 2  | 1    | 2     | 2     |
| 8     | 2  | 3  | .    | 1     | 1     |
| 9     | 2  | 1  | 1    | 1     | 2     |
| 9     | 2  | 2  | 1    | 2     | 2     |
| 9     | 2  | 3  | .    | 1     | 1     |
| 9     | 2  | 3  | .    | 2     | 1     |

## ARRAY

```
*array1.sas;
data studenti;
infile datalines eof=stop;
array esame(10) esame1-esame10;
*nb: condizione while valutata a inizio ciclo;
do i=1 to 10 while(matric=nextmatric);
  input matric 1-5 @7 esame(i) mmddy8.;
  input nextmatric 1-5 @@;
end;
drop nextmatric i;
*output scrive obs nel data set quando while e'
falsa;
*(serve per ultimo studente del file originale);
stop: output;
datalines;
11111 12/02/97
11111 02/15/98
11111 06/20/98
11111 09/15/98
22222 12/05/97
33333 12/02/97
33333 02/14/98
33333 06/12/98
44444 12/08/97
44444 02/16/98
44444 06/01/98
44444 07/15/98
44444 09/12/98
44444 12/12/98
;
run;
```

NOTE: The data set WORK.STUDENTI has 4 observations and 11 variables.

NOTE: DATA statement used:  
real time 0.10 seconds

```
proc print data=studenti;
format esame1-esame10 date7.;
title 'Esempio lettura di più record per u.s.:
data esami per studente';
run;
```

Esempio lettura di più record per u.s.: data esami per studente

| Obs | esame1  | esame2  | esame3  | esame4  | esame5  | esame6  | esame7 | esame8 | esame9 | esame10 | matric |
|-----|---------|---------|---------|---------|---------|---------|--------|--------|--------|---------|--------|
| 1   | 02DEC97 | 15FEB98 | 20JUN98 | 15SEP98 | .       | .       | .      | .      | .      | .       | 11111  |
| 2   | 05DEC97 | .       | .       | .       | .       | .       | .      | .      | .      | .       | 22222  |
| 3   | 02DEC97 | 14FEB98 | 12JUN98 | .       | .       | .       | .      | .      | .      | .       | 33333  |
| 4   | 08DEC97 | 16FEB98 | 01JUN98 | 15JUL98 | 12SEP98 | 12DEC98 | .      | .      | .      | .       | 44444  |

## GENERAZIONE DI NUMERI CASUALI

*Distribuzione uniforme in (0,1).*

**RANUNI(***seme***)**

*seme* intero  $< 2^{31} - 1$ . Se  $seme \leq 0$ , la sequenza di numeri casuali viene inizializzata utilizzando come seme l'ora dell'orologio interno.

Utilizza un generatore a modulo moltiplicativo con  $m=2^{31} - 1$  e  $a=397204094$ :

$$X_{n+1} = (aX_n) \bmod(2^{31} - 1)$$

- $seme = X_0$ ;
- $X_0$  resta invariato per tutta l'esecuzione;
- Per variare  $X_0$  usare CALL.

## NUMERI CASUALI

### *Distribuzione normale*

**RANNOR**(*seme*)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu)^2 / 2\sigma^2\right]$$
$$\mu = 0, \sigma^2 = 1$$

Trasformazione Box-Muller da uniforme.

### *Distribuzione gamma*

**RANGAM**(*seme, a*)

$$f(x) = \left(\frac{x}{b}\right)^{a-1} \frac{\exp(-x/b)}{b\Gamma(a)} I_{(0,\infty)}(x)$$
$$b > 0, a > 0$$

- per  $b=1 \rightarrow$  Gamma standard
- per  $2*a$  intero  $\rightarrow$  chi-quadro con  $2*a$  g.l.

Metodo di accettazione/rifiuto (Cheng, 1977; Fishman, 1978).

## NUMERI CASUALI

### *Distribuzione binomiale*

**RANBIN**(*seme, n, p*)

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\dots,n\}}(x)$$
$$0 \leq p \leq 1, n \in \mathcal{N}$$

- se  $n \leq 50$  o  $np \leq 5 \rightarrow$  trasformazione inversa da uniforme
- se  $n > 50$  e  $np > 5 \rightarrow$  approssimazione normale  $\rightarrow$  trasformazione Box-Muller da uniforme

### *Distribuzione esponenziale*

**RANEXP**(*seme*)

$$f(x) = e^{-x} I_{(0,\infty)}(x)$$

Metodo della trasformazione inversa da uniforme.

## FUNZIONI SAS PER LA GENERAZIONE DI NUMERI CASUALI

NORMAL normale standard  
RANNOR normale standard  
RANBIN binomiale  
RANCAU Cauchy  
RANEXP esponenziale standard  
RANGAM gamma standard  
RANPOI Poisson  
RANTBL distribuzione discreta  
RANTRI distribuzione triangolare  
RANUNI uniforme (0,1)  
UNIFORM uniforme (0,1)

## CALCOLO DELLA FUNZIONE DI RIPARTIZIONE

|          |                    |
|----------|--------------------|
| POISSON  | Poisson            |
| PROBBETA | Beta               |
| PROBBNML | Binomiale          |
| PROBCHI  | Chi-quadrato       |
| PROBF    | F di Fisher        |
| PROBGAM  | Gamma              |
| PROBHYP  | Ipergeometrica     |
| PROBNEG  | Binomiale negativa |
| PROBNORM | Normale standard   |
| PROBT    | t di Student       |

## QUANTILI

|         |                  |
|---------|------------------|
| BETAINV | Beta             |
| CINV    | Chi-quadrato     |
| FINV    | F di Fisher      |
| GAMINV  | Gamma            |
| PROBIT  | Normale standard |
| TINV    | t di Student     |

## LETTURA ARCHIVI GERARCHICI

(vedi diapositiva 13 di istruzioni condizionali.ppt)

```
*censimento2.sas;

options nocenter nodate nonumber;
title 'Lettura di file gerarchico: crea archivio sas che legge
info da livello diverso';
title2 'utilizzo RETAIN per associare info abitazioni ad ogni
individuo';

*****;
*** archivio ASCII cens81.dat: 5 tracciati record diversi ***;
***   col.1=1==>abitazioni ***;
***   col.1=2==>residenti in famiglie ***;
***   archivio ordinato per codice famiglia: ***;
***   prima record abitazione e poi records componenti ***;
*****;

*selezione info relative a residenti e abitazione della famiglia;
*uso RETAIN per associare ad ogni record individuale info abitazione;
*uso RETURN per inizio nuovo step:
  se tr=1 SAS legge valore stanze e servizi e torna a inizio step per lettura
nuovo record;

data famiglie;
retain stanze bagno;
infile 'v:\didat\carla\sas\cens81t.txt' missover;
input tr 1. @;
if tr=1 then do;
  input @25 stanze 2. @36 bagno 1.;
  return;
end;
else if tr=2 then do;
  input comp 15-16 sesso 18;
  output;
end;
drop tr;
run;
proc print data=famiglie (obs=50);
run;
```

## LETTURA ARCHIVI GERARCHICI: LOG

NOTE: The infile 'v:\didat\carla\sas\cens81t.txt' is:  
File Name=v:\didat\carla\sas\cens81t.txt,  
RECFM=V,LRECL=256

NOTE: 2267 records were read from the infile  
'v:\didat\carla\sas\cens81t.txt'.

The minimum record length was 33.

The maximum record length was 71.

NOTE: The data set WORK.FAMIGLIE has 808 observations and  
4 variables.

NOTE: DATA statement used:

real time 0.32 seconds

## LETTURA ARCHIVI GERARCHICI: OUTPUT

Lettura di file gerarchico: crea archivio sas che legge  
info da livello diverso  
utilizzo RETAIN per associare info abitazioni ad ogni  
individuo

| Obs | stanze | bagno | comp | sezzo |
|-----|--------|-------|------|-------|
| 1   | 4      | 1     | 1    | 1     |
| 2   | 4      | 1     | 2    | 2     |
| 3   | 4      | 1     | 3    | 2     |
| 4   | 4      | 1     | 4    | 2     |
| 5   | 4      | 1     | 5    | 1     |

## ISTRUZIONE SUM

```
*sum.sas;
*****;
*** esempio di utilizzo sum statment ***;
*****;
data dati;
input stanze case;
cumula+(stanze*case);
totfreq+case;
cards;
1 100
2 230
3 245
4 180
5 95
6 20
;
proc print;
sum case;
run;
```

Letture di file gerarchico: crea archivio sas che legge info da livello diverso  
utilizzo RETAIN per associare info abitazioni ad ogni individuo

| Obs | stanze | case | cumula | totfreq |
|-----|--------|------|--------|---------|
| 1   | 1      | 100  | 100    | 100     |
| 2   | 2      | 230  | 560    | 330     |
| 3   | 3      | 245  | 1295   | 575     |
| 4   | 4      | 180  | 2015   | 755     |
| 5   | 5      | 95   | 2490   | 850     |
| 6   | 6      | 20   | 2610   | 870     |
|     |        | ==== |        |         |
|     |        | 870  |        |         |

## ISTRUZIONE SUM

```
data reddito;
input tasso @@;
retain reddito 2000 prod 1 ;
reddito=reddito*(1+tasso);
prod=prod*(1+tasso);
cumulata+reddito;
t+1;
*media aritmetica del montante annuo;
media=cumulata/_n_;
*media geometrica dei tassi di interesse;
tmedio=prod**((1/t)-1);

cards;
0.08 0.072 0.064 0.053 0.048 0.042 0.040 0.032
0.028
proc print data=reddito;
id t;
format tmedio 6.4;
run;
```

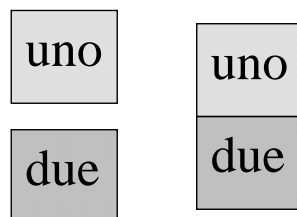
| t | tasso | reddito | prod    | cumulata | media   | tmedio |
|---|-------|---------|---------|----------|---------|--------|
| 1 | 0.080 | 2160.00 | 1.08000 | 2160.00  | 2160.00 | 0.0800 |
| 2 | 0.072 | 2315.52 | 1.15776 | 4475.52  | 2237.76 | 0.0760 |
| 3 | 0.064 | 2463.71 | 1.23186 | 6939.23  | 2313.08 | 0.0720 |
| 4 | 0.053 | 2594.29 | 1.29715 | 9533.52  | 2383.38 | 0.0672 |
| 5 | 0.048 | 2718.82 | 1.35941 | 12252.34 | 2450.47 | 0.0633 |
| 6 | 0.042 | 2833.01 | 1.41650 | 15085.35 | 2514.22 | 0.0597 |
| 7 | 0.040 | 2946.33 | 1.47316 | 18031.67 | 2575.95 | 0.0569 |
| 8 | 0.032 | 3040.61 | 1.52030 | 21072.28 | 2634.04 | 0.0538 |
| 9 | 0.028 | 3125.75 | 1.56287 | 24198.03 | 2688.67 | 0.0509 |

## TRATTAMENTO DI SAS DATA SET

Concatenazione    istruzione SET

(a) uno sull'altro

```
Data tot;
set uno due;
run;
```



(b) interleaving

```
Data tot;
set uno due;
by anno;
run;
```

|      |      |      |
|------|------|------|
| 1991 | 1991 | 1991 |
| 1992 | 1993 | 1991 |
| 1993 | 1994 | 1992 |
| 1994 | 1995 | 1993 |
|      |      | 1994 |
|      |      | 1994 |
|      |      | 1995 |

## CONCATENAMENTO DI SAS DATA SET

```
*set1.sas;
*****;
***          Istruzione SET          ***;
*** concatenamento verticale di due data set ***;
*****;
* TESTO;
* Il Parco XX ha due ingressi, a SUD e a NORD.
  Vengono raccolti dati sui visitatori:
  numero del biglietto, numero componenti il
  gruppo, età del visitatore, numero del parcheggio
  (un solo parcheggio per l'ingresso SUD);

*INPUT DEI DATI;
data sud;
input entrata $ nbigl ngruppo eta;
cards;
S 43 3 27
S 44 3 24
S 45 3 2
;
proc print;
run;
data NORD;
input entrata $ nbigl ngruppo AGE park;
cards;
N 21 5 41 1
N 87 4 33 3
N 65 2 67 1
N 66 2 7 1
;
proc print;
run;
```



```

*CREA UN DATA SET CONCATENADO NORD E SUD;
*CREA LA VARIABILE SPESA (in dollari), in
base alla variabile ETA;
data unisci;
set sud nord(RENAME=(AGE=ETA));
if eta=. then spesa=.;
else if eta < 3 then spesa=0;
else if eta < 65 then spesa=17;
else spesa=12;
run;
*NB: valore missing per PARK per le obs
provenienti da SUD;
proc print;
run;

```

```
Obs  entrata  nbigl  gruppo  eta
```

```
1    S      43    3      27
2    S      44    3      24
3    S      45    3      2
```

```
Obs  entrata  nbigl  ngruppo  AGE  park
```

```
1    N      21    5      41    1
2    N      87    4      33    3
3    N      65    2      67    1
4    N      66    2      7     1
```

```
Obs  entrata  nbigl  ngruppo  eta  park  spesa
```

```
1    S      43    3      27    .    17
2    S      44    3      24    .    17
3    S      45    3      2     .    0
4    N      21    5      41    1    17
5    N      87    4      33    3    17
6    N      65    2      67    1    12
7    N      66    2      7     1    17
```

```

*set2.sas;
*****;
***          Istruzione SET          ***;
*** concatenamento ordinato di due data set ***;
*****;
* TESTO;
* Il Parco XX ha due ingressi, a SUD e a NORD. Vengono
raccolti dati sui visitatori:
  numero del biglietto,  numero componenti il gruppo, eta
del visitatore,
  numero del parcheggio (un solo parcheggio per
l'ingresso SUD);

options nodate nonumber nocenter;
*ORDINA GLI ARCHIVI IN BASE A NBIGL (numero del
biglietto);
proc sort data=sud;
by nbigl;
run;
PROC PRINT;
RUN;

```

```
Obs  entrata  nbigl  ngruppo  eta
1    S      43    3      27
2    S      44    3      24
3    S      45    3      2
```

```

proc sort data=nord;
by nbigl;
run;
PROC PRINT;
RUN;

```

```
Obs  entrata  nbigl  ngruppo  AGE  park
1    N      21    5      41    1
2    N      65    2      67    1
3    N      66    2      7     1
4    N      87    4      33    3
```

```

*CREA UN DATA SET CONCATENADO ORDINATAMENTE PER NBIGL
NORD E SUD;
TITLE 'Concatenamento verticale archivi in
base al numero del biglietto';
data unisci;
set sud nord(RENAME=(AGE=ETA));
BY NBIGL;
run;
*NB: valore missing per PARK per le obs
provenienti da SUD;
proc print;
run;

```

Concatenamento verticale archivi in base al numero del biglietto

| Obs | entrata | nbigl | ngruppo | eta | park |
|-----|---------|-------|---------|-----|------|
| 1   | N       | 21    | 5       | 41  | 1    |
| 2   | S       | 43    | 3       | 27  | .    |
| 3   | S       | 44    | 3       | 24  | .    |
| 4   | S       | 45    | 3       | 2   | .    |
| 5   | N       | 65    | 2       | 67  | 1    |
| 6   | N       | 66    | 2       | 7   | 1    |
| 7   | N       | 87    | 4       | 33  | 3    |

## TRATTAMENTO DI SAS DATA SET

*Fusione*

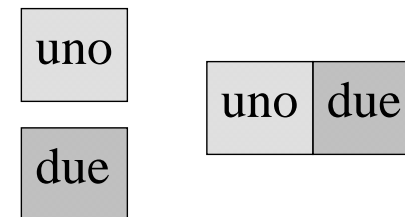
istr. SET e MERGE

(a) uno a uno

```

Data tot;
set uno;
set due;
run;

```



oppure

```

data tot;
merge uno due;
run;

```

(b) match-merging

```

data tot;
merge uno due;
by anno;
run;

```

|      |      |      |   |
|------|------|------|---|
| 91 a | 91 1 | 91 a | 1 |
| 92 b | 91 2 | 91 a | 2 |
| 93 c | 93 3 | 92 b | . |
| 94 d | 94 4 | 93 c | 3 |
|      |      | 94 d | 4 |

## CONCATENAMENTO ORIZZONTALE SAS DATA SETS: MERGE

\*mergel.sas;

\* Un cioccolataio registra i dati relativi al tipo di cioccolato venduto ogni giorno. Il tipo di cioccolata e il n. di pezzi venduti ogni giorno sono registrati in un file. In un file separato sono registrati il nome e la descrizione di ogni tipo di cioccolata e il numero di codice corrispondente. Al fine di stampare tutte le informazioni, e' necessario unire i due archivi in base al numero di codice del tipo di cioccolata;

```
data vendite;
input codice $ 1-4 pezzi 6-7;
cards;
C865 15
K086 9
A536 21
S163 34
K014 1
A206 12
B713 29
;
```

```
DATA descr;
input codice $ 1-4 nome $ 6-14 descr $15-60;
cards;
A206 Mokka      Cioccolato fondente al caffè
A536 Nocciola   Cioccolato fondente con nocciole a pezzi
B713 Mirtilli   Cioccolato al latte con crema ai mirtilli
C865 Vaniglia   Cioccolato alla vaniglia
K014 Menta      Cioccolato al latte con crema alla menta
K086 Konig      Cioccolato al latte con crema di nocciole
M315 Piramide   Cioccolato bianco ripieno di cioccolato
fondente
S163 Orbais     Cioccolato fondente con ripieno di crema
;
```

```
*ordina archivio vendite in base a codice
prodotto;
proc sort data=vendite;
by codice;
run;

*unisce archivi in base a codice prodotto;
data unisci;
merge descr vendite;
by codice;
run;

proc print;
id codice;
run;
```

| codice | nome     | descr                                         | pezzi |
|--------|----------|-----------------------------------------------|-------|
| A206   | Mokka    | Cioccolato fondente al caffè                  | 12    |
| A536   | Nocciola | Cioccolato fondente con nocciole a pezzi      | 21    |
| B713   | Mirtilli | Cioccolato al latte con crema ai mirtilli     | 29    |
| C865   | Vaniglia | Cioccolato alla vaniglia                      | 15    |
| K014   | Menta    | Cioccolato al latte con crema alla menta      | 1     |
| K086   | Konig    | Cioccolato al latte con crema di nocciole     | 9     |
| M315   | Piramide | Cioccolato bianco ripieno di cioccolato fonde | .     |
| S163   | Orbais   | Cioccolato fondente con ripieno di crema      | 34    |

