



Centro di Riferimento per l'Epidemiologia
e la Prevenzione Oncologica in Piemonte

www.cpo.it

Evento EFR N.35653

**Accesso e utilizzo del patrimonio informativo sanitario pseudonimizzato
tramite l'ambiente 'Clone DWH' su piattaforma SAS Enterprise Guide**

PROGRAMMAZIONE DI BASE IN AMBIENTE SAS SYSTEM E SAS ENTERPRISE GUIDE

Daniela Di Cuonzo

A.O.U. Città della Salute e della Scienza di Torino
S.S.D. Epidemiologia Clinica e Valutativa - CPO
C.P.O. Piemonte

Torino – 29 gennaio 2020

POTENZIALITA' DEL SISTEMA BASE

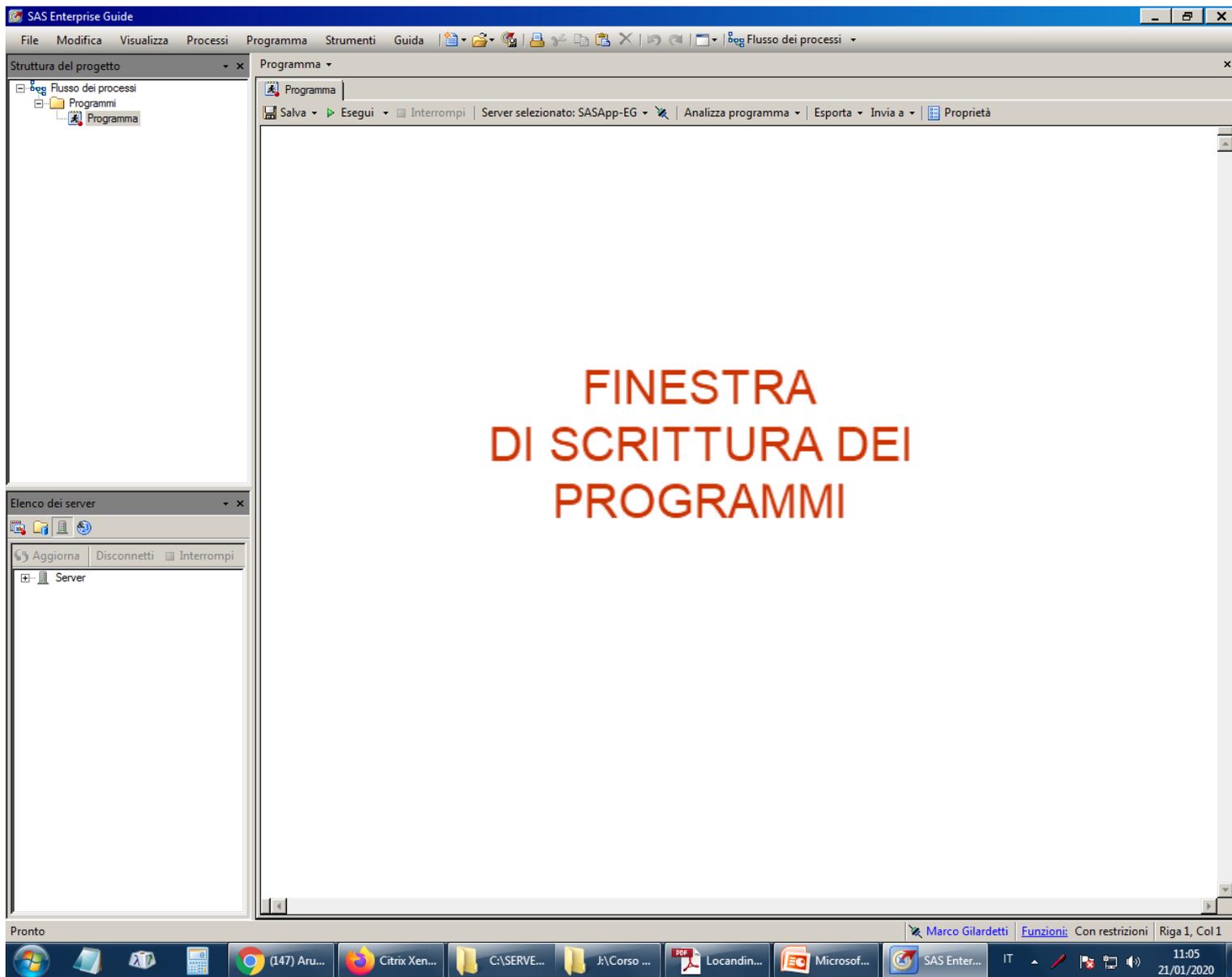
- MEMORIZZARE E RITROVARE INFORMAZIONI
- ELABORARE DATI CON SEMPLICI PROGRAMMI
- PRODURRE TABULATI STANDARD
- ANALIZZARE I DATI CON PROCEDURE DI USO GENERALE
- GESTIRE GLI ARCHIVI (ACCODARE, FONDERE, ...)

