



Statistica computazionale

a.a. 2008/09

Appunti lezione del 2/3/09

Carla Rampichini
rampichini@ds.unifi.it



Creazione SAS data set da dati esterni

- Istruzioni SAS nel passo di DATA;
- External File Interface (EFI);
- Import Wizard



Creazione di SAS Data set da file ASCII

- Istruzione INFILE

DATA *dataset*;

INFILE '*percorso e nome file*' [*opzioni*]

INPUT *lista*;

....

vedi `valuta1.sas`

Osservare la finestra LOG



Esempio lettura a lista da file esterno

Il file ASCII `esempio_lista.txt` contiene i dati relativi a un'indagine sui laureati dell'ateneo fiorentino a circa 2 anni dalla laurea:

- Sesso (1=maschio, 0=femmina)
- Voto di maturità
- Voto medio agli esami
- Età alla laurea
- Codice corso di laurea
- Lavora stabilmente (1=sì, 0=no)

`esempio_lista.sas`

Esercizio 1:

lettura dati da programma

- Scrivere un programma SAS che generi un SAS data set temporaneo di nome DATI e che contenga i dati relativi a anno di nascita, sesso e titolo di studio di un gruppo di 10 studenti (vedi lucido successivo);
- Creare la variabile età, sapendo che i dati sono stati rilevati nel 2003;
- Stampare il contenuto dell'archivio;
- Stampare il contenuto della DIA.

Dati esercizio 1

- Anno di nascita
- Sesso (1=M, 2=F)
- Titolo di studio
 - 1=liceo classico
 - 2=liceo scientifico
 - 3=istituto tecnico
 - 4=istituto professionale
 - 5=altro

1982	2	5
1983	2	3
1983	1	3
1981	2	2
1981	1	2
1983	1	1
1984	2	2
1983	1	2
1983	1	2
1983	1	3
1984	1	5
1983	2	1
1983	1	2
1983	1	3
1976	1	2
1975	1	2
1983	2	3
1981	1	2
1980	2	5

Esercizio 2:

lettura a lista da file ASCII

- Creare un archivio permanente SAS dell'archivio in formato testo v:\carla\sas\studenti.txt;
- L'archivio contiene 4 variabili numeriche: leggere solo le prime 3;
- Stampare il contenuto dell'archivio.

Lettura file tipo .xls

Import wizard consente di leggere in maniera quasi automatica file di vario tipo (.csv, delimited, .txt, .xls)

- Dal menu **FILE** scegliere **IMPORT DATA**
- Nel campo **data source** scegliere Microsoft Excel 97 or 2000 (*.xls)
- Selezionare il file di interesse
- Assegnare un nome al sas data set su cui scrivere
- [creare file contenente istruzioni PROC IMPORT]
- Premere **finish**



Esercizio 3

- Importare i dati contenuti nel file STUDENTI.XLS, salvandoli nell'archivio temporaneo STUDENTI
- Scrivere un passo di DATA che generi la variabile ETA, contenente l'età degli studenti, sapendo che i dati sono stati rilevati nel 2003.



Regole per l'importazione di file CSV

- (1) I file EXCEL o altri files possono essere salvati in formato CSV (Comma Separated Values)
- (2) Assicurarsi che nelle impostazioni della macchina raggiungibili tramite:

Menù Start→**Impostazioni**→**Pannello di controllo**

Opzioni internazionali→**Numeri**

come **separatore di elenco** sia impostata la **virgola [,]**, mentre come **separatore decimale** sia impostato il **punto [.]**.



Lettura file tipo .csv

- Dal menu FILE scegliere IMPORT DATA
- Nel campo data source scegliere comma separated values (.csv)
- Selezionare il file di interesse
- Assegnare un nome al sas data set su cui scrivere
- Premere finish



Esercizio 4

- Importare i dati contenuti nel file V:\CARLA\SAS\KAZUAKI.CSV, salvandoli nell'archivio temporaneo KAZUAKI
- Stampare il contenuto della DIA;
- Stampare i dati.



Input a colonna

INPUT var [\$] col1[-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 osservazioni
- Variabili in forma numerica o carattere standard



Vantaggi input a colonna


- non occorre che i campi corrispondenti alle variabili siano separati da *blanks*;
- non è necessario indicare esplicitamente i valori mancanti (*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



Esercizio 5: lettura a colonna

- 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.
- Creare un archivio SAS permanente a partire dai dati contenuti nel file ASCII 'v:\didat\carla\sas\biscotti.txt'.
- Creare una variabile che contiene, per ciascuna osservazione, la media dei punteggi.



Variabili e posizione nel file v:\carla\sas\biscotti.txt

- nome 1-13 nome dei biscotti
- anno 14-15 anno inizio produzione
- tipo 16-17 al cioccolato (sì o no)
- data 18-27 giorno della rilevazione
- punti1 29-31 punteggio 1° assaggiatore
- punti2 33-35 punteggio 2° assaggiatore
- punti3 37-39 punteggio 3° assaggiatore
- punti4 41-43 punteggio 4° assaggiatore
- punti5 45-47 punteggio 5° assaggiatore



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.



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:

- Il primo record contiene la città e lo stato;
- Il secondo la temperatura media min e max;
- Il terzo la temperatura "record" min e max.

`piurec.sas`