## CONCATENAMENTO UNO A MOLTI

```
*merge2.sas;  
*TESTO;  
* L' esercente di un distributore automatico di  
videocassette fa un prezzo speciale, scontando dal 20 al  
30% il prezzo regolare. Il distributore ha due archivi:  
  uno relativo alle videocassette (un record per ogni  
video: titolo, tipo, prezzo) e l'altro con i tassi di  
sconto per tipo di videocassetta (tipo e sconto);
```

```
data video;  
input nome $ 1-29 tipo $ prezzo;  
cards;
```

|                 |          |       |
|-----------------|----------|-------|
| Pokemon         | cartoni  | 22000 |
| Pane e Tulipani | commedia | 40000 |
| Tom e Jerry     | cartoni  | 29000 |
| Amici miei      | commedia | 40000 |
| Soldato blu     | western  | 35000 |
| Io e Annie      | commedia | 30000 |
| Corvo Rosso     | western  | 38000 |

```
;  
run;
```

```
data sconti;  
input tipo $ sconto ;  
cards;
```

|          |      |
|----------|------|
| cartoni  | 0.20 |
| commedia | 0.25 |
| western  | 0.30 |

```
;  
run;
```

```
*ordina archivio videocassette in base a  
tipo;
```

```
proc sort data=video;  
by tipo;  
run;
```

```
*unisce archivi in base a tipo;  
*calcola prezzo scontato;  
data unisci;  
merge video sconti;  
by tipo;  
prezzo_n=round(prezzo-prezzo*sconto);  
run;
```

```
proc print;  
run;
```

| Obs | nome            | tipo     | prezzo | sconto | prezzo_n |
|-----|-----------------|----------|--------|--------|----------|
| 1   | Pokemon         | cartoni  | 22000  | 0.20   | 17600    |
| 2   | Tom e Jerry     | cartoni  | 29000  | 0.20   | 23200    |
| 3   | Pane e Tulipani | commedia | 40000  | 0.25   | 30000    |
| 4   | Amici miei      | commedia | 40000  | 0.25   | 30000    |
| 5   | Io e Annie      | commedia | 30000  | 0.25   | 22500    |
| 6   | Soldato blu     | western  | 35000  | 0.30   | 24500    |
| 7   | Corvo Rosso     | western  | 38000  | 0.30   | 26600    |

## FUSIONE UNO A UNO CON BY E IN=

```

title 'fusione uno a uno con istruzioni BY e
opzione IN=';
title2 'archivio NOMI';
proc sort data=nomi;
by nome;
proc print data=nomi;
title2 'archivio INFO';
proc sort data=info;
by nome;
proc print data=info;

title2 'archivio UNOAUNO';
data unoauno;
merge nomi(in=n) info(in=i);
NOMI=N; INFO=I;
by nome;
*if n and i;
*if n=1 and I=1 then output;
run;

proc print data=unoauno;
run;

```

### Archivio NOMI

| Obs | nome     | eta | voto | codfac |
|-----|----------|-----|------|--------|
| 1   | anna     | 23  | 30   | 2      |
| 2   | franco   | 21  | 22   | 2      |
| 3   | giuseppe | 24  | 30   | 1      |
| 4   | laura    | 23  | 18   | 2      |
| 5   | marco    | 22  | 28   | 1      |
| 6   | maria    | 22  | 26   | 1      |

### Archivio INFO

| Obs | nome     | voto2 | voto3 | anno |
|-----|----------|-------|-------|------|
| 1   | franco   | 23    | 22    | 99   |
| 2   | giuseppe | 24    | 28    | 98   |
| 3   | laura    | 20    | 18    | 98   |
| 4   | luisa    | 28    | 30    | 98   |
| 5   | marco    | 29    | 30    | 99   |
| 6   | maria    | 24    | 25    | 98   |

### Archivio UNOAUNO

| Obs | nome     | eta | voto | codfac | voto2 | voto3 | anno | NOMI | INFO |
|-----|----------|-----|------|--------|-------|-------|------|------|------|
| 1   | franco   | 21  | 22   | 2      | 23    | 22    | 99   | 1    | 1    |
| 2   | giuseppe | 24  | 30   | 1      | 24    | 28    | 98   | 1    | 1    |
| 3   | laura    | 23  | 18   | 2      | 20    | 18    | 98   | 1    | 1    |
| 4   | marco    | 22  | 28   | 1      | 29    | 30    | 99   | 1    | 1    |
| 5   | maria    | 22  | 26   | 1      | 24    | 25    | 98   | 1    | 1    |

## FUSIONE UNO A MOLTI CON BY E IN=

```

title 'fusione uno a molti con istruzioni BY
e opzione IN=';
proc sort data=nomi;
by codfac;
proc sort data=facolta;
by codfac;
data unoamolti;
merge nomi (in=n) facolta;
by codfac;
if n ;
run;
proc print data=unoamolti;
run;

```

### ARCHIVIO NOMI

| Obs | nome     | eta | voto | codfac |
|-----|----------|-----|------|--------|
| 1   | giuseppe | 24  | 30   | 1      |
| 2   | marco    | 22  | 28   | 1      |
| 3   | maria    | 22  | 26   | 1      |
| 4   | anna     | 23  | 30   | 2      |
| 5   | franco   | 21  | 22   | 2      |
| 6   | laura    | 23  | 18   | 2      |

### ARCHIVIO FACOLTA

| Obs | codfac | descrizione       | gruppo         |
|-----|--------|-------------------|----------------|
| 1   | 1      | Economia          | socio-economic |
| 2   | 2      | Scienze Politiche | socio-economic |
| 3   | 3      | Medicina          | medico         |

### ARCHIVIO UNOAMOLTI

| Obs | nome     | eta | voto | codfac | descrizione       | gruppo         |
|-----|----------|-----|------|--------|-------------------|----------------|
| 1   | giuseppe | 24  | 30   | 1      | Economia          | socio-economic |
| 2   | marco    | 22  | 28   | 1      | Economia          | socio-economic |
| 3   | maria    | 22  | 26   | 1      | Economia          | socio-economic |
| 4   | anna     | 23  | 30   | 2      | Scienze Politiche | socio-economic |
| 5   | franco   | 21  | 22   | 2      | Scienze Politiche | socio-economic |
| 6   | laura    | 23  | 18   | 2      | Scienze Politiche | socio-economic |

## FUSIONE MOLTI A MOLTI CON BY E IN=

```

proc sort data=unoauno;
by codfac anno;
proc sort data=matricole;
by codfac anno;

data moltiamolti;
merge unoauno matricole;
by codfac anno;
run;

```

### archivio UNOAUNO

| Obs | nome     | eta | voto | codfac | voto2 | voto3 | anno | NOMI | INFO |
|-----|----------|-----|------|--------|-------|-------|------|------|------|
| 1   | giuseppe | 24  | 30   | 1      | 24    | 28    | 98   | 1    | 1    |
| 2   | maria    | 22  | 26   | 1      | 24    | 25    | 98   | 1    | 1    |
| 3   | marco    | 22  | 28   | 1      | 29    | 30    | 99   | 1    | 1    |
| 4   | laura    | 23  | 18   | 2      | 20    | 18    | 98   | 1    | 1    |
| 5   | franco   | 21  | 22   | 2      | 23    | 22    | 99   | 1    | 1    |

### archivio MATRICOLE

| Obs | codfac | anno | matricole |
|-----|--------|------|-----------|
| 1   | 1      | 98   | 600       |
| 2   | 1      | 99   | 560       |
| 3   | 2      | 98   | 760       |
| 4   | 2      | 99   | 650       |

### archivio MOLTIAMOLTI

| Obs | nome     | eta | voto | codfac | voto2 | voto3 | anno | NOMI | INFO | matric |
|-----|----------|-----|------|--------|-------|-------|------|------|------|--------|
| 1   | giuseppe | 24  | 30   | 1      | 24    | 28    | 98   | 1    | 1    | 600    |
| 2   | maria    | 22  | 26   | 1      | 24    | 25    | 98   | 1    | 1    | 600    |
| 3   | marco    | 22  | 28   | 1      | 29    | 30    | 99   | 1    | 1    | 560    |
| 4   | laura    | 23  | 18   | 2      | 20    | 18    | 98   | 1    | 1    | 760    |
| 5   | franco   | 21  | 22   | 2      | 23    | 22    | 99   | 1    | 1    | 650    |

## PROCEDURE SAS BASE

Programmi generalizzati che leggono ed elaborano, di norma, SAS Data Set.

**PROC** *nome* [*opzioni*];

### *Procedure di utilità*

CONTENTS

FORMAT

SORT

### *Procedure di report*

PRINT

### *Procedure Statistiche*

FREQ

MEANS

UNIVARIATE

TABULATE

REG, CORR

## ISTRUZIONI PROC STEP

```
PROC nome [opzioni];  
...  
(altre istruzioni)  
...  
RUN;
```

- Istruzioni *proprie* di una specifica procedura → agiscono solo in una determinata procedura;
- Istruzioni *tipiche* del passo di procedura (*proc step*) → possono essere utilizzate in molte procedure con lo stesso significato, agiscono a livello di singolo *proc step*;
- Istruzioni del linguaggio SAS *comuni* ai passi di Data e Proc → possono essere utilizzate ovunque (es. TITLE).

## ISTRUZIONI COMUNI PER IL FORMATO DELL'OUTPUT

**TITLE** *n* 'descrizione max 132 caratteri';

**FOOTNOTE** *n* 'descrizione max 132 car.';

**LABEL** *var*='descr. max 40 car.' ...;

*esempio*

```
TITLE1 'Popolazione residente al 1988 e superficie';  
TITLE2 'Regioni italiane per ripartizione.';  
FOOTNOTE 'Fonte: Istat 1989, Compendio Statistico Italiano';  
PROC PRINT DATA=dati.regioni LABEL;  
WHERE RIPARIZ LE 2;  
LABEL RIPARTIZ='ripartizione geografica';  
RUN;
```

Popolazione residente al 1988 e superficie  
Regioni italiane per ripartizione.

| OBS | ripartizione geografica | REGIONE               | SUPERF   | POP88   | MASCHI  |
|-----|-------------------------|-----------------------|----------|---------|---------|
| 1   | 1                       | PIEMONTE              | 25398.94 | 4377229 | 2118931 |
| 2   | 1                       | VALLE D'AOSTA         | 3262.26  | 114325  | 56630   |
| 3   | 1                       | LOMBARDIA             | 23858.55 | 8886402 | 4287700 |
| 4   | 2                       | TRENTINO-ALTO ADIGE   | 13618.31 | 881986  | 430461  |
| 5   | 2                       | VENETO                | 18364.56 | 4374911 | 2120827 |
| 6   | 2                       | FRIULI-VENEZIA GIULIA | 7844.13  | 1210242 | 577195  |
| 7   | 1                       | LIGURIA               | 5417.97  | 1749572 | 830641  |
| 8   | 2                       | EMILIA ROMAGNA        | 22123.18 | 3924199 | 1895131 |

Fonte: ISTAT 1989, COMPENDIO STATISTICO ITALIANO



## ISTRUZIONI DI USO GENERALE

BY → analisi gruppi omogenei di obs  
WHERE → scelta osservazioni  
VAR → scelta delle variabili  
FORMAT → definizione formati report  
WEIGHT → ponderazione delle obs  
FREQ → ponderazione delle obs

## ISTRUZIONE BY

Esecuzione separata della procedura per gruppi omogenei di obs. Le obs *devono essere ordinate* in base alle variabili BY.

```
proc sort data=dati.regioni;  
by ripartiz;  
PROC PRINT DATA=dati.regioni;  
by ripartiz;  
RUN;
```

```
RIPARTIZ=1
```

| OBS | REGIONE       | SUPERF   | POP88   | MASCHI  |
|-----|---------------|----------|---------|---------|
| 1   | PIEMONTE      | 25398.94 | 4377229 | 2118931 |
| 2   | VALLE D'AOSTA | 3262.26  | 114325  | 56630   |
| 3   | LOMBARDIA     | 23858.55 | 8886402 | 4287700 |
| 4   | LIGURIA       | 5417.97  | 1749572 | 830641  |

## SELEZIONE DELLE OSSERVAZIONI: WHERE E WHERE=

Agisce prima che l'osservazione sia caricata nel PDV → le osservazioni escluse dalla WHERE non sono disponibili nel DATA Step.

**WHERE** espressione;  
**SET** data\_set (**WHERE**=(espressione));

- Non si possono usare funzioni SAS;
- Sono disponibili operatori 'speciali'

## ISTRUZIONE WHERE

Seleziona le osservazioni in fase di esecuzione della procedura, lasciando inalterato l'archivio da cui legge i dati.

### NB:

(a) WHERE viene eseguita *dopo* che hanno avuto effetto le opzioni relative al SAS Data Set.

```
PROC PRINT DATA=dati.regioni (KEEP=regione
    ripartiz pop88 rename=(ripartiz=zona));
WHERE zona=3;
RUN;
```

(b) WHERE viene eseguita *prima* dell'istr. BY.

```
PROC PRINT DATA=dati.regioni;
by ripartiz;
where ripartiz ne 2;
RUN;
```

## OPERATORI WHERE

WHERE espr **BETWEEN** val1 **AND** val2

WHERE var **CONTAINS** 'stringa'

WHERE var1 **CONTAINS** TRIM(var2)

WHERE var **IS NULL|IS MISSING**

WHERE var **LIKE** 'stringa'

WHERE var=**\*** 'stringa' (sound-like)

WHERE SAME AND espressione

[where.sas](#)

## SELEZIONE DELLE OSSERVAZIONI CON OPERATORE BETWEEN-AND

```
proc print data=regioni n noobs;
where pop88 not between 1046114
and 4376070;
run;
```

log

NOTE: There were 10 observations read from the data set WORK.REGIONI.

```
WHERE not ((pop88>=1046114 and
pop88<=4376070));
```

output

| rip | regione             | superf   | pop88   | maschi  |
|-----|---------------------|----------|---------|---------|
| 1   | PIEMONTE            | 25398.94 | 4377229 | 2118931 |
| 1   | VALLE D'AOSTA       | 3262.26  | 114325  | 56630   |
| 1   | LOMBARDIA           | 23858.55 | 8886402 | 4287700 |
| 2   | TRENTINO-ALTO ADIGE | 13618.31 | 881986  | 430461  |
| 3   | UMBRIA              | 8456.04  | 818226  | 398466  |
| 3   | LAZIO               | 17202.74 | 5137270 | 2488272 |
| 4   | MOLISE              | 4437.64  | 334680  | 163641  |
| 4   | CAMPANIA            | 13595.33 | 5731426 | 2830845 |
| 4   | BASILICATA          | 9992.27  | 621506  | 307826  |
| 5   | SICILIA             | 25707.23 | 5141343 | 2512105 |

N = 10

## SELEZIONE DELLE OSSERVAZIONI CON OPERATORE CONTAINS

```
proc print n noobs
data=regioni;
where regione contains 'IA';
run;
```

LOG

NOTE: There were 9 observations read from the data set WORK.REGIONI.

```
WHERE regione contains 'IA';
```

NOTE: PROCEDURE PRINT used:

```
real time          0.02 seconds
cpu time           0.02 seconds
```

OUTPUT

| rip | regione               | superf   | pop88   | maschi  |
|-----|-----------------------|----------|---------|---------|
| 1   | LOMBARDIA             | 23858.55 | 8886402 | 4287700 |
| 2   | FRIULI-VENEZIA GIULIA | 7844.13  | 1210242 | 577195  |
| 1   | LIGURIA               | 5417.97  | 1749572 | 830641  |
| 2   | EMILIA ROMAGNA        | 22123.18 | 3924199 | 1895131 |
| 3   | UMBRIA                | 8456.04  | 818226  | 398466  |
| 4   | CAMPANIA              | 13595.33 | 5731426 | 2830845 |
| 4   | PUGLIA                | 19357.25 | 4042996 | 1970864 |
| 4   | CALABRIA              | 15080.32 | 2146724 | 1060391 |
| 5   | SICILIA               | 25707.23 | 5141343 | 2512105 |

N = 9

## SELEZIONE DELLE OSSERVAZIONI CON OPERATORE LIKE

% specifica che questa posizione può essere occupata da un qualunque numero/tipo di caratteri;

```
proc print data= regioni;
where regione like 'L%A';
run;
```

log

NOTE: There were 2 observations read from the data set WORK.REGIONI.

```
WHERE regione like 'L%A';
```

NOTE: PROCEDURE PRINT used:

```
real time          0.05 seconds
cpu time           0.01 seconds
```

output

| Obs | ripartiz | regione   | superf   | pop88   | maschi  |
|-----|----------|-----------|----------|---------|---------|
| 3   | 1        | LOMBARDIA | 23858.55 | 8886402 | 4287700 |
| 7   | 1        | LIGURIA   | 5417.97  | 1749572 | 830641  |

