



Next: [guessingrows](#) **Up:** [Reading in datasets](#) **Previous:** [Reading in datasets](#)

Quick and dirty: convert proc import log output to an input statement

Reading in CSV or otherwise delimited text files using proc import does not give you the control you want over how the variables are defined, but using an input statement means you have to type out all the (possibly tens of) variable names and format definitions. It also means that you have to know all your variable names and what exactly SAS's name for the delimiter is and a whole host of other things.

A useful way to have control while not doing too much work is to use proc import on a small subsample first, and then use what proc import writes to the log to construct your own input statement, changing variable names or formats as you like.

For instance, I have the MSF dataset stored as a CSV file: text.csv. The first two lines are

```
PERMNO,DATE,RET
10000,19860228,-0.257143
```

A proc import would read:

```
proc import file='text.csv' out=out dbms=csv;
run;
```

and produce the following in the log file:

```
11          data WORK.OUT                                ;
12          %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
13          infile 'text' delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2 ;
14              informat PERMNO best32. ;
15              informat DATE best32. ;
16              informat RET best32. ;
17              format PERMNO best12. ;
18              format DATE best12. ;
19              format RET best12. ;
20          input
21              PERMNO
22              DATE
23              RET
24          ;
25          if _ERROR_ then call symputx('_EFIERR_',1); /* set ERROR detection macro variable */
26          run;
```

Notice that the date variable is stored as a number, and it would have been a hassle to get it into the correct format.

But working from this log, it is trivial to produce the following clean input data step:

```
data WORK.OUT;
infile 'text' delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2;
informat PERMNO best32. ;
informat DATE YYMMDD8. ;
informat RET best32. ;
format PERMNO best12. ;
format DATE date9. ;
format RET best12. ;
```

```
input
  PERMNO
  DATE
  RET
;
run;
```

Notice the change in the informat and format for date, so that the output dataset has date correctly coded.

Next: [guessingrows](#) **Up:** [Reading in datasets](#) **Previous:** [Reading in datasets](#)

Andre de Souza 2012-11-19