PACCHETTI SAS PER SETTORI SPECIFICI

- FINANZIARIO
- INDUSTRIA
- COMMERCIO E SERVIZI
- PUBBLICA AMMINISTRAZIONE E SANITA'



Le finestre di SAS Enterprise Guide (1)



Le finestre di SAS Enterprise Guide (2)

The screenshot displays the SAS Enterprise Guide interface. The main window shows a log of SAS code and its execution output. The code includes a macro call to submit a task to a grid node. The output shows the start of the task and various system messages.

**FINESTRA
DI COMMENTO ALLE
ISTRUZIONI E MESSAGGI
DI ERRORE**

```
1 The SAS System 11:06
1 rsubmit task1;
NOTE: Remote submit to TASK1 commencing.
1471 %put NOTE: utilizzo del grid node &SYSHOSTNAME per eseguire il processo.;
NOTE: utilizzo del grid node sasgridprod2 per eseguire il processo.
2 The SAS System 11:07

1472 %let _EGRC=0;
1473 ;*;*;*;/quit;run;
1474 OPTIONS PAGENO=MIN;
1475 %LET _CLIENTTASKLABEL='Programma';
1476 %LET _CLIENTPROJECTPATH='';
1477 %LET _CLIENTPROJECTNAME='';
1478 %LET _SASPROGRAMFILE=;
1479
1480 ODS _ALL_ CLOSE;
1481 OPTIONS DEV=ACTIVEVEX;
NOTE: Procedures may not support all options or statements for all devices. For details, see the documentat
1482 %let sasworkklocation="%_eg_gridworkpath/";
1483 GOPTIONS XPIXELS=0 YPIXELS=0;
1484 FILENAME EGSR "%_eg_gridworkpath/%sysfunc(uuidgen(0))_eggen";
1485 ODS tagsets.sasreport13(ID=EGSR) FILE=EGSR STYLE=HtmlBlue
1485! STYLESHEET=(URL="file:///c:/program%20files/sashome/sasenterpriseiguide/5.1/Styles/HtmlBlue.css") NOGT
1485! GPATH=%sasworkklocation ENCODING=UTF8 options(rolap="on");
NOTE: Writing TAGSETS.SASREPORT13(EGSR) Body file: EGSR
1486
1487 GOPTIONS ACCESSIBLE;
1488 *;
1489
1490
1491 GOPTIONS NOACCESSIBLE;
1492 %LET _CLIENTTASKLABEL=;
1493 %LET _CLIENTPROJECTPATH=;
1494 %LET _CLIENTPROJECTNAME=;
1495 %LET _SASPROGRAMFILE=;
1496
1497 ;*;*;*;/quit;run;
1498 ODS _ALL_ CLOSE;
1499
1500
1501 QUIT; RUN;
1502 %sysrput _EGRCGRID=&_EGRC;
NOTE: Remote submit to TASK1 complete.
```

The interface also shows a project structure on the left and a server list at the bottom left. The taskbar at the bottom includes various application icons and system information.

SAS

File Modifica Visualizza Strumenti Dati Soluzioni Finestra ?

VIEWTABLE: Sasuser.Navigators

	ID	LastName	FirstName	City	State	Gender	JobCode	Salary	Birth	Hired	HomePhone
1	1269	CASTON	FRANKLIN	STAMFORD	CT	M	NA1	41690.00	06MAY60	01DEC80	203/781-3335
2	1935	FERNANDEZ	KATRINA	BRIDGEPORT	CT		NA2	51081.00	31MAR42	19OCT69	203/675-2962
3	1417	NEWKIRK	WILLIAM	PATERSON	NJ	.	NA2	52270.00	30JUN52	10MAR77	201/732-6611
4	1839	NORRIS	DIANE	NEW YORK	YN	F	NA1	43433.00	02DEC58	06JUL81	718/384-1767
5	1111	RHODES	JEREMY	PRINCETON	NJ	M	NA1	40586.00	17JUL61	03NOV80	201/812-1837
6	1352	RIVERS	SIMON	NEW YORK	NY	M	NA2	5379.80	05DEC48	19OCT74	718/383-3345
7	1332	STEPHENSON	ADAM	BRIDGEPORT	CT	M	NA1	42178.00	20SEP58	07JUN79	203/675-1497
8	1443	WELLS	AGNES	STAMFORD	CT	F	NA1	422.74	20NOV56	01SEP79	203/781-5546

OSSERVAZIONE O UNITA' D'ANALISI (RIGA)

VARIABILE ALFANUMERICA (COLONNA)

VARIABILE NUMERICA (COLONNA)

Output - (Senza nome) Log - (Senza nome) Editor - Senza nome1 * Explorer VIEWTABLE: Sasuser.N...

