

## 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

120

## 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).

121

## 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

122

## 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

123

## 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'

124

## 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;
```

125

## 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](#)

126

## 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

127

## 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

128

## 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

129

## 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

130

## 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

131

## SELEZIONE DELLE OSSERVAZIONI: WHERE O SUBSETTING IF?

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

132

## 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	<b>WHERE</b>
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

133

## 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		

134

## 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;
```

135

## 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.

136

## 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.

137

## 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.!!!

138

## 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 *E<sub>w</sub>*

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;
```

139

## FORMATI DI SCRITTURA SAS

### Formati di data

DDMMYYw. (*ggmmaa*), MMDDYYw. (*mmggaa*), YYMMDDw. (*aammgg*): mese in numeri;

DATEw.: *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;
```

140

## 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.

141

```
*FORMATI.SAS;
*****
*** crea formati per dati valutazione (valuta1.sas) ***;
*****
*libname library 'd:\dip\computaz';
*proc format library=library; *creazione libreria permanente;
proc format fmlib; *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 scprovf 1='Liceo Classico'
2='Liceo Scientific.'
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;
```

142

### output di formati.sas (sintesi) opzione FMTLIB

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

FORMAT NAME: VOTOMAT LENGTH: 5 NUMBER OF VALUES: 5			
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		

143

```

*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 scprov. dom6 ancorsof.
      dom7 posizf. Dom5 Votomat.;
run;
title 'DATA STEP: associazione permanente formati' ;
data dati;
set dati.valuta;
format corso nomec. dom1 sesso. dom4 scprov. 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

```

DATA STEP: associazione permanente formati          43
15:42 Friday, May 14, 1999

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	NOMEC.	
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		