## COMBINAZIONE DI ESPRESSIONI TRAMITE OPERATORI LOGICI

**WHERE** espr1 *operatore* espr2

| espr1 | espr2 |  | AND |  | OR |
|-------|-------|--|-----|--|----|
| v     | f     |  | f   |  | v  |
| v     | v     |  | v   |  | v  |
| f     | f     |  | f   |  | f  |

**NOT** negazione, può essere utilizzata in combinazione con ogni operatore WHERE e/o logico

## ORDINE DI ESECUZIONE WHERE CON PIÙ CONDIZIONI

- Espressioni **NOT**
- Espressioni unite da **AND**
- Espressioni unite da **OR**

L'ordine di esecuzione può essere modificato ricorrendo alle parentesi

## SELEZIONE DELLE OSSERVAZIONI: WHERE O SUBSETTING IF?

|                                | Subsetting <b>IF</b> | <b>WHERE</b>                                            |
|--------------------------------|----------------------|---------------------------------------------------------|
| <i>utilizzo</i>                | DATA step            | DATA step<br>PROC step<br>opzione nome                  |
| <i>Verifica<br/>condizione</i> | dopo lettura<br>PDV  | prima della<br>lettura nel<br>PDV                       |
| <i>Uso funzioni</i>            | sì                   | no                                                      |
| <i>operatori</i>               | espressioni<br>SAS   | espressioni<br>SAS+operatori<br>speciali (LIKE<br>etc.) |

## PROCESSI CHE RICHIEDONO NECESSARIAMENTE IF O WHERE

| <i>Se si deve ...</i>                                            | <i>usare...</i> |
|------------------------------------------------------------------|-----------------|
| Selezionare in una procedura                                     | <b>WHERE</b>    |
| Usare DATA set indicizzato                                       | WHERE           |
| Usare operatori 'speciali'                                       | <b>WHERE</b>    |
| Selezionare in base a valori calcolati o assegnati nel data step | subst IF        |
| Selezionare in un punto del DATA step che non sia l'inizio       | subst IF        |
| Eseguire una selezione condizionata                              | subs IF         |

## ISTRUZIONI E OPZIONI CHE CONTROLLANO LETTURA E SCRITTURA DI VARIABILI E OSSERVAZIONI

| Task                 | Statements    | Data set options | System options |
|----------------------|---------------|------------------|----------------|
| Control variables    | DROP          | DROP=            |                |
|                      | KEEP          | KEEP=            |                |
|                      | RENAME        | RENAME=          |                |
| Control observations | WHERE         | WHERE=           | FIRSTOBS=      |
|                      | subsetting IF | FIRSTOBS=        | OBS=           |
|                      | DELETE        | OBS=             |                |
|                      | REMOVE        |                  |                |
|                      | OUTPUT        |                  |                |

## SCELTA DELLE VARIABILI

Di norma la procedura viene eseguita su tutte le variabili presenti nel Data Set.

### *Selezione delle variabili*

- Opzioni KEEP= e DROP=;
- Istruzione VAR.

### **VAR** *elenco variabili;*

#### *Esempi*

```
(1)
title 'Selezione delle variabili';
PROC PRINT DATA=dati.regioni(keep=regione pop88);
PROC PRINT DATA=dati.regioni(drop=ripartiz maschi
superf);
PROC PRINT DATA=dati.regioni;
var regione pop88;
RUN;
```

```
(2)
PROC PRINT DATA=dati.regioni;
where ripartiz ne 2;
var regione pop88;
RUN;
```

## PONDERAZIONE DELLE OSSERVAZIONI

### *Pesi*

### **WEIGHT** *variabile;*

*variabile* contiene valori *non negativi*, non necessariamente intero. Se *variabile=0* l'obs viene esclusa dall'elaborazione.

### *Frequenze*

### **FREQ** *var;*

*var* contiene valori *interi* non negativi. Per *var=n*, l'obs corrispondente è considerata *n* volte. Se *variabile<1* l'obs viene esclusa dall'elaborazione.

## FORMATI DI SCRITTURA

Descrivono la forma di rappresentazione esterna dei dati:

- (a) formati SAS:
  - numerici;
  - carattere;
  - date e tempi.
- (b) formati definiti dall'utente.

## ASSOCIAZIONE ALLE VARIABILI DEI FORMATI DI SCRITTURA

**FORMAT** *variabili formato. ...;*

Definizione di associazione di formato:

- *permanente*: se usata nel Data step.  
Inserita nella DIA;
- *temporanea*: se usata nel Proc step.

## FORMATI SAS (SELEZIONE)

BEST<sub>w</sub> SAS sceglie notazione che restituisce la maggiore informazione sul dato, scritto nelle colonne disponibili (*w*).

```
*best.sas;  
title 'Scrittura dati in formato BESTw.';  
data pippo;  
a=16895450;  
B=A; C=A; d=a;  
run;  
title2 'formato di default (best12.), best5., best3. e best2.';  
proc print data=pippo;  
format b best5. c best3. d best2.;  
run;
```

Attenzione alla nota per best2.!!!



## FORMATI DI SCRITTURA SAS

### Formato *w.d*

Indica che il valore deve occupare un campo di *w* colonne con *d* punti decimali.

```
*wd.sas;
title 'Scrittura dati in formato w.d';
data pippo;
a=1689.5453;B=A; C=A; d=a;
run;
title2 'formato di default (8.2), 9.4, 8.0 e 6.4';
*attenzione alla nota(W.D. format 6.4 troppo piccolo);
proc print data=pippo;
format b 9.4 c 8.0 d 6.4;
run;
```

### Formato *Ew*.

Indica che il valore deve essere scritto in notazione scientifica. NB: 4 posizioni sempre occupate dall'esponente, una dal punto decimale e una dal segno.

```
*Ew.sas;
title 'Scrittura dati in formato Ew.';
data pippo;
a=16895453; B=A;
title2 'formato di default (best12.) e E9.';
proc print data=pippo;
format b E9.;
run;
```

## FORMATI DI SCRITTURA SAS

### Formati di data

DDMMYY $w$ . (*ggmmaa*), MMDDYY $w$ . (*mmggaa*), YYMMDD $w$ . (*aammgg*): mese in numeri;

DATE $w$ .: *ggmmaa* se  $w=7$ , *ggmmaaaa* se  $w=9$  (mese in lettere).

```
*data.sas;
title 'formato di scrittura delle date';
data pippo;
input data1 date9. data2 date9. a$ b 25-30;
giorni=data2-data1+1;
cards;
  1 jan 8931 dec 89 abcd 759.76
  1 jan 9031 dec 90 defg 45.71
;
run;
title2 'nb: variabili data sono considerate numeriche (giorni da 1/1/60)';
proc contents position; proc print; run;
title2 'data in formato date9., formato troppo piccolo per a, arrotondamento b';
proc print;
format data1 data2 date9. a $3. b 5.1;
run;
title2 'nb. formato data output diverso da input';
proc print;
format data1 data2 ddmmyy8. a $5. b 6.2;
run;
title2 'scrittura formati nella DIA';
data pippo;
input data1 date9. data2 date9. a$ b 25-30 ;
giorni=data2-data1+1;
format data1 data2 ddmmyy8. a $5. b 6.2;
cards;
  1 jan 8931 dec 89 abcd 759.76
  1 jan 9031 dec 90 defg 45.71
;
run;
proc contents; proc print; run;
```

## PROCEDURA FORMAT FORMATI DEFINITI DALL'UTENTE

- Decodifica delle modalità di variabili qualitative;
- Raggruppamento in classi di modalità e/o valori delle variabili.

**PROC FORMAT** *opzioni*;  
**VALUE** *nomefmt mod1='descr'*  
*mod2='descr' ...;*

...  
 (altre istruzioni value)

...  
**RUN;**

**VALUE** definisce un formato utente;  
*nomefmt* nome al più di 8 caratteri, non deve terminare con un numero;  
*mod* può essere un valore singolo, una lista di valori o un intervallo;  
*descr* stringa alfanumerica max 16 caratteri.

```
*FORMATI.SAS;
*****;
*** crea formati per dati valutazione (valuta1.sas) ***;
*****;
*libname library 'd:\dip\computaz';
*proc format library=library;      *creazione libreria permanente;
proc format fmtlib;                *creazione libreria temporanea;
  value nomec 10='Organ. internaz.'
              60='Dir.amministrat.'
              70='Statistica (A-G)'
              71='Statistica (H-Z)'
              213='Soc. (A-G)+(H-Z)'
              250='St.contemp.(A-G)'
              251='St.contemp.(H-Z)'
              270='St.dot.pol.(A-G)'
              271='St.dot.pol.(H-Z)'
              280='Is.dir.pub.(A-G)'
              281='Is.dir.pub.(H-Z)'
              290='Econ. pol. (A-G)'
              292='Econ. pol. (H-Z)'
              340='Dir.cos.it.comp.'
              350='Is. dir. privato'
              360='Sc.politica(A-G)'
              361='Sc.politica(H-Z)'
              370='Pol. econ. (A-G)'
              371='Pol. econ. (H-Z)';
  value sesso1 1='maschio' 2='femmina';
  value scprov1 1='Liceo Classico'
                2='Liceo Scientif.'
                3='Liceo Linguist.'
                4='Liceo Artistico'
                5='Ist. Magistrale'
                6='Ist. Geometri'
                7='Ist.Tec.industr'
                8='Ist.Tec.Commerc'
                9='Altre scuole it'
                10='Scuole straniere';
  value ancorsof 1='1^ anno'
                 2='2^ anno'
                 3='3^ anno'
                 4='4^ anno';
  value posizf 1='Regolare'
               2='Ripetente'
               3='Fuori corso';
  value votomat 36-40='36-40'
                41-45='41-45'
                46-50='46-50'
                51-54='51-54'
                55-59='55-59';

run;
```

# output di formati.sas (sintesi)

## opzione FMTLIB

| FORMAT NAME: SCPROVF LENGTH: 16 NUMBER OF VALUES: 10<br>MIN LENGTH: 1 MAX LENGTH: 40 DEFAULT LENGTH 16 FUZZ: STD |     |                                    |
|------------------------------------------------------------------------------------------------------------------|-----|------------------------------------|
| START                                                                                                            | END | LABEL (VER. 6.12 14MAY99:15:22:14) |
| 1                                                                                                                |     | 1 Liceo Classico                   |
| 2                                                                                                                |     | 2 Liceo Scientif.                  |
| 3                                                                                                                |     | 3 Liceo Linguist.                  |
| 4                                                                                                                |     | 4 Liceo Artistico                  |
| 5                                                                                                                |     | 5 Ist. Magistrale                  |
| 6                                                                                                                |     | 6 Ist. Geometri                    |
| 7                                                                                                                |     | 7 Ist.Tec.industr                  |
| 8                                                                                                                |     | 8 Ist.Tec.Commerc                  |
| 9                                                                                                                |     | 9 Altre scuole it                  |
| 10                                                                                                               |     | 10 Scuole straniere                |

| FORMAT NAME: VOTOMAT LENGTH: 5 NUMBER OF VALUES: 5<br>MIN LENGTH: 1 MAX LENGTH: 40 DEFAULT LENGTH 5 FUZZ: STD |     |                                    |
|---------------------------------------------------------------------------------------------------------------|-----|------------------------------------|
| START                                                                                                         | END | LABEL (VER. 6.12 14MAY99:15:22:14) |
| 36                                                                                                            | 40  | 36-40                              |
| 41                                                                                                            | 45  | 41-45                              |
| 46                                                                                                            | 50  | 46-50                              |
| 51                                                                                                            | 54  | 51-54                              |
| 55                                                                                                            | 59  | 55-59                              |

## Proc contents

CONTENTS PROCEDURE

-----Directory-----

Libref: WORK  
 Engine: V612  
 Physical Name: C:\SAS\SASWORK\#TD70549

| # | Name    | Memtype | Indexes |
|---|---------|---------|---------|
| 1 | FORMATS | CATALOG |         |

```
*formati2.sas;
*****;
*** associazione formati ai dati ***;
*****;
libname dati 'd:\dip\computaz';
title 'PROC STEP: associazione temporanea formati';
proc freq data=dati.valuta (drop=id docente dom2 dom3);
format corso nomec. dom1 sesso. dom4 scprovf. dom6 ancorsof.
      dom7 posizf. Dom5 Votomat.;
run;
title 'DATA STEP: associazione permanente formati' ;
data dati;
set dati.valuta;
format corso nomec. dom1 sesso. dom4 scprovf. dom6 ancorsof.
      dom7 posizf. Dom5 Votomat.;
run;
proc contents;
run;
proc freq data=dati(drop=id docente dom2 dom3);
run;
```

PROC STEP: associazione temporanea formati 40  
 May 14, 1999 15:42 Friday,

| CORSO            | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|------------------|-----------|---------|----------------------|--------------------|
| Statistica (A-G) | 93        | 9.4     | 93                   | 9.4                |
| Statistica (H-Z) | 181       | 18.2    | 274                  | 27.6               |
| Soc. (A-G)+(H-Z) | 99        | 10.0    | 373                  | 37.6               |
| St.contemp.(A-G) | 124       | 12.5    | 497                  | 50.1               |
| St.contemp.(H-Z) | 23        | 2.3     | 520                  | 52.4               |
| St.dot.pol.(A-G) | 26        | 2.6     | 546                  | 55.0               |
| St.dot.pol.(H-Z) | 16        | 1.6     | 562                  | 56.6               |
| Is.dir.pub.(H-Z) | 27        | 2.7     | 589                  | 59.3               |
| Econ. pol. (A-G) | 37        | 3.7     | 626                  | 63.0               |
| Econ. pol. (H-Z) | 43        | 4.3     | 669                  | 67.4               |
| Dir.cos.it.comp. | 88        | 8.9     | 757                  | 76.2               |
| Is. dir. privato | 44        | 4.4     | 801                  | 80.7               |
| Sc.politica(A-G) | 77        | 7.8     | 878                  | 88.4               |
| Sc.politica(H-Z) | 44        | 4.4     | 922                  | 92.8               |
| Pol. econ. (A-G) | 38        | 3.8     | 960                  | 96.7               |
| Pol. econ. (H-Z) | 33        | 3.3     | 993                  | 100.0              |

## CONTENTS PROCEDURE

```

Data Set Name: WORK.DATI      Observations:    993
Member Type:  DATA          Variables:       11
Engine:       V612           Indexes:         0
Created:      15:48 Friday, May 14, 1999  Observation Length: 88
Last Modified: 15:48 Friday, May 14, 1999 Deleted Observations: 0
Protection:                               Compressed:      NO
Data Set Type:                               Sorted:         NO
Label:

```

## -----Engine/Host Dependent Information-----

```

Data Set Page Size:    8192
Number of Data Set Pages: 11
File Format:           607
First Data Page:      1
Max Obs per Page:     92
Obs in First Data Page: 73

```

## -----Alphabetic List of Variables and Attributes-----

| #  | Variable | Type | Len | Pos | Format    | Label                       |
|----|----------|------|-----|-----|-----------|-----------------------------|
| 2  | CORSO    | Num  | 8   | 8   | NOME.C.   |                             |
| 3  | DOCENTE  | Num  | 8   | 16  |           |                             |
| 4  | DOM1     | Num  | 8   | 24  | SESSOF.   | Sesso                       |
| 5  | DOM2     | Num  | 8   | 32  |           | Anno di nascita             |
| 6  | DOM3     | Num  | 8   | 40  |           | Prima immatricolazione      |
| 7  | DOM4     | Num  | 8   | 48  | SCPROVF.  | Scuola di provenienza       |
| 8  | DOM5     | Num  | 8   | 56  | VOTOMAT.  | Voto riportato al Diploma   |
| 9  | DOM6     | Num  | 8   | 64  | ANCORSOF. | Anno di corso               |
| 10 | DOM7     | Num  | 8   | 72  | POSIZF.   | Posizione                   |
| 11 | DOM8     | Num  | 8   | 80  |           | Corsi frequentati nell'anno |
| 1  | ID       | Num  | 8   | 0   |           |                             |

## TABELLE: PROCEDURA FREQ

**PROC FREQ** *opzioni*;  
**TABLES** *espressione* [*opzioni*];

...

(*altre istruzioni*)

....

