

Week 2 - Three ways to get your data into SAS

Importing data into SAS is the first step to creating your program from analysis. There are three main ways to do this. The first way is useful for very small datasets (like the ones you might get in your take-home exam, for example). The second and third ways are better for larger data sets. It is more likely that these are already in Excel or some other database, and so we import from those existing files.

1. Using the CARDS command. For this method you type your data straight into the editor window. We will use a fictional data set (the body weights of children in two US states). Here we have separated each column with a space.

```
Title 'Body weights of children';
data fatkids;
input state $1-2 obs 4 weight 6-8;
cards;
ny 1 125
ny 2 115
ny 3 137
la 1 315
la 2 225
la 3 145;
run;
```

2. Using the INFILE statement. This imports your data from an existing file. Again, you need to tell the program what each column is called using Input.

```
Title 'Body weights of children';
data fatkids;
infile 'C:\Documents and Settings\user\My Documents\fatkids.egc';
Options nocenter nonumber nodate ls=72 ps=66;
Input state $1-2 obs 4 weight 6-8;
run;
```

3. Using the PROC IMPORT command. This imports from an existing file and allows you to bring column names across using getnames, if your dataset already has them in the first row.

```
Title 'Body weights of children';
data fatkids;
proc import datafile='C:\Documents and Settings\user\My Documents\fatkids.csv'
out=fatkids dbms=csv replace;
getnames=yes;
run;
proc print data=fatkids;
run;
```

Note that in this code the first semicolon after proc import is after the replace statement. Proc print is an optional extra, it is not a necessary part of proc import, and you could end with the first 'run' if you chose.

Explanations of options statements and other common statements:

```
nocenter – left justifies the output
nonumber – page headers in output are not numbered
nodate – date not listed in page headers
```

linesize (ls) – number of columns in the output window
pagesize (ps) – number of rows in the output window
dbms – database management system. Tells SAS what kind of delimiter to expect
replace – overwrites previous file with this name if it exists
getnames – tells SAS to use whatever is in row 1 of the imported dataset as the
column names
proc print – the data will be shown (printed) in the output window.