NOTE: La tabella è stata aperta in modalità Browse. C:\Documents and Settings\Administrator

STRUTTURA DI UN DATASET SAS

Le finestre di SAS Enterprise Guide (3)

The screenshot displays the SAS Enterprise Guide interface. On the left, the 'Struttura del progetto' (Project Structure) pane shows a project named 'Corso CloneDWH SAS EG - Gilardetti - Formati e SASdate'. Below it, the 'Elenco dei server' (Server List) pane shows a tree view with 'Server' > 'SASApp-EG' > 'Librerie' > 'GRIDWORK' > 'LIB_EGSANCPO909_DBORA_cpo909in'. A red arrow points from this library name to the data table in the center. The data table has columns 'ID_ANONIMO_RIC' and 'PESO_BAMBINO_KG' and contains 48 rows of data. The bottom status bar shows the user 'Marco Gilardetti' and the date '21/01/2020'.

FINESTRA DI VISUALIZZAZIONE LIBRERIE

FILE NELLA LIBRERIA cpo909in.

Il database viene identificato con *cpo909in .nomefile*

LIBRERIA cpo909in è la directory in cui sono allocati in modalità di sola lettura i dati nativi

	ID_ANONIMO_RIC	PESO_BAMBINO_KG
1	117482403	3.65
2	117744609	2.93
3	115619809	2.076
4	117555363	2.85
5	117375031	3.72
6	117744396	3.52
7	115619550	3.13
8	117743897	2.285
9	117818663	3.61
10	117747549	3.56
11	116125863	2.82
12	117478082	3.59
13	117647133	2.66
14	115993685	2.96
15	115997139	3.74
16	117381339	3.27
17	115626152	3.28
18	117377015	3.4
19	115528444	2.93
20	115622920	3.65
21	117812578	3.37
22	117655628	3.65
23	117650721	3.92
24	117903602	2.75
25	117905068	3.93
26	117815441	3.51
27	117374327	2.99
28	117357488	4
29	116120406	3.42
30	115999151	3.23
32	117481497	1.739
33	117331640	3.13
34	117556655	3.445
35	117649062	3.441
36	117378775	3.125
37	117377088	3.21
38	117908529	3.52
39	117908005	3.38
40	117480756	3.59
41	117380849	3.29
42	117375958	2.91
43	117907574	2.77
44	116120615	2.75
45	115620712	2.7
46	115619479	2.77
47	117356445	3.39
48	117814451	2.92

Le finestre di SAS Enterprise Guide (4)

The screenshot displays the SAS Enterprise Guide interface. On the left, the 'Struttura del progetto' (Project Structure) pane shows a project named 'Corso CloneDWH SAS EG - Gilardetti - Formati e SASdate'. Below it, the 'Elenco dei server' (Server List) pane shows a tree view of servers, with 'LIB EG SANCP0909_DBORA_cpo909ou' highlighted. A red arrow points from this server to the data table. The main window shows a data table with columns 'ID_ANONIMO_RIC' and 'PESO_BAMBINO_KG'. The table contains 48 rows of data. The bottom status bar shows the user 'Marco Gilardetti' and the date '21/01/2020'.

FINESTRA DI VISUALIZZAZIONE LIBRERIE

FILE NELLA LIBRERIA cpo909ou.