**RUN**;

*espressione* definisce la tabella

- tabelle univariate

**TABLES** var1 [var2]... [*opzioni*];

- tabelle bivariate

**TABLES** v1\*v2[\*v3]... [*opzioni*];

*opzioni di TABLES (selezione)*

**NOFREQ** niente freq assolute

**NOPERCENT** niente freq percentuali

**NOCUM** niente freq cumulate

**NOROW** no freq. perc. su tot di riga

**NOCOL** no freq. perc. su tot di col.

```

*freq1.sas;
*****;
*** distribuzioni di frequenza ***;
*****;
libname dati 'v:\didat\carla\sas';
proc freq data=dati.turismo;
*tabella semplice (distrib. di freq.
univariata);
tables studio;
*tabella doppia (distrib. di freq.
bivariata);
tables sesso*sci;
RUN;

```

The FREQ Procedure

| titolo di studio |           |         |                         |                       |
|------------------|-----------|---------|-------------------------|-----------------------|
| STUDIO           | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
| 1                | 2         | 0.21    | 2                       | 0.21                  |
| 2                | 32        | 3.39    | 34                      | 3.60                  |
| 3                | 174       | 18.43   | 208                     | 22.03                 |
| 4                | 500       | 52.97   | 708                     | 75.00                 |
| 5                | 236       | 25.00   | 944                     | 100.00                |

| Table of SESSO by SCI |                                |                              |                              |                                |               |
|-----------------------|--------------------------------|------------------------------|------------------------------|--------------------------------|---------------|
| SESSO                 | SCI(pratica dello sci)         |                              |                              |                                |               |
| Frequency             | 1                              | 2                            | 3                            | 4                              | Total         |
| Percent               |                                |                              |                              |                                |               |
| Row Pct               |                                |                              |                              |                                |               |
| Col Pct               |                                |                              |                              |                                |               |
| F                     | 252<br>26.98<br>54.31<br>46.67 | 40<br>4.28<br>8.62<br>45.45  | 64<br>6.85<br>13.79<br>47.76 | 108<br>11.56<br>23.28<br>62.79 | 464<br>49.68  |
| M                     | 288<br>30.84<br>61.28<br>53.33 | 48<br>5.14<br>10.21<br>54.55 | 70<br>7.49<br>14.89<br>52.24 | 64<br>6.85<br>13.62<br>37.21   | 470<br>50.32  |
| Total                 | 540<br>57.82                   | 88<br>9.42                   | 134<br>14.35                 | 172<br>18.42                   | 934<br>100.00 |

FREQUENCY MISSING = 10

```

*freq2.sas;
*****;
*** distribuzioni di frequenza ***;
*** associazione di formati e titoli **;
*****;
libname dati 'v:\didat\carla\sas';
libname library 'v:\didat\carla\sas';
title1 'Indagine sul turismo montano';
proc freq data=dati.turismo;
tables studio eta;
run;

```

The FREQ Procedure

| titolo di studio |           |         |                         |                       |
|------------------|-----------|---------|-------------------------|-----------------------|
| STUDIO           | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
| 1                | 2         | 0.21    | 2                       | 0.21                  |
| 2                | 32        | 3.39    | 34                      | 3.60                  |
| 3                | 174       | 18.43   | 208                     | 22.03                 |
| 4                | 500       | 52.97   | 708                     | 75.00                 |
| 5                | 236       | 25.00   | 944                     | 100.00                |

The FREQ Procedure

| TA  | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-----|-----------|---------|-------------------------|-----------------------|
| 5   | 2         | 0.21    | 2                       | 0.21                  |
| 14  | 2         | 0.21    | 4                       | 0.42                  |
| 15  | 6         | 0.64    | 10                      | 1.06                  |
| 16  | 18        | 1.91    | 28                      | 2.97                  |
| 17  | 16        | 1.69    | 44                      | 4.66                  |
| 18  | 24        | 2.54    | 68                      | 7.20                  |
| 19  | 24        | 2.54    | 92                      | 9.75                  |
| 20  | 32        | 3.39    | 124                     | 13.14                 |
| 21  | 20        | 2.12    | 144                     | 15.25                 |
| 22  | 28        | 2.97    | 172                     | 18.22                 |
| 23  | 20        | 2.12    | 192                     | 20.34                 |
| 24  | 24        | 2.54    | 216                     | 22.88                 |
| 25  | 34        | 3.60    | 250                     | 26.48                 |
| 26  | 24        | 2.54    | 274                     | 29.03                 |
| 27  | 26        | 2.75    | 300                     | 31.78                 |
| 28  | 30        | 3.18    | 330                     | 34.96                 |
| 29  | 30        | 3.18    | 360                     | 38.14                 |
| ... |           |         |                         |                       |
| 66  | 2         | 0.21    | 938                     | 99.36                 |
| 72  | 2         | 0.21    | 940                     | 99.58                 |
| 73  | 4         | 0.42    | 944                     | 100.00                |

```

title2 'decodifica e raggruppamento in
classi delle variabili';
proc freq data=dati.turismo;
tables studio eta;
format studio istrfmt. eta etafmt.;
run;

```

titolo di studio

| STUDIO          | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-----------------|-----------|---------|-------------------------|-----------------------|
| nessun titolo   | 2         | 0.21    | 2                       | 0.21                  |
| lic. elementare | 32        | 3.39    | 34                      | 3.60                  |
| licenza media   | 174       | 18.43   | 208                     | 22.03                 |
| diploma         | 500       | 52.97   | 708                     | 75.00                 |
| laurea          | 236       | 25.00   | 944                     | 100.00                |

| ETA        | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|------------|-----------|---------|-------------------------|-----------------------|
| 0 - 13     | 2         | 0.21    | 2                       | 0.21                  |
| 14 - 29    | 358       | 37.92   | 360                     | 38.14                 |
| 30-59      | 560       | 59.32   | 920                     | 97.46                 |
| 60 E OLTRE | 24        | 2.54    | 944                     | 100.00                |

```

*freq3.sas;
*****;
*** distribuzioni di frequenza ***;
*** opzioni istruzione tables ***;
*****;
title1 'Indagine sul turismo montano';
title2 'tabelle di default';
proc freq data=dati.turismo;
tables sesso sci*sesso;
run;

```

The FREQ Procedure

| SESSO | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-------|-----------|---------|-------------------------|-----------------------|
| F     | 470       | 49.79   | 470                     | 49.79                 |
| M     | 474       | 50.21   | 944                     | 100.00                |

Table of SCI by SESSO  
SCI(pratica dello sci)

|           |                                | SESSO                          |               | Total |
|-----------|--------------------------------|--------------------------------|---------------|-------|
| Frequency | Percent                        | F                              | M             |       |
| Row Pct   |                                |                                |               |       |
| Col Pct   |                                |                                |               |       |
| 1         | 252<br>26.98<br>46.67<br>54.31 | 288<br>30.84<br>53.33<br>61.28 | 540<br>57.82  |       |
| 2         | 40<br>4.28<br>45.45<br>8.62    | 48<br>5.14<br>54.55<br>10.21   | 88<br>9.42    |       |
| 3         | 64<br>6.85<br>47.76<br>13.79   | 70<br>7.49<br>52.24<br>14.89   | 134<br>14.35  |       |
| 4         | 108<br>11.56<br>62.79<br>23.28 | 64<br>6.85<br>37.21<br>13.62   | 172<br>18.42  |       |
| Total     | 464<br>49.68                   | 470<br>50.32                   | 934<br>100.00 |       |

Frequency Missing = 10

```

title2 'opzioni istruzione TABLES';
title3 'opzione nocum';
proc freq data=dati.turismo;
tables sesso/nocum;
run;

```

The FREQ Procedure

| SESSO | Frequency | Percent |
|-------|-----------|---------|
| F     | 470       | 49.79   |
| M     | 474       | 50.21   |

```

title3 'opzione nopercent';
proc freq data=dati.turismo;
*esclude percentuali;
tables sesso sci*sesso/nopercent;
RUN;

```

The FREQ Procedure

| SESSO | Frequency | Cumulative Frequency |
|-------|-----------|----------------------|
| F     | 470       | 470                  |
| M     | 474       | 944                  |

Table of SCI by SESSO  
SCI(pratica dello sci)

| SESSO   |         | Frequency |       | Total |
|---------|---------|-----------|-------|-------|
| Row Pct | Col Pct | F         | M     |       |
| 1       |         | 252       | 288   | 540   |
|         |         | 46.67     | 53.33 |       |
|         |         | 54.31     | 61.28 |       |
| 2       |         | 40        | 48    | 88    |
|         |         | 45.45     | 54.55 |       |
|         |         | 8.62      | 10.21 |       |
| 3       |         | 64        | 70    | 134   |
|         |         | 47.76     | 52.24 |       |
|         |         | 13.79     | 14.89 |       |
| 4       |         | 108       | 64    | 172   |
|         |         | 62.79     | 37.21 |       |
|         |         | 23.28     | 13.62 |       |
| Total   |         | 464       | 470   | 934   |

FREQUENCY MISSING = 10

```

title3 'opzioni nocum e nopercent';
proc freq data=dati.turismo;
tables sesso sci*sesso/nocum nopercent;
RUN;

```

opzioni istruzione TABLES  
opzioni nocum e nopercent

The FREQ Procedure

| SESSO | Frequency |
|-------|-----------|
| F     | 470       |
| M     | 474       |

Table of SCI by SESSO

SCI(pratica dello sci)  
SESSO

| Frequency<br>Row Pct<br>Col Pct | SESSO |       | Total |
|---------------------------------|-------|-------|-------|
|                                 | F     | M     |       |
| 1                               | 252   | 288   | 540   |
|                                 | 46.67 | 53.33 |       |
|                                 | 54.31 | 61.28 |       |
| 2                               | 40    | 48    | 88    |
|                                 | 45.45 | 54.55 |       |
|                                 | 8.62  | 10.21 |       |
| 3                               | 64    | 70    | 134   |
|                                 | 47.76 | 52.24 |       |
|                                 | 13.79 | 14.89 |       |
| 4                               | 108   | 64    | 172   |
|                                 | 62.79 | 37.21 |       |
|                                 | 23.28 | 13.62 |       |
| Total                           | 464   | 470   | 934   |

FREQUENCY MISSING = 10

```

title3 'opzioni norow e nocol';
proc freq data=dati.turismo;
*esclude percentuali di riga;
tables sci*sex/norow;
*esclude percentuali di colonna;
tables sci*sex/nocol;
run;

```

opzione norow  
SCI(pratica dello sci)  
SESSO

| Frequency<br>Percent<br>Col Pct | SESSO                 |                       | Total         |
|---------------------------------|-----------------------|-----------------------|---------------|
|                                 | F                     | M                     |               |
| 1                               | 252<br>26.98<br>54.31 | 288<br>30.84<br>61.28 | 540<br>57.82  |
| 2                               | 40<br>4.28<br>8.62    | 48<br>5.14<br>10.21   | 88<br>9.42    |
| 3                               | 64<br>6.85<br>13.79   | 70<br>7.49<br>14.89   | 134<br>14.35  |
| 4                               | 108<br>11.56<br>23.28 | 64<br>6.85<br>13.62   | 172<br>18.42  |
| Total                           | 464<br>49.68          | 470<br>50.32          | 934<br>100.00 |

opzioni norow e nocol  
SCI(pratica dello sci)  
SESSO

| Frequency<br>Percent | SESSO        |              | Total         |
|----------------------|--------------|--------------|---------------|
|                      | F            | M            |               |
| 1                    | 252<br>26.98 | 288<br>30.84 | 540<br>57.82  |
| 2                    | 40<br>4.28   | 48<br>5.14   | 88<br>9.42    |
| 3                    | 64<br>6.85   | 70<br>7.49   | 134<br>14.35  |
| 4                    | 108<br>11.56 | 64<br>6.85   | 172<br>18.42  |
| Total                | 464<br>49.68 | 470<br>50.32 | 934<br>100.00 |

FREQUENCY MISSING = 10

```

*freq4.sas;
*****;
*** distribuzioni di frequenza ***;
*** trattamento dei dati mancanti ***;
*****;
title1 'Indagine sul turismo montano';
title2 'tabella di default: esclusione
dei dati mancanti';
proc freq data=dati.turismo;
tables volte;
run;

```

Indagine sul turismo montano  
tabella di default: esclusione dei dati mancanti

The FREQ Procedure

| VOLTE | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-------|-----------|---------|-------------------------|-----------------------|
| 1     | 18        | 2.91    | 18                      | 2.91                  |
| 2     | 86        | 13.92   | 104                     | 16.83                 |
| 3     | 114       | 18.45   | 218                     | 35.28                 |
| 4     | 50        | 8.09    | 268                     | 43.37                 |
| 5     | 294       | 47.57   | 562                     | 90.94                 |
| 6     | 14        | 2.27    | 576                     | 93.20                 |
| 7     | 4         | 0.65    | 580                     | 93.85                 |
| 8     | 8         | 1.29    | 588                     | 95.15                 |
| 9     | 30        | 4.85    | 618                     | 100.00                |

FREQUENCY MISSING = 326



```

title2 'opzione MISSPRINT: stampa
missing, senza entrare nelle
percentuali';
proc freq data=dati.turismo;
tables volte/ missprint;
RUN;

```

Indagine sul turismo montano  
opzione MISSPRINT: stampa missing, senza entrare nelle percentuali

The FREQ Procedure

| VOLTE | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-------|-----------|---------|-------------------------|-----------------------|
| .     | 96        | .       | .                       | .                     |
| A     | 230       | .       | .                       | .                     |
| 1     | 18        | 2.91    | 18                      | 2.91                  |
| 2     | 86        | 13.92   | 104                     | 16.83                 |
| 3     | 114       | 18.45   | 218                     | 35.28                 |
| 4     | 50        | 8.09    | 268                     | 43.37                 |
| 5     | 294       | 47.57   | 562                     | 90.94                 |
| 6     | 14        | 2.27    | 576                     | 93.20                 |
| 7     | 4         | 0.65    | 580                     | 93.85                 |
| 8     | 8         | 1.29    | 588                     | 95.15                 |
| 9     | 30        | 4.85    | 618                     | 100.00                |

FREQUENCY MISSING = 326

```

title2 'opzione MISSING: missing
considerati validi';
proc freq data=dati.turismo;
tables volte/ missing;
RUN;

```

Indagine sul turismo montano  
opzione MISSING: missing considerati validi

The FREQ Procedure

| VOLTE | Frequency | Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-------|-----------|---------|-------------------------|-----------------------|
| .     | 96        | 10.17   | 96                      | 10.17                 |
| A     | 230       | 24.36   | 326                     | 34.53                 |
| 1     | 18        | 1.91    | 344                     | 36.44                 |
| 2     | 86        | 9.11    | 430                     | 45.55                 |
| 3     | 114       | 12.08   | 544                     | 57.63                 |
| 4     | 50        | 5.30    | 594                     | 62.92                 |
| 5     | 294       | 31.14   | 888                     | 94.07                 |
| 6     | 14        | 1.48    | 902                     | 95.55                 |
| 7     | 4         | 0.42    | 906                     | 95.97                 |
| 8     | 8         | 0.85    | 914                     | 96.82                 |
| 9     | 30        | 3.18    | 944                     | 100.00                |

## PROC FREQ: NOTAZIONE E DEFINIZIONE DEGLI INDICI STATISTICI

|                                                                        |                                    |
|------------------------------------------------------------------------|------------------------------------|
| $n_{i.} = \sum_j n_{ij}$                                               | (row totals)                       |
| $n_{.j} = \sum_i n_{ij}$                                               | (column totals)                    |
| $n = \sum_i \sum_j n_{ij}$                                             | (overall total)                    |
| $p_{ij} = n_{ij}/n$                                                    | (cell percentages)                 |
| $p_{i.} = n_{i.}/n$                                                    | (row percentages)                  |
| $p_{.j} = n_{.j}/n$                                                    | (column percentages)               |
| $R_i = \text{score for row } i$                                        |                                    |
| $C_j = \text{score for column } j$                                     |                                    |
| $\bar{R} = \sum_i n_{i.} R_i / n$                                      | (average row score)                |
| $A_{ij} = \sum_{k>i} \sum_{l>j} n_{kl} + \sum_{k<i} \sum_{l<j} n_{kl}$ |                                    |
| $\bar{C} = \sum_j n_{.j} C_j / n$                                      | (average column score)             |
| $A_{ij} = \sum_{k>i} \sum_{l>j} n_{kl} + \sum_{k<i} \sum_{l<j} n_{kl}$ |                                    |
| $D_{ij} = \sum_{k>i} \sum_{l<j} n_{kl} + \sum_{k<i} \sum_{l>j} n_{kl}$ |                                    |
| $P = \sum_i \sum_j n_{ij} A_{ij}$                                      | (twice the number of concordances) |
| $Q = \sum_i \sum_j n_{ij} D_{ij}$                                      | (twice the number of discordances) |

## PROC FREQ: OPZIONE CHISQ

### TABLES $v1*v2$ /CHISQ;

Calcola misure di associazione per tabelle bivariate basate sull'indice chi-quadrato (Chi-Square):

$$Q_P = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad e_{ij} = \frac{n_{i.} n_{.j}}{n}$$

Coefficiente  $\phi$  (Phi Coefficient):

$$\phi = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}} \quad \text{for } 2 \times 2 \text{ tables}$$

$$\phi = \sqrt{Q_P/n} \quad \text{otherwise.}$$

Indice di Cramer (Cramer's V):

$$V = \phi \quad \text{for } 2 \times 2 \text{ tables}$$

$$V = \sqrt{\frac{Q_P/n}{\min(R-1, C-1)}} \quad \text{otherwise.}$$

## PROC FREQ: ESEMPIO OPZIONE CHISQ

```
*freq9.sas;
libname dati 'v:\didat\carla\sas';
proc freq data=dati.turismo;
tables sci*sexo/CHISQ nopercnt norow nocol ;
format sesso $1. sci 4.;
run;
```

SCI(pratica dello sci)

| Frequency | SESSO |     | Total |
|-----------|-------|-----|-------|
|           | F     | M   |       |
| 1         | 252   | 288 | 540   |
| 2         | 40    | 48  | 88    |
| 3         | 64    | 70  | 134   |
| 4         | 108   | 64  | 172   |
| Total     | 464   | 470 | 934   |

Frequency Missing = 10

Statistics for Table of SCI by SESSO

| Statistic                   | DF | Value   | Prob   |
|-----------------------------|----|---------|--------|
| Chi-Square                  | 3  | 14.6138 | 0.0022 |
| Likelihood Ratio Chi-Square | 3  | 14.7422 | 0.0021 |
| Mantel-Haenszel Chi-Square  | 1  | 10.1746 | 0.0014 |
| Phi Coefficient             |    | 0.1251  |        |
| Contingency Coefficient     |    | 0.1241  |        |
| Cramer's V                  |    | 0.1251  |        |

Effective Sample Size = 934

Frequency Missing = 10

## PROC FREQ: OPZIONE MEASURES

### TABLES v1\*v2 /MEASURES;

Calcola misure di associazione per tabelle bivariate. Gli indici seguenti sono adatti per variabili ordinali.

