

# Introduzione a SAS Base

*Federico Viscioletti*

*12 novembre 2015*

In questo articolo inizieremo ad esplorare il mondo SAS, iniziando dal mattone fondamentale: il linguaggio di programmazione.

## Introduzione

SAS, acronimo che sta per **Statistical Analysis System**, è uno degli strumenti software per condurre analisi su dati più utilizzati. L'azienda, ha sede a Cary (North Carolina), produce una grande varietà di strumenti software, ma ognuno di questi si basa su un motore di calcolo: SAS Base.

Questi è un vero e proprio linguaggio di programmazione. Tramite SAS Base è infatti possibile rendere automatica, e quindi rieseguibile, qualsiasi analisi statistica. Il linguaggio si divide in due elementi logicamente ben distinti fra loro:

- **Data Step** - Permette di effettuare operazioni di ETL e data manipulation. Queste comprendono operazioni di estrazione, trasformazione e caricamento di dati;
- **Proc Step** - Incapsula tutte le procedure SAS (proc è infatti l'abbreviazione di procedure). Tra le procedure più utilizzate ci sono ad esempio la PROC PRINT, che permette di riportare a video i primi n record di un dataset; la PROC SQL, che permette di eseguire codice SQL all'interno dell'ambiente SAS e così via.

## Interfaccia

L'interfaccia grafica di SAS Base è minimale e pulita. Presenta due pannelli principali, il primo sulla sinistra, necessario ad esplorare le risorse provenienti da librerie SAS o dal file system. Nel secondo invece, quello che occupa la maggior parte dello schermo, vengono precaricate tre viste:

- **Output** - In questo pannello vengono riportati i report, siano essi grafici o tabellari, delle procedure eseguite.
- **Log** - Consente di verificare la corretta esecuzione delle procedure lanciate
- **Editor** - Si tratta di un'editor di testo intelligente, dotato di alcune funzionalità comode come il syntax highlighting. Qui è possibile inserire il codice SAS da poter poi eseguire.

## Un esempio

Iniziamo ora a mettere mano al codice, sviluppando un piccolo esempio che ci sarà utile per esplorare le prime funzionalità del **data step**. Il seguente frammento di codice permette infatti di caricare un dataset

```
libname registry "data";  
  
data registry.people;
```

```

input id name $ surname $ sex $ age weight height;
datalines;
1 Julia Smith F 29 56 172
2 Mark Ronson M 47 78 182
3 Patrick Lane M 33 69 177
4 Allie White F 45 62 189
5 Owen Peterson M 39 71 176
;
run;

```

Una volta eseguito il codice il dataset SAS People\_registry è disponibile e può essere visualizzata cliccando sul pannello di sinistra al percorso Librerie > Work. Qui troveremo la nostra tabella SAS caricata e visibile.

Esercitiamoci ora anche nell'utilizzo di **variabili calcolate** all'interno di un data step. Calcoliamo dunque il Body Mass Index (Indice di Massa Corporea) di ciascun soggetto, che si esprime come

$$BMI = \frac{mass_{kg}}{height_m^2}$$

Ecco dunque che il seguente estratto di codice SAS prende in input il dataset *people\_registry* e crea una nuova variabile calcolata *bmi*, che calcola il valore espresso dalla precedente formula.

```

libname registry "data";

data registry.people;
  set registry.people;
  bmi = weight / ( ( height / 100 ) ** 2 );
run;

```