Il database viene identificato con *cpo909ou .nomefile*

LIBRERIA cpo909ou è la directory su cui è possibile salvare i dati in modo permanente (questi non si cancellano con la chiusura del programma)

ID_ANONIMO_RIC	PESO_BAMBINO_KG	
1	117482403	3.65
2	117744608	2.93
3	115619809	2.076
4	117555363	2.85
5	117375031	3.72
6	117744396	3.52
7	115619550	3.13
8	117743897	2.285
9	117818663	3.61
10	117747549	3.56
11	116125863	2.82
12	117478082	3.59
13	117647133	2.66
14	115993685	2.96
15	115997139	3.74
16	117381339	3.27
17	115626152	3.28
18	117377015	3.4
19	115528444	2.93
20	115622920	3.65
21	117812578	3.37
22	117655628	3.65
23	117650721	3.92
24	117903602	2.75
25	117905068	3.93
26	117815441	3.51
27	117374327	2.99
28	117357488	4
29	116120406	3.42
30	115999151	3.23
31	117375147	3.72
33	117331640	3.13
34	117556655	3.445
35	117649062	3.441
36	117378775	3.125
37	117377088	3.21
38	117908529	3.52
39	117908005	3.38
40	117480756	3.59
41	117380849	3.29
42	117375958	2.91
43	117907574	2.77
44	116120615	2.75
45	115620712	2.7
46	115619479	2.77
47	117356445	3.39
48	117814451	2.92

Le finestre di SAS Enterprise Guide (5)

FINESTRA DI VISUALIZZAZIONE LIBRERIE

FILE TEMPORANEO NELLA LIBRERIA RTMWORK.

Il database viene identificato con *nomefile*

LIBRERIA RTMWORK temporanea che si “svuota” con la chiusura della sessione di lavoro

	ID_ANONIMO_RIC	PESO_BAMBINO_KG
1	117482403	3.65
2	117744608	2.93
3	115619809	2.076
4	117555363	2.85
5	117375031	3.72
6	117744396	3.52
7	115619550	3.13
8	117743897	2.285
9	117818663	3.61
10	117747549	3.56
11	116125863	2.82
12	117478082	3.59
13	117647133	2.66
14	115993685	2.96
15	115997139	3.74
16	117381339	3.27
17	115626152	3.28
18	117377015	3.4
19	115528444	2.93
20	115622920	3.65
21	117812578	3.37
22	117655628	3.65
23	117650721	3.92
24	117903602	2.75
25	117905068	3.93
26	117815441	3.51
27	117374327	2.99
28	117357488	4
29	116120406	3.42
30	115999151	3.23
31	117375147	3.72
32	117481497	1.739
33	117331640	3.13
34	117556655	3.445
36	117378775	3.125
37	117377088	3.21
38	117908529	3.52
39	117908005	3.38
40	117480756	3.59
41	117380849	3.29
42	117375958	2.91
43	117907574	2.77
44	116120615	2.75
45	115620712	2.7
46	115619479	2.77
47	117356445	3.39
48	117814451	2.92

PROGRAMMA
SAS

DATA Step

PROC Step

DATA Step

PROC Step

PROC Step

..... Step

CREA O MODIFICA UN
ARCHIVIO SAS

ANALIZZA E
PROCESSA I DATI
CONTENUTI IN UN
ARCHIVIO SAS

CREAZIONE DI UN ARCHIVIO SAS (1)



```
data sdo18;
```

```
set cpo909ou.sdo18; run;
```

```
/*Crea una copia temporanea del  
file sdo18 allocato nella libreria  
cpo909ou; il file viene salvato  
nella directory RTMWORK*/
```

ESTENSIONE DELLA DICHIARAZIONE SET

Permette di specificare più archivi SAS.

Variabili (con lo stesso nome) presenti su più archivi devono però essere compatibili per tipo e dimensione:

```
data tre;
```

```
set uno due;
```

```
run;
```

```
/* il dataset tre include le  
osservazioni di uno e quelle di  
due*/
```

TRATTAMENTO E MODIFICA DI VARIABILI (1)

Dichiarazione di assegnazione:

variabile = espressione;

```
X=3; /*assegno alla variabile numerica X il valore3*/  
A='abc'; /*assegno alla variabile alfanumerica A  
        la stringa 'abc'*/
```

Espressioni aritmetiche fondamentali:

```
+, -, *, /, ** /*elevamento a potenza*/
```

Funzioni matematiche:

```
X=SQRT(Y); /*radice quadrata*/
```

```
X=ABS(Y); /*valore assoluto*/
```

```
X=SUM(Y,Z); /*somma*/
```

```
X=MEAN(Y,Z); /*media*/
```

```
X=LOG(Y); /*logaritmo*/
```

TRATTAMENTO E MODIFICA DI VARIABILI (2)

Dichiarazione `length`:

definisce la lunghezza e il tipo di variabile da generare

```
data sdo18 ; set sdo18;  
length sesso $ 1;  
sesso='F';  
length gravita 3;  
gravita=6;  
run;
```

Funzione `length` (*variabile alfanumerica*):

restituisce la posizione del carattere diverso da “spazio” posto più a destra.

```
data sdo18 ; set sdo18;  
N=length(nome); /* ad es. a marco corrisponde N=5*/  
run;
```