#### Tau-b di Kendall:

$$t_b = \frac{(P - Q)}{\sqrt{w_r w_c}}$$

$$w_r = n^2 - \sum_i n_i^2$$

$$w_c = n^2 - \sum_j n_j^2$$

P coppie concordanti, Q coppie discordanti

#### Coefficiente di correlazione di Pearson:

$$r = \frac{U}{W} = \frac{SS_{rc}}{\sqrt{SS_r SS_c}}$$

$$SS_r = \sum_i \sum_j n_{ij} (R_i - \bar{R})^2$$

$$SS_c = \sum_i \sum_j n_{ij} (C_j - \bar{C})^2$$

$$SS_{rc} = \sum_i \sum_j n_{ij} (R_i - \bar{R}) (C_j - \bar{C})$$

I punteggi di riga  $R_i$  e di colonna  $C_j$  vengono calcolati automaticamente da SAS: sono pari alle modalità della var se questa ha modalità numeriche, alla riga o alla colonna corrispondente se la var ha modalità carattere.

## Coefficiente di correlazione di Spearman:

i ranghi  $R_i$  e  $C_j$  vengono calcolati automaticamente da SAS  
(per modifiche si veda l'opzione SCORE= di TABLES)

$$r_s = \frac{U}{W}$$

con:

$$v = \sum_i \sum_j n_{ij} R(i) C(j)$$

$$w = \frac{1}{12} \sqrt{FG}$$

$$F = n^3 - \sum_i n_i^3$$

$$G = n^3 - \sum_j n_j^3$$

$$R(i) = RI_i - \frac{n}{2}$$

$$C(j) = CI_j - \frac{n}{2}$$

Row scores :

$$RI_i = \sum_{k < i} n_{k.} + (n_{i.} + 1) / 2 \quad i = 1, 2, \dots, R$$

Column scores :

$$CI_j = \sum_{k < j} n_{.k} + (n_{.j} + 1) / 2 \quad j = 1, 2, \dots, C$$

## PROC FREQ: ESEMPIO OPZIONE MEASURES

```
PROC FORMAT ;
VALUE VOLTE 1='1' 2-5='2-5' 6-HIGH='6 e oltre';
value studio 1,2,3='1 obbligo' 4='2 diploma' 5='3
laurea';
proc freq data=dati.turismo;
tables VOLTE*STUDIO/MEASURES nopercnt norow nocol
;
format VOLTE volte. STUDIO studio.;
run;
```

The FREQ Procedure

Table of VOLTE by STUDIO

| VOLTE     | STUDIO(titolo di studio) |               |          | Total |
|-----------|--------------------------|---------------|----------|-------|
|           | 1 obblig<br>o            | 2 diplom<br>a | 3 laurea |       |
| 1         | 2                        | 14            | 2        | 18    |
| 2-5       | 126                      | 276           | 142      | 544   |
| 6 e oltre | 20                       | 20            | 16       | 56    |
| Total     | 148                      | 310           | 160      | 618   |

Frequency Missing = 326

Statistics for Table of VOLTE by STUDIO

| Statistic                         | Value   | ASE    |
|-----------------------------------|---------|--------|
| Gamma                             | -0.0785 | 0.1039 |
| Kendall's Tau-b                   | -0.0293 | 0.0389 |
| Stuart's Tau-c                    | -0.0161 | 0.0215 |
| Somers' D C R                     | -0.0498 | 0.0661 |
| Somers' D R C                     | -0.0173 | 0.0229 |
| Pearson Correlation               | -0.0721 | 0.0451 |
| Spearman Correlation              | -0.0315 | 0.0416 |
| Lambda Asymmetric C R             | 0.0000  | 0.0000 |
| Lambda Asymmetric R C             | 0.0000  | 0.0000 |
| Lambda Symmetric                  | 0.0000  | 0.0000 |
| Uncertainty Coefficient C R       | 0.0090  | 0.0052 |
| Uncertainty Coefficient R C       | 0.0216  | 0.0122 |
| Uncertainty Coefficient Symmetric | 0.0127  | 0.0073 |

Effective Sample Size = 618

Frequency Missing = 326

WARNING: 35% OF THE DATA ARE MISSING.

## PROC FREQ: ESEMPIO CREAZIONE DI TABELLE A N DIMENSIONI

```
*freq5.sas;
*****;
***      distribuzioni di frequenza      ***;
***      creazione di tabelle a n dimensioni      ***;
*****;
title1 'Indagine sul turismo montano';
title2 'tabella a 3 dimensioni specificata in
TABLES';
*memoria RAM occupata pari a 2*4*4=32 celle*8
bytes=256 bytes;
proc freq data=dati.turismo;
tables sesso*studio*sci/nopercent norow nocol;
format studio istrfmt.;
run;
```

The FREQ Procedure

Table 1 of STUDIO by SCI

Controlling for SESSO=Femmina

| STUDIO(titolo di studio)<br>Frequency | SCI(pratica dello sci) |                |          |         | Total |
|---------------------------------------|------------------------|----------------|----------|---------|-------|
|                                       | SOLO ALP<br>INO        | SOLO FON<br>DO | ENTRAMBI | NESSUNO |       |
| nessun titolo                         | 0                      | 0              | 2        | 0       | 2     |
| lic. elementare                       | 8                      | 0              | 0        | 7       | 15    |
| licenza media                         | 54                     | 10             | 12       | 27      | 103   |
| diploma                               | 132                    | 20             | 34       | 58      | 244   |
| laurea                                | 58                     | 10             | 16       | 16      | 100   |
| Total                                 | 252                    | 40             | 64       | 108     | 464   |

Frequency Missing = 6

```

title2 'tabella a 3 dimensioni con istruzione BY';
*memoria RAM occupata pari a 4*4=16 celle*8
bytes=128 bytes;
proc sort data=dati.turismo;
by sesso;
proc freq data=dati.turismo;
tables studio*sci/nopercent norow nocol;
by sesso;
format studio istrfmt.;
run;

```

Indagine sul turismo montano  
tabella a 3 dimensioni con istruzione BY

SESSO=Femmina

The FREQ Procedure

Table of STUDIO by SCI

| STUDIO(titolo di studio) | SCI(pratica dello sci) |                |          |         | Total |
|--------------------------|------------------------|----------------|----------|---------|-------|
| Frequency                | SOLO ALP<br>INO        | SOLO FON<br>DO | ENTRAMBI | NESSUNO |       |
| nessun titolo            | 0                      | 0              | 2        | 0       | 2     |
| lic. elementare          | 8                      | 0              | 0        | 7       | 15    |
| licenza media            | 54                     | 10             | 12       | 27      | 103   |
| diploma                  | 132                    | 20             | 34       | 58      | 244   |
| laurea                   | 58                     | 10             | 16       | 16      | 100   |
| Total                    | 252                    | 40             | 64       | 108     | 464   |

FREQUENCY MISSING = 6

## PROC FREQ: PONDERAZIONE DELLE OSSERVAZIONI

```

*freq8.sas;
*****;
*** utilizzo procedura FREQ per archivi di frequenze ***;
*****;
title1 'Popolazione in eta' 14 anni e oltre per
condizione';
title2 'utilizzo dell''istruzione WEIGHT';
footnote 'Fonte: ISTAT - Rilevazione Forze Lavoro,
media 1985';

proc format;
value xfmt 1='14-29' 2='30-70' 3='71 e oltre';
value yfmt 1='Occupati' 2='In cerca di occ.'
3='non FF.LL.';
run;

data ffl1;
input x y num;
label x='classe di eta'' y='condizione'
num='frequenze (migliaia)';
cards;
1 1 5280
1 2 1864
1 3 6321
2 1 20610
2 2 2470
2 3 18723
3 1 132
3 2 1
3 3 4643
;
RUN;

```

```
*programma corretto;
proc freq data=ff11;
weight num;
table x*y/chisq norow nocol;
format x xfmt. y yfmt.;
run;
```

| x(classe di eta') | y(condizione)  |                  |                | Total           |
|-------------------|----------------|------------------|----------------|-----------------|
| Frequency         | Occupati       | In cerca di occ. | non FF.L.      |                 |
| Percent           |                |                  |                |                 |
| 14-29             | 5280<br>8.79   | 1864<br>3.10     | 6321<br>10.53  | 13465<br>22.43  |
| 30-70             | 20610<br>34.32 | 2470<br>4.11     | 18723<br>31.18 | 41803<br>69.62  |
| 71 e oltre        | 132<br>0.22    | 1<br>0.00        | 4643<br>7.73   | 4776<br>7.95    |
| Total             | 26022<br>43.34 | 4335<br>7.22     | 29687<br>49.44 | 60044<br>100.00 |

Statistics for Table of x by y

| Statistic                   | DF | Value     | Prob   |
|-----------------------------|----|-----------|--------|
| Chi-Square                  | 4  | 5875.5533 | <.0001 |
| Likelihood Ratio Chi-Square | 4  | 6878.9383 | <.0001 |
| Mantel-Haenszel Chi-Square  | 1  | 1019.6557 | <.0001 |
| Phi Coefficient             |    | 0.3128    |        |
| Contingency Coefficient     |    | 0.2986    |        |
| Cramer's V                  |    | 0.2212    |        |

Sample Size = 60044

```
*programma errato;
title2 'programma errato (manca WEIGHT)';
proc freq data=ff11;
table x*y/chisq norow nocol;
format x xfmt. y yfmt.;
run;
```

| x(classe di eta') | y(condizione) |                  |            | Total       |
|-------------------|---------------|------------------|------------|-------------|
| Frequency         | Occupati      | In cerca di occ. | non FF.L.  |             |
| Percent           |               |                  |            |             |
| 14-29             | 1<br>11.11    | 1<br>11.11       | 1<br>11.11 | 3<br>33.33  |
| 30-70             | 1<br>11.11    | 1<br>11.11       | 1<br>11.11 | 3<br>33.33  |
| 71 e oltre        | 1<br>11.11    | 1<br>11.11       | 1<br>11.11 | 3<br>33.33  |
| Total             | 3<br>33.33    | 3<br>33.33       | 3<br>33.33 | 9<br>100.00 |

Statistics for Table of x by y

| Statistic                   | DF | Value  | Prob   |
|-----------------------------|----|--------|--------|
| Chi-Square                  | 4  | 0.0000 | 1.0000 |
| Likelihood Ratio Chi-Square | 4  | 0.0000 | 1.0000 |
| Mantel-Haenszel Chi-Square  | 1  | 0.0000 | 1.0000 |
| Phi Coefficient             |    | 0.0000 |        |
| Contingency Coefficient     |    | 0.0000 |        |
| Cramer's V                  |    | 0.0000 |        |

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 9

### *Procedure Statistiche elementari per tipo*

| Per fare ...            | Utilizza la PROC .. | Che calcola ...                                                                                                                 |
|-------------------------|---------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Statistiche Descrittive | CORR                | statistiche descrittive univariate                                                                                              |
|                         | MEANS o SUMMARY     | statistiche descrittive univariate; stampa output e crea data set. Per default, PROC MEANS stampa e PROC SUMMARY crea data set. |
|                         | REPORT              | Simile a PROC TABULATE                                                                                                          |
|                         | SQL                 | statistiche descrittive per dati in uno o più DBMS; può stampare o creare un SAS data set.                                      |
|                         | TABULATE            | tabelle che contengono statistiche descrittive; può creare un sas data set.                                                     |
|                         | UNIVARIATE          | Il più ampio numero di statistiche descrittive; può creare un sas data set.                                                     |

### *Procedure Statistiche elementari per tipo*

| Per fare ...                         | Utilizza la PROC .. | Che calcola ...                                                                                                                                                                                        |
|--------------------------------------|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Tabelle di frequenza e altre tabelle | FREQ                | Tabelle univariate o a n-entrate; riporta le frequenze; indici e test di associazione per tabelle a doppia entrata o stratificate; può calcolare test esatti e asintotici; può creare un sas data set. |
|                                      | TABULATE            | tabelle univariate e; può creare un sas data set.                                                                                                                                                      |
|                                      | UNIVARIATE          | tabelle univariate                                                                                                                                                                                     |
| Analisi di correlazione              | CORR                | Correlazione e correlazione parziale di Pearson, di Spearman e di Kendall, D di Hoeffding e alpha di Cronbach.                                                                                         |
| Analisi distribuzioni                | UNIVARIATE          | tests sulla media e test di normalità della distribuzione.                                                                                                                                             |
|                                      | FREQ                | test per la binomiale e test di adattamento in tabelle univariate; test chi-quadrato per tabelle bivariate.                                                                                            |



## Procedure Statistiche elementari per tipo

| Per fare ...                      | Utilizza la PROC .. | Che calcola ...                                                                                                                                                            |
|-----------------------------------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ranghi Standardizzazione dei dati | RANK                | ranghi per una o più variabili numeriche in base ai valori registrati nel SAS data set, crea un sas data set; può calcolare punteggi normali o altri ranghi.               |
|                                   | STANDARD            | crea un sas data set che contiene le variabili standardizzate per una data media e deviazione standard.                                                                    |
| Grafici a bassa risoluzione *     |                     |                                                                                                                                                                            |
|                                   | CHART               | Grafici che mostrano una delle statistiche seguenti per la variabile di interesse: frequenze, percentuali, frequenze cumulate, percentuali cumulate, totali o valori medi. |
|                                   | UNIVARIATE          | grafici descrittivi tipo: stem and leaf, box plot e normal probability plot.                                                                                               |

\* Per produrre grafici a alta risoluzione, utilizzare il Software SAS/GRAPH.

## SINTESI DI DISTRIBUZIONI SEMPLICI

### Indici statistici

| Chiave   | Indice                            | MEANS | UNIVARIATE | TABULATE |
|----------|-----------------------------------|-------|------------|----------|
| CSS      | Devianza                          | X     | X          | X        |
| CV       | Coeff. variazione                 | X     | X          | X        |
| KURTOSIS | Curtosi                           | X     | X          |          |
| MAX      | Valore massimo                    | X     | X          | X        |
| MEAN     | Media aritmetica                  | X     | X          | X        |
| MEDIAN   | Mediana                           | X     | X          |          |
| MIN      | Valore minimo                     | X     | X          | X        |
| MODE     | Moda                              |       | X          |          |
| N        | Num. casi validi                  | X     | X          | X        |
| NMISS    | Num. casi mancanti                | X     | X          | X        |
| PRT      | Significatività test <i>t</i>     | X     | X          | X        |
| QUARTILE | 1° e 3° quartile                  | X     | X          | X        |
| RANGE    | MIN-MAX                           | X     | X          | X        |
| STD      | Deviazione standard               | X     | X          | X        |
| SKEWNESS | Indice di simmetria               | X     | X          | X        |
| STDERR   | Errore standard media             | X     | X          | X        |
| SUM      | Somma dei valori                  | X     | X          | X        |
| SUMWGT   | Somma dei pesi                    | X     | X          | X        |
| T        | Test <i>t</i> per $H_0$ : media=0 | X     | X          | X        |
| USS      | Somma valori al quadr.            | X     | X          | X        |
| VAR      | Varianza                          | X     | X          | X        |

## PROCEDURA MEANS

PROC MEANS [*opzioni*];  
CLASS *variabile*;  
BY *variabile*;  
VAR *elenco variabili*;  
FREQ *variabile*;  
WEIGHT *variabile*;  
ID *variabile*;  
OUTPUT OUT=*archivio*  
    [*chiave=lista nomi*]  
    [*chiave(lista var)=lista nomi*];  
RUN;

```
*MEANS6.SAS;
*****;
***  proc means: ponderazione degli indici ***;
***                istruzione WEIGHT          ***;
*****;
title1 'Procedura MEANS: ponderazione degli
indici';
title2 'Istruzione WEIGHT';

data dati;
input tipo $10. prezzi quantita;
pq=prezzi*quantita; *valore;
cards;
arance      23695 5528
mandarini   31750 1591
limoni      43280 2235
;
run;
title3 'stampa valori';
proc print;
sum prezzi quantita pq;
run;
title3 'media aritmetica semplice dei prezzi';
proc means sum sumwgt mean;
var prezzi;
run;
title3 'media aritmetica ponderata dei prezzi per
le quantita'';
proc means sum sumwgt mean;
var prezzi;
weight quantita;
run;
```

Procedura MEANS: ponderazione degli indici  
 Istruzione WEIGHT  
 stampa valori

| Obs | tipo      | prezzi | quantita | pq        |
|-----|-----------|--------|----------|-----------|
| 1   | arance    | 23695  | 5528     | 130985960 |
| 2   | mandarini | 31750  | 1591     | 50514250  |
| 3   | limoni    | 43280  | 2235     | 96730800  |
|     |           | =====  | =====    | =====     |
|     |           | 98725  | 9354     | 278231010 |

media aritmetica semplice dei prezzi

The MEANS Procedure

Analysis Variable : prezzi

| Sum      | Sum Wgts  | Mean     |
|----------|-----------|----------|
| 98725.00 | 3.0000000 | 32908.33 |

media aritmetica ponderata dei prezzi per le quantita'

The MEANS Procedure

Analysis Variable : prezzi

| Sum       | Sum Wgts | Mean     |
|-----------|----------|----------|
| 278231010 | 9354.00  | 29744.60 |

```
*MEANS7.SAS;
*****;
*** proc means: analisi per gruppi di obs omogenee ***;
***          istruzione CLASS          ***;
*****;
title1 'Procedura MEANS: analisi per gruppi di obs
omogenee';
title2 'Istruzione CLASS';

proc means data=dati.turismo sum mean std;
var eta;
class sesso;
run;
```

Procedura MEANS: analisi per gruppi di obs omogenee  
 Istruzione CLASS

The MEANS Procedure

Analysis Variable : ETA

| SESSO | N   |  | Sum      | Mean       | Std Dev    |
|-------|-----|--|----------|------------|------------|
|       | Obs |  |          |            |            |
| F     | 470 |  | 16202.00 | 34.4723404 | 11.8355055 |
| M     | 474 |  | 16252.00 | 34.2869198 | 11.4965735 |

```

*MEANS8.SAS;
*****;
***  proc means: analisi per gruppi di obs omogenee ***;
***          istruzione BY          ***;
*****;
title1 'Procedura MEANS: analisi per gruppi di obs
omogenee';
title2 'Istruzione BY';

PROC sort data=dati.turismo;
by sesso;
proc means data=dati.turismo sum mean std;
var eta;
by sesso;
run;

```

Procedura MEANS: analisi per gruppi di obs omogenee  
Istruzione BY

SESSO=F

The MEANS Procedure

Analysis Variable : ETA

| Sum      | Mean       | Std Dev    |
|----------|------------|------------|
| 16202.00 | 34.4723404 | 11.8355055 |

SESSO=M

Analysis Variable : ETA

| Sum      | Mean       | Std Dev    |
|----------|------------|------------|
| 16252.00 | 34.2869198 | 11.4965735 |

```

*MEANS9.SAS;
*****;
***  proc means: analisi per gruppi di obs omogenee ***;
***          istruzioni BY e CLASS per 2 o più var          ***;
*****;
title1 'Procedura MEANS: analisi per gruppi di obs
omogenee';
title2 'istruzioni BY e CLASS per 2 o piu' var';

PROC sort data=dati.turismo;
by sesso;
proc means data=dati.turismo sum mean std;
var eta;
by sesso sci;
run;

```

Procedura MEANS: analisi per gruppi di obs omogenee  
istruzioni BY e CLASS per 2 o piu' var

SESSO=F pratica dello sci=.

The MEANS Procedure

Analysis Variable : ETA

| Sum         | Mean       | Std Dev    |
|-------------|------------|------------|
| 244.0000000 | 40.6666667 | 15.8198188 |

SESSO=F pratica dello sci=1

Analysis Variable : ETA

| Sum     | Mean       | Std Dev    |
|---------|------------|------------|
| 8012.00 | 31.7936508 | 10.8325299 |

*eccetera!!!*

```

*MEANS9.SAS;
*****;
***  proc means: analisi per gruppi di obs omogenee ***;
***  istruzioni BY e CLASS per 2 o più var      ***;
*****;
title1 'Procedura MEANS: analisi per gruppi di obs
omogenee';
title2 'istruzioni BY e CLASS per 2 o piu' var';
proc means data=dati.turismo sum mean std;
var eta;
class sesso sci;
run;

```

Procedura MEANS: analisi per gruppi di obs omogenee  
istruzioni BY e CLASS per 2 o piu' var

The MEANS Procedure

Analysis Variable : ETA

| pratica         | N   |     |         |            |            |
|-----------------|-----|-----|---------|------------|------------|
| SESSO dello sci | Obs | Sum | Mean    | Std Dev    |            |
| F               | 1   | 252 | 8012.00 | 31.7936508 | 10.8325299 |
|                 | 2   | 40  | 1320.00 | 33.0000000 | 9.4787211  |
|                 | 3   | 64  | 2122.00 | 33.1562500 | 10.7853792 |
|                 | 4   | 108 | 4504.00 | 41.7037037 | 12.3335719 |
| M               | 1   | 288 | 9280.00 | 32.2222222 | 10.8009836 |
|                 | 2   | 48  | 1944.00 | 40.5000000 | 11.3924200 |
|                 | 3   | 70  | 2326.00 | 33.2285714 | 10.3813816 |
|                 | 4   | 64  | 2542.00 | 39.7187500 | 12.7285066 |

```

*MEANS10.SAS;
*****;
***  proc means: creazione SAS data set ***;
***  istruzione OUTPUT          ***;
*****;
title1 'Procedura MEANS: creazione SAS data set';
title2 'istruzione OUTPUT';
libname dati 'v:\didat\carla\sas';
title3 'default';
proc means data=dati.turismo maxdec=2 fw=8;
var eta spesa;
output;
run;
proc print;
run;

```

Procedura MEANS: creazione SAS data set  
istruzione OUTPUT  
default

| Obs | _TYPE_ | _FREQ_ | _STAT_ | ETA     | SPESA   |
|-----|--------|--------|--------|---------|---------|
| 1   | 0      | 944    | N      | 944.000 | 944.00  |
| 2   | 0      | 944    | MIN    | 5.000   | 0.00    |
| 3   | 0      | 944    | MAX    | 73.000  | 8000.00 |
| 4   | 0      | 944    | MEAN   | 34.379  | 571.15  |
| 5   | 0      | 944    | STD    | 11.661  | 991.5   |

```

title3 'opzione OUT= e scelta indici';
proc means data=dati.turismo maxdec=2 fw=8
noprnt;
var eta spesa;
output out=indici mean=meta mspesa var=veta
vspesa;
run;
proc print;
run;

```

Procedura MEANS: creazione SAS data set  
 istruzione OUTPUT  
 opzione OUT= e scelta indici

| Obs | _TYPE_ | _FREQ_ | meta    | mspesa  | veta    | vspesa    |
|-----|--------|--------|---------|---------|---------|-----------|
| 1   | 0      | 944    | 34.3792 | 571.150 | 135.973 | 983192.75 |

```

*MEANS11.SAS;
*****;
*** proc means: trattamento dei casi mancanti ***;
*****;
title1 'Procedura MEANS: trattamento dei casi mancanti';

*DATI;
DATA DATI;
INPUT X Y @@;
CARDS;
10 1 15 1 23 . 8 1 7 1 . 1
;
PROC PRINT;
SUM X Y;
RUN;

title2 'DEFAULT: esclude obs con valori mancanti da tutti
gli indici';
proc means;
run;

```

*output*

Procedura MEANS: trattamento dei casi mancanti

| OBS | X  | Y |
|-----|----|---|
| 1   | 10 | 1 |
| 2   | 15 | 1 |
| 3   | 23 | . |
| 4   | 8  | 1 |
| 5   | 7  | 1 |
| 6   | .  | 1 |
|     | == | = |
|     | 63 | 5 |

| Variable | N | Mean      | Std Dev   | Minimum  | Maximum   |
|----------|---|-----------|-----------|----------|-----------|
| X        | 5 | 12.600000 | 6.5802736 | 7.000000 | 23.000000 |
| Y        | 5 | 1.000000  | 0         | 1.000000 | 1.000000  |

```
TITLE2 'missing nella WEIGHT: esclude
obs con peso miss dalla media';
```

```
proc means;
weight y;
run;
```

*output*

Procedura MEANS: trattamento dei casi mancanti  
missing nella WEIGHT: esclude obs con peso miss dalla media

The MEANS Procedure

Analysis Variable : X

| N | Mean       | Std Dev   | Minimum   | Maximum    |
|---|------------|-----------|-----------|------------|
| 4 | 10.0000000 | 3.5590261 | 7.0000000 | 15.0000000 |

```
title2 'missing nella FREQ: esclude obs
con valori mancanti da tutti gli
indici';
```

```
proc means;
freq y;
run;
```

*output*

Procedura MEANS: trattamento dei casi mancanti  
missing nella FREQ: esclude obs con valori mancanti da tutti gli indici

Analysis Variable : X

| N | Mean       | Std Dev   | Minimum   | Maximum    |
|---|------------|-----------|-----------|------------|
| 4 | 10.0000000 | 3.5590261 | 7.0000000 | 15.0000000 |

```
*MEANS12.SAS;
*****;
*** proc means: trattamento dei casi mancanti ***;
*** opzione MISSING ***;
*****;
title1 'Procedura MEANS: trattamento dei casi mancanti';
title2 'Missing nella BY: missing categoria a parte';
```

```
proc sort data=dati.turismo;
by sci;
run;
proc means data=dati.turismo fw=6 maxdec=2;
var spesa;
by sci;
run;
```

*output*

Procedura MEANS: trattamento dei casi mancanti  
Missing nella BY: missing categoria a parte

Analysis Variable : SPESA spesa pro-capite

pratica dello sci=.

| N  | Mean   | Std Dev | Minimum | Maximum |
|----|--------|---------|---------|---------|
| 10 | 185.00 | 474.96  | 0.00    | 1500.0  |

pratica dello sci=SOLO ALPINO

| N   | Mean   | Std Dev | Minimum | Maximum |
|-----|--------|---------|---------|---------|
| 540 | 535.81 | 913.49  | 0.00    | 5500.0  |

```

title2 'Missing nella CLASS: obs missing escluse
dal calcolo';
proc means data=dati.turismo fw=6 maxdec=2;
var spesa;
class sci;
run;

```

*output*

Procedura MEANS: trattamento dei casi mancanti  
Missing nella CLASS: obs missing escluse dal calcolo  
Analysis Variable : SPESA spesa pro-capite

| SCI         | N Obs | N   | Mean   | Std Dev | Minimum | Maximum |
|-------------|-------|-----|--------|---------|---------|---------|
| SOLO ALPINO | 540   | 540 | 535.81 | 913.49  | 0.00    | 5500.0  |
| SOLO FONDO  | 88    | 88  | 573.52 | 982.46  | 0.00    | 6030.0  |
| ENTRAMBI    | 134   | 134 | 582.78 | 846.52  | 0.00    | 3500.0  |
| NESSUNO     | 172   | 172 | 694.29 | 1305.1  | 0.00    | 8000.0  |

```

title2 'Missing nella CLASS: opzione MISSING';
proc means data=dati.turismo fw=6 maxdec=2 missing;
var spesa;
class sci;
run;

```

*output*

Procedura MEANS: trattamento dei casi mancanti  
Missing nella CLASS: opzione MISSING  
Analysis Variable : SPESA spesa pro-capite

| SCI         | N Obs | N   | Mean   | Std Dev | Minimum | Maximum |
|-------------|-------|-----|--------|---------|---------|---------|
| .           | 10    | 10  | 185.00 | 474.96  | 0.00    | 1500.0  |
| SOLO ALPINO | 540   | 540 | 535.81 | 913.49  | 0.00    | 5500.0  |
| SOLO FONDO  | 88    | 88  | 573.52 | 982.46  | 0.00    | 6030.0  |
| ENTRAMBI    | 134   | 134 | 582.78 | 846.52  | 0.00    | 3500.0  |
| NESSUNO     | 172   | 172 | 694.29 | 1305.1  | 0.00    | 8000.0  |

## PROCEDURA UNIVARIATE

Sintesi di distribuzioni semplici attraverso indici statistici, grafici e tabelle.

```

PROC UNIVARIATE [opzioni];
BY variabile;
CLASS var1 [[var2]...];
VAR elenco variabili;
FREQ variabile;
WEIGHT variabile;
ID variabile;
OUTPUT OUT=archivio
      [elenco chiavi]
      [PCTLPTS=percen PCTLPRE=nome];
      [PCTLNAMES=suffisso]
RUN;

```



```

*univar1.SAS;
*****;
***  proc UNIVARIATE: output di default ***;
*****;
options nocenter nodate nonumber;

libname dati 'v:\didat\carla\sas';
title1 'Procedura UNIVARIATE: output di
default';
proc univariate data=dati.VALUTA;
var DOM5;
run;

```

*output di univar1.sas*

Procedura UNIVARIATE: output di default

The UNIVARIATE Procedure

Variable: dom5 (Voto riportato al Diploma)

Moments

|                 | 962        | Sum Weights      | 962        |
|-----------------|------------|------------------|------------|
| N               |            |                  |            |
| Mean            | 49.0488565 | Sum Observations | 47185      |
| Std Deviation   | 7.11059532 | Variance         | 50.5605658 |
| Skewness        | -0.0445876 | Kurtosis         | -1.0061163 |
| Uncorrected SS  | 2362959    | Corrected SS     | 48588.7037 |
| Coeff Variation | 14.4969645 | Std Error Mean   | 0.22925479 |

*Segue output di univar1.sas*

Basic Statistical Measures

| Location |          | Variability         |          |
|----------|----------|---------------------|----------|
| Mean     | 49.04886 | Std Deviation       | 7.11060  |
| Median   | 49.00000 | Variance            | 50.56057 |
| Mode     | 60.00000 | Range               | 24.00000 |
|          |          | Interquartile Range | 12.00000 |

Tests for Location: Mu0=0

| Test        | -Statistic- | -----p Value----- |        |
|-------------|-------------|-------------------|--------|
| Student's t | t 213.9491  | Pr >  t           | <.0001 |
| Sign        | M 481       | Pr >=  M          | <.0001 |
| Signed Rank | S 231601.5  | Pr >=  S          | <.0001 |

Quantiles (Definition 5)

| Quantile   | Estimate |
|------------|----------|
| 100% Max   | 60       |
| 99%        | 60       |
| 95%        | 60       |
| 90%        | 60       |
| 75% Q3     | 55       |
| 50% Median | 49       |
| 25% Q1     | 43       |
| 10%        | 40       |
| 5%         | 37       |
| 1%         | 36       |
| 0% Min     | 36       |

Segue output di univar1.sas

The UNIVARIATE Procedure

Variable: dom5 (Voto riportato al Diploma)

Extreme Observations

| ----Lowest---- |     | ----Highest--- |     |
|----------------|-----|----------------|-----|
| Value          | Obs | Value          | Obs |
| 36             | 974 | 60             | 965 |
| 36             | 957 | 60             | 975 |
| 36             | 951 | 60             | 984 |
| 36             | 943 | 60             | 985 |
| 36             | 942 | 60             | 988 |

Missing Values

| Missing Value | Count | -----Percent Of----- |             |
|---------------|-------|----------------------|-------------|
|               |       | All Obs              | Missing Obs |
| .             | 31    | 3.12                 | 100.00      |

## CALCOLO DEI QUANTILI

$x_p$  quantile di ordine  $p$

$$x_p : \begin{cases} F(x_p) \leq p \\ 1 - F(x_p) \geq p \end{cases}$$

$N$  n. casi validi per  $X$

$x_1, x_2, \dots, x_N$  valori ordinati di  $X$

$Np = j + g$   $j$  = parte intera,  $g$  = parte dec.

$$x_p = (x_j + x_{j+1}) / 2 \quad \text{se } g = 0$$

$$x_p = x_{j+1} \quad \text{se } g > 0$$

```

*univar2.SAS;
*****;
*** proc UNIVARIATE: grafici con opzione PLOT ***;
*****;
proc univariate data=dati.valuta plot;
var dom5;
run;

```

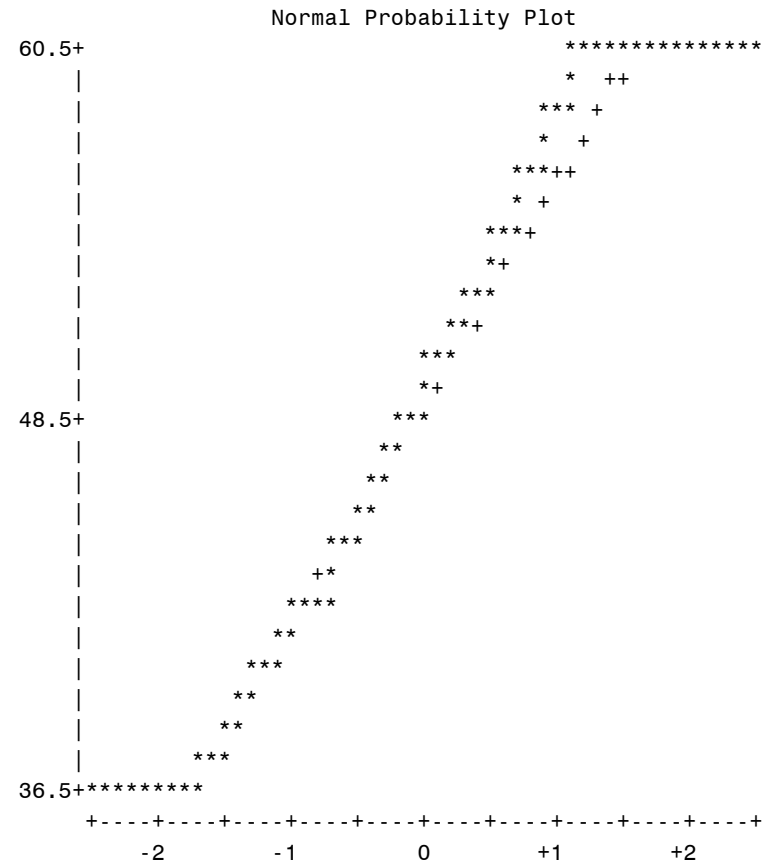
*output di univar2.sas*

Univariate Procedure

Variable=DOM5      Voto riportato al Diploma

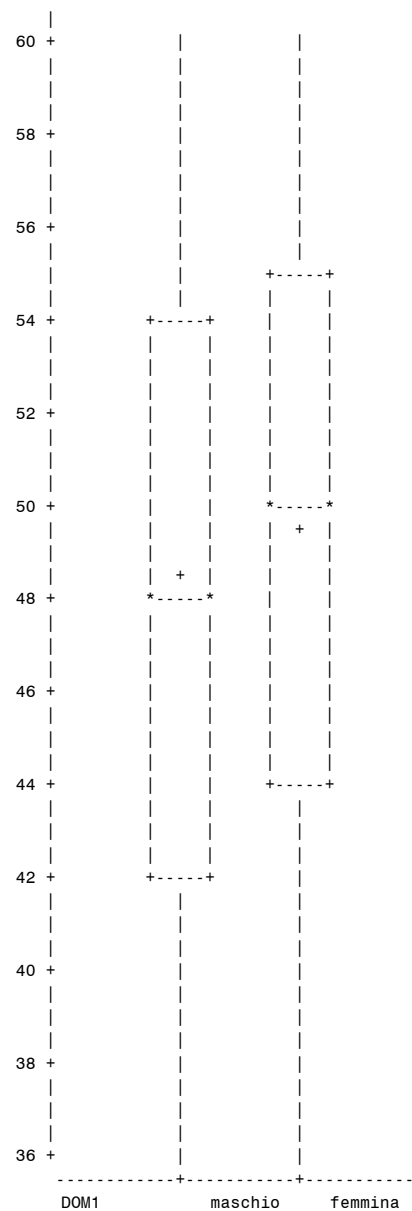
| Histogram  | #   | Boxplot  |
|------------|-----|----------|
| 60.5+***** | 122 |          |
| .*         | 2   |          |
| .*****     | 44  |          |
| .****      | 12  |          |
| .*****     | 51  |          |
| .*****     | 14  | +-----+  |
| .*****     | 38  |          |
| .****      | 11  |          |
| .*****     | 88  |          |
| .****      | 15  |          |
| .*****     | 76  |          |
| .*****     | 18  | *- - - * |
| 48.5+***** | 83  |          |
| .*****     | 26  |          |
| .*****     | 45  |          |
| .*****     | 31  |          |
| .*****     | 41  |          |
| .*****     | 19  | +-----+  |
| .*****     | 75  |          |
| .*****     | 20  |          |
| .*****     | 37  |          |
| .****      | 15  |          |
| .*****     | 20  |          |
| .*****     | 16  |          |
| 36.5+***** | 43  |          |