TRATTAMENTO E MODIFICA DI VARIABILI (3)

Funzione `substr` (*variabile alfanumerica, posizione, n*):
estrae dalla stringa una sottostringa di n caratteri a partire da una posizione specificata.

```
data sdo18 ; set sdo18;  
codice= substr (cognome,1,2);  
run;
```

Operatore di concatenazione (`stringa1||stringa2`):

```
A='MARIO   ' ;  
B='ROSSI' ;  
VAR=A||B; /*==> VAR='MARIO   ROSSI'*/
```

Funzione `trim` (*variabile alfanumerica*):
elimina gli spazi alla destra della variabile .

```
VAR=TRIM(A)||B; /*==> VAR='MARIOROSSI'*/
```

TRATTAMENTO E MODIFICA DI VARIABILI (4)

Dichiarazione `drop` < lista di variabili > :

permette di eliminare le variabili che non si vogliono mantenere nell'archivio di uscita, anche se tali variabili sono presenti nel data set da elaborare e possono essere utilizzate per eventuali calcoli.

```
data sdo18 ;  
set sdo18;  
drop istituto sesso;  
run;
```

Dichiarazione `keep` < lista di variabili > :

specifica le variabili che si vogliono mantenere nell'archivio di uscita (la scelta fra le due dichiarazioni è solo una questione di risparmio di scrittura).

```
data sdo18 ;  
set sdo18;  
keep eta istituto;  
run;
```

Dichiarazione `rename` <vecchio nome>= <nuovo nome>
cambia il nome ad una o più variabili.

ISTRUZIONI LOGICHE E CONDIZIONALI (1)

Dichiarazione IF-THEN-ELSE:

```
if <espressione condizionale>  
then <dichiarazione eseguibile>;  
else <dichiarazione eseguibile>;
```

Se l'espressione condizionale risulta vera
allora si esegue la dichiarazione che segue il `then`
altrimenti si esegue la dichiarazione che segue la
parola `else`

Per scrivere un'espressione condizionale si possono
utilizzare gli operatori di confronto:

= , < , <= , > , >= , NE (diverso da)
gli operatori logici: OR , AND, NOT
e le parentesi.

ISTRUZIONI LOGICHE E CONDIZIONALI (2)

```
if A<B OR C=10 /*se ALMENO UNO DEI DUE confronti  
risulta vero verrà eseguita l'istruzione1, altrimenti  
verrà eseguita l'istruzione2*/  
then istruzione1;  
else istruzione2;
```

```
if A<B AND C=10 /*se ENTRAMBI i confronti risultano  
veri verrà eseguita l'istruzione1, altrimenti verrà  
eseguita l'istruzione2*/  
then istruzione1;  
else istruzione2;
```

```
if NOT( stringa='abc' ) /*verrà eseguita l'istruzione1 SE  
NON RISULTA VERA LA CONDIZIONE, altrimenti verrà  
eseguita l'istruzione2*/  
then istruzione1;  
else istruzione2;
```

ISTRUZIONI LOGICHE E CONDIZIONALI (3)

Nel caso di blocchi costituiti da più istruzioni:

```
if <espressione condizionale>  
then do;  
<dichiarazione eseguibile1>;  
<dichiarazione eseguibile2>;  
.....;  
                end;  
  
else do;  
<dichiarazione eseguibile1>;  
<dichiarazione eseguibile2>;  
.....;  
                end;
```

SELEZIONE/RIDUZIONE DELLE OSSERVAZIONI (1)

Dichiarazione IF subsetting:

```
if <espressione condizionale>;
```

Conserva solo le osservazioni per cui l'espressione condizionale risulta vera e non deve essere seguita dal `then`.

```
if sesso='F'; /* il dataset include solo le  
              osservazioni delle femmine */
```

Dichiarazione `delete`:

```
if <espressione condizionale> then delete;
```

Elimina le osservazioni per cui l'espressione condizionale risulta vera.

```
if sesso='F' then delete; /* il dataset non  
contiene le osservazioni delle femmine */
```

SELEZIONE/RIDUZIONE DELLE OSSERVAZIONI (2)

La variabile automatica `_N_`:

E' generata dal sistema SAS ed è utilizzabile all'interno di un passo di data ma non viene riportata sugli archivi in uscita. Identifica la posizione dell'osservazione all'interno del dataset.

```
if _N_ <= 5; /* il dataset include solo le  
prime 5 osservazioni*/
```