-----+  
\* may represent up to 3 counts





Procedura UNIVARIATE: istruzione BY  
 Univariate Procedure Schematic Plots  
 Variable=DOM5 Voto riportato al Diploma



```
*univar4.SAS;
*****;
*** proc UNIVARIATE: opzione FREQ ***;
*****;
title1 'Procedura UNIVARIATE: opzione FREQ';
```

```
proc univariate data=dati.valuta FREQ;
var dom5;
run;
```

*output di univar4.sas*

Frequency Table

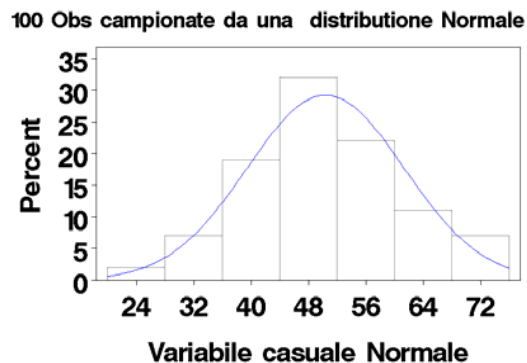
| Value | Count | Percents |      | Value | Count | Percents |       |
|-------|-------|----------|------|-------|-------|----------|-------|
|       |       | Cell     | Cum  |       |       | Cell     | Cum   |
| 36    | 43    | 4.5      | 4.5  | 49    | 18    | 1.9      | 50.8  |
| 37    | 16    | 1.7      | 6.1  | 50    | 76    | 7.9      | 58.7  |
| 38    | 20    | 2.1      | 8.2  | 51    | 15    | 1.6      | 60.3  |
| 39    | 15    | 1.6      | 9.8  | 52    | 88    | 9.1      | 69.4  |
| 40    | 37    | 3.8      | 13.6 | 53    | 11    | 1.1      | 70.6  |
| 41    | 20    | 2.1      | 15.7 | 54    | 38    | 4.0      | 74.5  |
| 42    | 75    | 7.8      | 23.5 | 55    | 14    | 1.5      | 76.0  |
| 43    | 19    | 2.0      | 25.5 | 56    | 51    | 5.3      | 81.3  |
| 44    | 41    | 4.3      | 29.7 | 57    | 12    | 1.2      | 82.5  |
| 45    | 31    | 3.2      | 33.0 | 58    | 44    | 4.6      | 87.1  |
| 46    | 45    | 4.7      | 37.6 | 59    | 2     | 0.2      | 87.3  |
| 47    | 26    | 2.7      | 40.3 | 60    | 122   | 12.7     | 100.0 |
| 48    | 83    | 8.6      | 49.0 |       |       |          |       |

## ISTRUZIONE HISTOGRAM

```
HISTOGRAM <var> </ option(s)>;
```

crea un diagramma a barre ad alta risoluzione e volendo sovrappone curve di densità stimate in modo parametrico o non parametrico

```
title '100 Obs campionate da una distribuzione Normale';
goptions htitle=4 htext=3 ftext=swissb
ftitle=swissb;
data distrdata;
  drop n;
  label Normal_x='Variabile casuale Normale';
  do n=1 to 100;
    Normal_x=10*rannor(53124)+50;
    output;
  end;
run;
proc univariate data=distrdata noprint;
  var Normal_x;
histogram Normal_x /normal(noprint) cbarline=grey
;run;
```



## PROCEDURA BOXPLOT

```
PROC BOXPLOT < options > ;
```

```
PLOT analysis-variable*group-  
variable < (block-variables ) > <  
=symbol-variable > </ options > ;
```

```
BY variables;
```

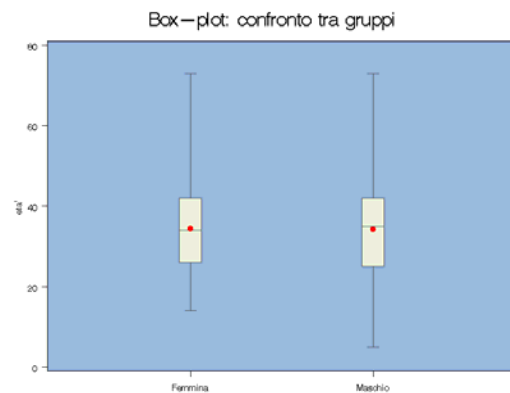
```
ID variables;
```

```
*boxplot.sas;  
title 'Box-plot: confronto tra gruppi';  
goptions ftext=swiss;  
axis1 minor=none color=black  
      label=(angle=90 rotate=0);  
symbol1 v=dot c=red;  
proc boxplot data=dati.turismo;  
  plot eta* sesso /  
  nohlabel  
  symbollegend = legend1  
  cboxes      = dagr  
  cboxfill    = ywh  
  cframe      = vligb  
  vaxis       = axis1;  
  legend1 label=('tipo di sci:')  
  cborder=black cframe=ligr;  
  label eta = 'età';  
  format sesso $sexfmt.;  
run;
```

Il box plot rappresenta sinteticamente la distribuzione della variabile età: ognuno dei 2 box-plot descrive la variabile per sesso.

Gli elementi del grafico e le statistiche in esso rappresentate sono i seguenti:

- l'altezza della *scatola* rappresenta lo scarto interquartile (differenza tra il 25-esimo e 75-esimo percentile)
- il pallino rosso interno alla scatola è la media aritmetica
- la linea orizzontale nella scatola è la mediana
- le linee verticali che partono dalla scatola si estendono fino al valore minimo e massimo della variabile di analisi



## LA PROCEDURA TABULATE

**PROC TABULATE** *<option(s)>*;

**BY** *<DESCENDING> var1*

*<...<DESCENDING> varn>*

*<NOTSORTED>*;

**CLASS** *variable(s) </ options>*;

**CLASSLEV** *variable(s) / style =<style-element-name | <PARENT>> <[style-attribute-specification(s)]>*;

**FREQ** *variable;*

**KEYLABEL** *keyword-1='description-1'*  
*<...keyword-n='description-n'>*;

**KEYWORD** *keyword(s) / style =<style-element-name | <PARENT>> <[style-attribute-specification(s)]>*;

**TABLE** *<<page-expression,> row-expression,> column-expression </ table-option(s)>*;

**VAR** *analysis-variable(s) </ options>*;

**WEIGHT** *variable;*

```
*tables1.sas;
*****;
*** Procedura TABULATE: tabelle di frequenza ***;
*****;
```

Title 'Procedura TABULATE: tabelle di frequenza:  
istruzione CLASS';

```
proc tabulate data=dati.valuta;
class dom5;
tables dom5;
format dom5 votomat.;
run;
```

*output di tables1.sas*

Procedura TABULATE: tabelle di frequenza: istruzione CLASS

| Voto riportato al Diploma |        |        |        |        |        |
|---------------------------|--------|--------|--------|--------|--------|
| 36-40                     | 41-45  | 46-50  | 51-54  | 55-59  | 60     |
| N                         | N      | N      | N      | N      | N      |
| 131.00                    | 186.00 | 248.00 | 152.00 | 123.00 | 122.00 |

```
proc tabulate data=dati.valuta;
class dom5 dom1;
tables dom5, dom1;
format dom5 votomat. dom1 sessoof.;
run;
```

*output*

Procedura TABULATE: tabelle di frequenza: istruzione CLASS

|                           | Sesso   |         |
|---------------------------|---------|---------|
|                           | maschio | femmina |
|                           | N       | N       |
| Voto riportato al Diploma |         |         |
| 36-40                     | 72.00   | 59.00   |
| 41-45                     | 66.00   | 118.00  |
| 46-50                     | 84.00   | 164.00  |
| 51-54                     | 55.00   | 97.00   |
| 55-59                     | 41.00   | 82.00   |
| 60                        | 48.00   | 74.00   |



```
*tables2.sas;
Title1 'Procedura TABULATE. Tabelle di intensita''
istruzione VAR';
title2 'Valutazione della didattica a SCPOL';
proc tabulate data=dati.valuta;
class dom1;
var dom5;
tables dom5,dom1;
tables dom5*dom1;
format dom5 votomat. dom1 sesso;.
run;
```

*output*

Procedura TABULATE. Tabelle di intensita' istruzione VAR  
 Valutazione della didattica a SCPOL

|                                      |     | Sesso    |          |
|--------------------------------------|-----|----------|----------|
|                                      |     | maschio  | femmina  |
| Voto<br>riporta-<br>to al<br>Diploma | SUM | 17674.00 | 29427.00 |

|                           |          |
|---------------------------|----------|
| Voto riportato al Diploma |          |
| SUM                       |          |
| Sesso                     |          |
| maschio                   | femmina  |
| 17674.00                  | 29427.00 |

## CONTENUTO DELLE CELLE

### 1. Default

- *frequenze* **N**, per variabili di classificazione dichiarate nella CLASS;
- *intensità* **SUM** se sono utilizzate una o più variabili di analisi, dichiarate nella VAR.

### 2. Indici statistici su richiesta.

## ISTRUZIONE TABLES

**TABLES** *espr1*[,*espr2*,*espr3*][/*opzioni*];

L'istruzione TABLE definisce contenuto e struttura tabella e contiene da una a 3 espressioni separate da virgole.

- Se ci sono tutte e 3 le espressioni queste definiscono (da sin. a dx):  
    *pagine, righe, colonne*
- Se ci sono 2 espressioni queste definiscono (da sin. a dx):  
    *righe, colonne*
- Se c'è 1 sola espressione questa definisce le colonne

## ESPRESSIONI DI TABLES

Ciascuna *espressione* è composta da elementi e operatori

### Elementi

- Variabili di analisi (dichiarate nell'istruzione VAR)
- Variabili di classificazione (dichiarate nell'istruzione CLASS)
- la variabile di sistema ALL, che calcola il totale
- keywords per gli indici statistici (MEANS, VAR, etc.)
- formati per i valori nelle celle  
    *F=formato*
- etichette *variable = 'descrizione'*

## OPERATORI DI TABLE

- \* incrocio
- (blank) concatenazione
- () raggruppamento
- < > denominatore delle percentuali  
(PCTN e PCTSUM)

## OPERATORE INCROCIO ‘\*\*’

```
*tables3.sas;
Title 'Procedura TABULATE. Operatore di incrocio';
title2 'Valutazione della didattica a SCPOL';
proc tabulate data=dati.valuta format=8.0;
class dom1 dom5 dom7;
tables dom5, dom1*dom7/rts=10;
format dom7 posizf. dom5 votomat. dom1 sessoof.;
run;
```

Procedura TABULATE. Operatore di incrocio  
Valutazione della didattica a SCPOL

|                                      | Sesso     |                |                |           |                |                |
|--------------------------------------|-----------|----------------|----------------|-----------|----------------|----------------|
|                                      | maschio   |                |                | femmina   |                |                |
|                                      | Posizione |                |                | Posizione |                |                |
|                                      | Regolare  | Ripeten-<br>te | Fuori<br>corso | Regolare  | Ripeten-<br>te | Fuori<br>corso |
|                                      | N         | N              | N              | N         | N              | N              |
| Voto<br>riporta-<br>to al<br>Diploma |           |                |                |           |                |                |
| 36-40                                | 61        | 2              | 7              | 49        | .              | 9              |
| 41-45                                | 55        | 1              | 8              | 97        | .              | 20             |
| 46-50                                | 81        | .              | 3              | 149       | 2              | 13             |
| 51-54                                | 51        | .              | 4              | 89        | 1              | 6              |
| 55-59                                | 40        | .              | 1              | 77        | .              | 4              |
| 60                                   | 45        | .              | 2              | 70        | 2              | 1              |

```
*tables4.sas;
proc tabulate data=dati.turismo;
class sci alloggio;
var pernotta;
tables sci, alloggio*pernotta;
format alloggio tipofmt.;
run;
```

|                   | tipo di alloggio   |                        |                    |                    |                    |
|-------------------|--------------------|------------------------|--------------------|--------------------|--------------------|
|                   | ALBERGHIERO        | EXTRA -<br>ALBERGHIERI | PROPRIETA'         | ALTRO              | PENDOLARI          |
|                   | pernottamen-<br>ti | pernottamen-<br>ti     | pernottamen-<br>ti | pernottamen-<br>ti | pernottamen-<br>ti |
|                   | SUM                | SUM                    | SUM                | SUM                | SUM                |
| pratica dello sci |                    |                        |                    |                    |                    |
| SOLO ALPINO       | 1528.00            | 1305.00                | 1274.00            | 60.00              | .                  |
| SOLO FONDO        | 204.00             | 276.00                 | 170.00             | 28.00              | .                  |
| ENTRAMBI          | 560.00             | 364.00                 | 366.00             | .                  | .                  |
| NESSUNO           | 546.00             | 236.00                 | 324.00             | 67.00              | .                  |

‘\*’ non può essere utilizzato tra variabili di analisi o tra indici!!!

```
169 proc tabulate data=dati.valuta;
170 class dom1;
171 var dom5 dom6;
172 tables dom1, dom5*dom6;
173 format dom1 sesso f.;
174 run;
```

ERROR: Esistono piu' variabili di analisi associate a una singola cella della tabella nella

seguente nidificazione: DOM1 \* DOM5 \* DOM6.

NOTE: SAS System ha interrotto l'elaborazione del passo a causa errori.

## CONCATENAZIONE

Si ottiene elencando gli elementi separati da uno o più spazi bianchi. E' valida tra qualsiasi tipo di variabile e/o indice.

```
*tables5.sas;
proc tabulate data=dati.valuta;
class dom1 dom7;
var dom5;
tables dom1 dom7;
format dom1 sesso f. dom7 posiz f.;
run;
```

| Sesso   |         | Posizione |           |             |
|---------|---------|-----------|-----------|-------------|
| maschio | femmina | Regolare  | Ripetente | Fuori corso |
| N       | N       | N         | N         | N           |
| 373.00  | 605.00  | 890.00    | 8.00      | 80.00       |

| Voto<br>riportato al<br>Diploma | Sesso   |         |
|---------------------------------|---------|---------|
|                                 | maschio | femmina |
| SUM                             | N       | N       |
| 46629.00                        | 373.00  | 605.00  |

## UTILIZZO DELLE PARENTESI

```
*tables6.sas;
proc tabulate data=dati.valuta;
class dom1 ;
var dom5 dom8;
tables dom1*(dom5 dom8);
tables dom1*dom5 dom1*dom8; *stesso risultato;
format dom1 sessof. ;
run;
```

| Sesso                           |                                   |                                 |                                   |
|---------------------------------|-----------------------------------|---------------------------------|-----------------------------------|
| maschio                         |                                   | femmina                         |                                   |
| Voto<br>riportato al<br>Diploma | Corsi<br>frequentati<br>nell'anno | Voto<br>riportato al<br>Diploma | Corsi<br>frequentati<br>nell'anno |
| SUM                             | SUM                               | SUM                             | SUM                               |
| 17674.00                        | 1293.00                           | 29427.00                        | 2273.00                           |

## INDICI

Calcolati per gruppi omogenei di osservazioni definiti dalle variabili di classificazione.

```
*tables7.sas;
proc tabulate data=dati.valuta;
class dom1;
var dom5 dom8;
tables dom1*(dom5 dom8)*mean;
format dom1 sessof. ;
run;
```

| Sesso                           |                                   |                                 |                                   |
|---------------------------------|-----------------------------------|---------------------------------|-----------------------------------|
| maschio                         |                                   | femmina                         |                                   |
| Voto<br>riportato al<br>Diploma | Corsi<br>frequentati<br>nell'anno | Voto<br>riportato al<br>Diploma | Corsi<br>frequentati<br>nell'anno |
| MEAN                            | MEAN                              | MEAN                            | MEAN                              |
| 48.29                           | 3.52                              | 49.54                           | 3.82                              |

*Indici disponibili:*

- Parole chiave come nelle procedure MEANS e UNIVARIATE;
- PCTN e PCTSUM per il calcolo delle %.

```
*tables8.sas;
proc tabulate data=dati.valuta;
class dom1 corso ;
var dom5;
tables corso all, (dom1 all)*dom5*mean;
format dom1 sessof. corso nomec.;
```

|                  | Sesso                           |                                 | ALL                             |
|------------------|---------------------------------|---------------------------------|---------------------------------|
|                  | maschio                         | femmina                         |                                 |
|                  | Voto<br>riportato al<br>Diploma | Voto<br>riportato al<br>Diploma | Voto<br>riportato al<br>Diploma |
|                  | MEAN                            | MEAN                            | MEAN                            |
| CORSO            |                                 |                                 |                                 |
| Statistica (A-G) | 48.94                           | 51.29                           | 50.48                           |
| Statistica (H-Z) | 45.37                           | 47.49                           | 46.79                           |
| Soc. (A-G)+(H-Z) | 47.64                           | 48.88                           | 48.43                           |
| St.contemp.(A-G) | 48.02                           | 49.96                           | 49.20                           |
| St.contemp.(H-Z) | 48.25                           | 48.55                           | 48.39                           |
| St.dot.pol.(A-G) | 52.70                           | 51.06                           | 51.69                           |
| St.dot.pol.(H-Z) | 46.17                           | 53.25                           | 50.21                           |
| Is.dir.pub.(H-Z) | 47.09                           | 51.94                           | 49.96                           |
| Econ. pol. (A-G) | 53.25                           | 50.45                           | 51.05                           |
| Econ. pol. (H-Z) | 48.92                           | 47.90                           | 48.21                           |
| Dir.cos.it.comp. | 48.69                           | 48.59                           | 48.63                           |
| Is. dir. privato | 48.63                           | 48.87                           | 48.76                           |
| Sc.politica(A-G) | 47.41                           | 50.80                           | 49.24                           |
| Sc.politica(H-Z) | 50.18                           | 51.46                           | 50.95                           |
| Pol. econ. (A-G) | 50.35                           | 50.20                           | 50.29                           |
| Pol. econ. (H-Z) | 52.17                           | 51.45                           | 51.72                           |
| ALL              | 48.29                           | 49.54                           | 49.06                           |

## VARIABILI DI ANALISI

```
Permanenza media per tipo di alloggio
*tables9.sas;
proc tabulate data=dati.turismo format=8.2;
class alloggio;
var pernotto;
tables alloggio all, pernotto*(n*f=8.0 mean var);
format alloggio tipofmt.;
```

|                     | pernottamenti |      |       |
|---------------------|---------------|------|-------|
|                     | N             | MEAN | VAR   |
| tipo di alloggio    |               |      |       |
| ALBERGHIERO         | 378           | 7.59 | 28.91 |
| EXTRA - ALBERGHIERI | 229           | 9.66 | 85.82 |
| PROPRIETA'          | 260           | 8.35 | 15.01 |
| ALTRO               | 18            | 6.22 | 8.65  |
| PENDOLARI           | 0             | .    | .     |
| ALL                 | 885           | 8.32 | 39.82 |

### Regole generali

1. non si possono incrociare tra loro variabili di analisi;
2. non si possono incrociare tra loro indici;
3. le variabili di analisi possono trovarsi in una qualunque delle dimensioni della tabella, ma tutte nella stessa.

## DISTRIBUZIONI PERCENTUALI

Chiavi PCTN e PCTSUM

**PCTN|PCTSUM**<denominatore>

```
*tables10.sas;
proc tabulate data=dati.turismo format=6.2;
class eta sesso;
tables eta, sesso*(N*f=6.0 PCTN)/rts=15;
keylabel all='TOT';
format eta etafmt. sesso $sexfmt.;
```

Stesso risultato esplicitando denominatore

```
tables eta, sesso*(N*f=6.0 PCTN<eta*sesso>)/rts=15;
```

|            | SESSO   |       |         |       |
|------------|---------|-------|---------|-------|
|            | Femmina |       | Maschio |       |
|            | N       | PCTN  | N       | PCTN  |
| ETA        |         |       |         |       |
| 0 - 13     | .       | .     | 2       | 0.21  |
| 14 - 29    | 174     | 18.43 | 184     | 19.49 |
| 30-59      | 276     | 29.24 | 284     | 30.08 |
| 60 E OLTRE | 20      | 2.12  | 4       | 0.42  |

## PERCENTUALI SU TOTALE DI COLONNA

Sommare i valori rispetto alle righe

```
tables eta all, (sesso all)*(N*f=6.0 PCTN<eta all>);
```

```
tables eta all, (sesso all)*(N*f=6.0 COLPCTN); *v9;
```

|            | SESSO   |        |         |        | TOT |        |
|------------|---------|--------|---------|--------|-----|--------|
|            | Femmina |        | Maschio |        |     |        |
|            | N       | PCTN   | N       | PCTN   | N   | PCTN   |
| ETA        |         |        |         |        |     |        |
| 0 - 13     | .       | .      | 2       | 0.42   | 2   | 0.21   |
| 14 - 29    | 174     | 37.02  | 184     | 38.82  | 358 | 37.92  |
| 30-59      | 276     | 58.72  | 284     | 59.92  | 560 | 59.32  |
| 60 E OLTRE | 20      | 4.26   | 4       | 0.84   | 24  | 2.54   |
| TOT        | 470     | 100.00 | 474     | 100.00 | 944 | 100.00 |

## PERCENTUALE SU TOTALE DI RIGA

Sommare i valori rispetto alle colonne

```
tables eta all, (sesso all)*(N*f=6.0 PCTN<sesso all>);
tables eta all, (sesso all)*(N*f=6.0 ROWPCTN); *v9;
```

|            | SESSO   |       |         |        | TOT |        |
|------------|---------|-------|---------|--------|-----|--------|
|            | Femmina |       | Maschio |        |     |        |
|            | N       | PCTN  | N       | PCTN   | N   | PCTN   |
| ETA        |         |       |         |        |     |        |
| 0 - 13     | .       | .     | 2       | 100.00 | 2   | 100.00 |
| 14 - 29    | 174     | 48.60 | 184     | 51.40  | 358 | 100.00 |
| 30-59      | 276     | 49.29 | 284     | 50.71  | 560 | 100.00 |
| 60 E OLTRE | 20      | 83.33 | 4       | 16.67  | 24  | 100.00 |
| TOT        | 470     | 49.79 | 474     | 50.21  | 944 | 100.00 |

## PERCENTUALI PER VARIABILI DI ANALISI

Utilizzare la parola chiave PCTSUM

```
*tables12.sas;
proc tabulate data=dati.turismo format=6.2;
class alloggio sesso;
var pernotto;
tables alloggio all, (sesso all)*pernotto*(N*f=6.0 SUM
PCTSUM<alloggio all>)/rts=15;
tables alloggio all, (sesso all)*pernotto*(N*f=6.0 SUM
COLPCTSUM)/rts=15; *v9;
keylabel all='TOT';
format alloggio tipofmt. sesso $sexfmt.;
run;
```

|                     | SESSO         |        |        |               |        |        | TOT           |        |        |
|---------------------|---------------|--------|--------|---------------|--------|--------|---------------|--------|--------|
|                     | Femmina       |        |        | Maschio       |        |        |               |        |        |
|                     | pernottamenti |        |        | pernottamenti |        |        | pernottamenti |        |        |
|                     | N             | SUM    | PCTSUM | N             | SUM    | PCTSUM | N             | SUM    | PCTSUM |
| tipo di alloggio    |               |        |        |               |        |        |               |        |        |
| ALBERGHIERO         | 198           | 1530.0 | 41.49  | 180           | 1338.0 | 36.41  | 378           | 2868.0 | 38.95  |
| EXTRA - ALBERGHIERI | 116           | 1068.0 | 28.96  | 113           | 1145.0 | 31.16  | 229           | 2213.0 | 30.06  |
| PROPRIETA'          | 116           | 1006.0 | 27.28  | 144           | 1164.0 | 31.67  | 260           | 2170.0 | 29.47  |
| ALTRO               | 12            | 84.00  | 2.28   | 6             | 28.00  | 0.76   | 18            | 112.00 | 1.52   |
| PENDOLARI           | 0             | .      | .      | 0             | .      | .      | 0             | .      | .      |
| TOT                 | 442           | 3688.0 | 100.00 | 443           | 3675.0 | 100.00 | 885           | 7363.0 | 100.00 |



## TRATTAMENTO DEI CASI MANCANTI

Per default esclusione di tutte le osservazione che presentano *missing* per *almeno una* delle variabili:

1. dichiarate in CLASS per le variabili di classificazione;
2. utilizzate nell'istruzione TABLES per le variabili di analisi.

Osservare i missing presenti nelle variabili!!!

| SESSO   | Frequency | Percent | Frequency | Percent |
|---------|-----------|---------|-----------|---------|
| Femmina | 470       | 49.8    | 470       | 49.8    |
| Maschio | 474       | 50.2    | 944       | 100.0   |

| Variable | Label             | N   | Nmiss |
|----------|-------------------|-----|-------|
| SCI      | pratica dello sci | 934 | 10    |
| PERNOTTO | pernottamenti     | 891 | 53    |
| ETA      |                   | 944 | 0     |

```
proc tabulate data=dati.turismo format=6.2;
class sesso sci;
var pernotto eta;
*esclude 10 obs miss in SCI anche se SESSO non ha miss;
tables sesso all;
*esclude 10 obs miss in sci e 53 obs miss in pernotto;
tables sci all, (pernotto eta)*(n*f=6.0 mean);
keylabel all='TOT';
format sesso $sexfmt.;
run;
```

| SESSO        |              | TOT    |
|--------------|--------------|--------|
| Femmi-<br>na | Masch-<br>io |        |
| N            | N            | N      |
| 464.00       | 470.00       | 934.00 |

|                   | pernottamenti |      | ETA |       |
|-------------------|---------------|------|-----|-------|
|                   | N             | MEAN | N   | MEAN  |
| pratica dello sci |               |      |     |       |
| SOLO ALPINO       | 501           | 8.36 | 540 | 32.02 |
| SOLO FONDO        | 88            | 7.70 | 88  | 37.09 |
| ENTRAMBI          | 130           | 9.92 | 134 | 33.19 |
| NESSUNO           | 162           | 7.27 | 172 | 40.97 |
| TOT               | 881           | 8.32 | 934 | 34.31 |

## OPZIONE MISSING

Include *missing* variabili di classificazione

```
proc tabulate data=dati.turismo format=6.2 MISSING;
class sesso sci;
var pernotto eta;
tables sesso all;
tables sci all, (pernotto eta)*(n*f=6.0 mean);
```

| SESSO        |              | TOT    |
|--------------|--------------|--------|
| Femmi-<br>na | Masch-<br>io |        |
| N            | N            | N      |
| 470.00       | 474.00       | 944.00 |

|                   | pernottamenti |      | ETA |       |
|-------------------|---------------|------|-----|-------|
|                   | N             | MEAN | N   | MEAN  |
| pratica dello sci |               |      |     |       |
| .                 | 10            | 9.80 | 10  | 40.40 |
| SOLO ALPINO       | 501           | 8.36 | 540 | 32.02 |
| SOLO FONDO        | 88            | 7.70 | 88  | 37.09 |
| ENTRAMBI          | 130           | 9.92 | 134 | 33.19 |
| NESSUNO           | 162           | 7.27 | 172 | 40.97 |
| TOT               | 891           | 8.34 | 944 | 34.38 |

## PAROLA CHIAVE NMISS

Indica numero valori mancanti per variabili di analisi

```
proc tabulate data=dati.turismo format=6.2 missing;
class sesso sci;
var pernotto eta;
tables sci all, (pernotto eta)*(n*f=6.0 mean
NMISS*f=6.0);
keylabel all='TOT';
format sesso $sexfmt.;
run;
```

|                   | pernottamenti |      |       | ETA |       |       |
|-------------------|---------------|------|-------|-----|-------|-------|
|                   | N             | MEAN | NMISS | N   | MEAN  | NMISS |
| pratica dello sci |               |      |       |     |       |       |
| .                 | 10            | 9.80 | 0     | 10  | 40.40 | 0     |
| SOLO ALPINO       | 501           | 8.36 | 39    | 540 | 32.02 | 0     |
| SOLO FONDO        | 88            | 7.70 | 0     | 88  | 37.09 | 0     |
| ENTRAMBI          | 130           | 9.92 | 4     | 134 | 33.19 | 0     |
| NESSUNO           | 162           | 7.27 | 10    | 172 | 40.97 | 0     |
| TOT               | 891           | 8.34 | 53    | 944 | 34.38 | 0     |

## PROCEDURA CORR

```
PROC CORR COV;  
VAR elenco variabili;  
RUN;
```

- calcola la matrice di correlazione
- opzione **COV** per ottenere anche matrice di covarianza

## LA PROCEDURA REG

```
PROC REG < opzioni > ;  
  <label:>MODEL vardip=<regressori>  
                                < / opzioni > ;  
ID variabili ;  
VAR variabili ;  
ADD variabili ;  
DELETE variabili ;  
OUTPUT<OUT=SAS-data-set>  
  chiave=nomi<... chiave=nomi> ;  
PAINT <condition | ALLOBS>  
  </opzioni> | < STATUS | UNDO> ;  
PLOT<yvar* xvar> <=simbolo>  
  <...yvar* xvar> <=simbolo>  
  </opzioni> ;  
REFIT;
```

## ESEMPIO REGRESSIONE

```

*reg1.sas;
title 'Regressione Lineare Semplice';
data Class;
input Name $ Height Weight Age @@;
datalines;
Alfred 69.0 112.5 14 Alice 56.5 84.0 13 Barbara 65.3 98.0 13
Carol 62.8 102.5 14 Henry 63.5 102.5 14 James 57.3 83.0 12
Jane 59.8 84.5 12 Janet 62.5 112.5 15 Jeffrey 62.5 84.0 13
John 59.0 99.5 12 Joyce 51.3 50.5 11 Judy 64.3 90.0 14
Louise 56.3 77.0 12 Mary 66.5 112.0 15 Philip 72.0 150.0 16
Robert 64.8 128.0 12 Ronald 67.0 133.0 15 Thomas 57.5 85.0 11
William 66.5 112.0 15
;
proc reg;
model Weight = Height;
run;
    
```

| Model: MODEL1              |    |                |             |         |        |  |
|----------------------------|----|----------------|-------------|---------|--------|--|
| Dependent Variable: Weight |    |                |             |         |        |  |
| Analysis of Variance       |    |                |             |         |        |  |
| Source                     | DF | Sum of Squares | Mean Square | F Value | Pr > F |  |
| Model                      | 1  | 7193.24912     | 7193.24912  | 57.08   | <.0001 |  |
| Error                      | 17 | 2142.48772     | 126.02869   |         |        |  |
| Corrected Total            | 18 | 9335.73684     |             |         |        |  |
| Root MSE                   |    | 11.22625       | R-Square    | 0.7705  |        |  |
| Dependent Mean             |    | 100.02632      | Adj R-Sq    | 0.7570  |        |  |
| Coeff Var                  |    | 11.22330       |             |         |        |  |

**MSE** Mean Square Error *varianza degli errori*

**RMSE** Root Mean Square Error *s.e. degli errori*

### Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1  | -143.02692         | 32.27459       | -4.43   | 0.0004  |
| Height    | 1  | 3.89903            | 0.51609        | 7.55    | <.0001  |

La tabella "Parameter Estimates" contiene le stime di  $\beta_0$  e  $\beta_1$ , la statistica  $t$  e il  $p$ -value corrispondente per il test sulla significatività di ciascun parametro.

Il modello stimato è

$$\text{Weight} = -143.0 + 3.9 \times \text{Height}$$

## OPZIONI DI PROC REG

- PLOT scatterplot
  - r. residui (errori)
  - p. valori previsti

*analisi dei RESIDUI e VALORI PREVISTI*

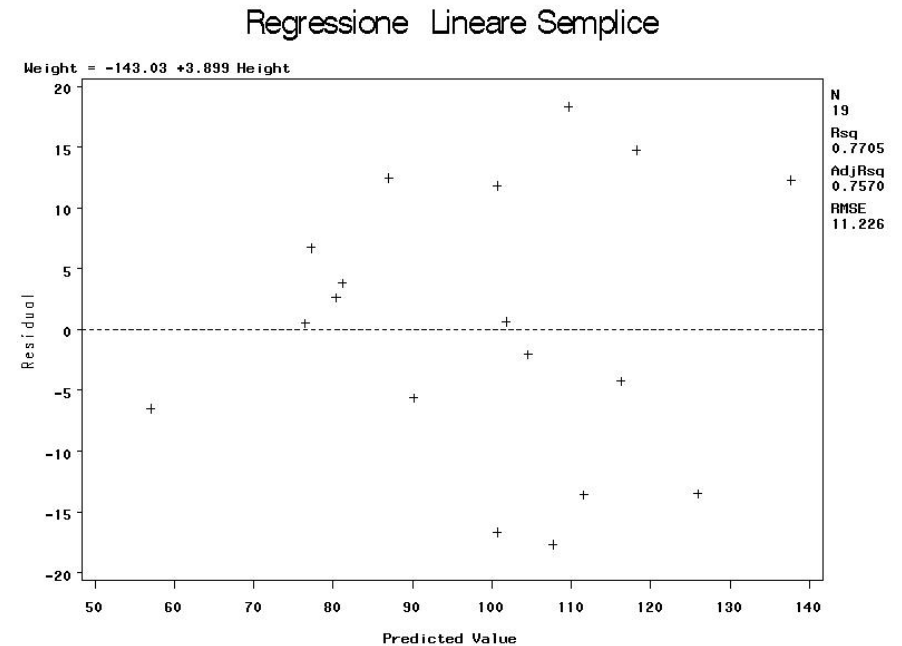
MODEL Y=X/R P;

SCATTERPLOT

PROC GPLOT;  
PLOT Y\*X;  
RUN;

## GRAFICI CON PROC REG

```
plot r.*p.;  
run;
```



NB:

La figura include alcune informazioni sul modello.