

Importing an Excel Worksheet into SAS

Preparing Data for a Statistics Package: From Excel

How to Set up the Excel File:

- Place the variable names in the first row. Be sure the names follow these rules:
 - variable names can be no more than **8 characters** long
 - variable names must **start with a letter**
 - variable names may only have letters, numbers, or underscores in them
 - do not use following characters in variable names:
%, \$, #, @, !, +, *, ~, ", ., -, .
 - **no blanks** in variable names
 - be sure that each variable name is **unique** (no duplicate variable names)
 - be sure variable names are on **the first row only!**
- Only include the **raw, un-summarized data**. Delete extraneous data in your Excel file, like row or column totals, graphs, comments, annotations, etc. To prevent "ghost" rows and columns, copy only the raw data onto a new worksheet, and save from there.
- Include a **unique identifying number for each case**. Sometimes you may have more than one identifier, such as Household ID and Subject ID; place these in separate columns. If you have several spreadsheets containing data on the same individuals, include their identifier(s) on each sheet.
- Only include **one value per cell**. Don't enter data such as "120/80" for blood pressure. Enter systolic blood pressure as one variable, and diastolic blood pressure as another variable. Don't enter data as "A,C,D" or "BDF" if there are three possible answers to a question. Include a separate column for each answer.
- Don't leave blank rows or columns in the data.
- Don't mix numeric and character values (e.g. names and ID numbers) in the same column.
- While character variables are allowed in statistical packages, they are not as flexible as numeric variables, which are preferred. **Use numeric values when feasible**.
- Date values are best entered in three columns: one for month, one for day, one for year. You can change them into date values in your statistics package later.
- If you have missing values, you can indicate them with a numeric code, such as 99 or 999, or you can leave the cell blank. Be sure, if you use a missing value code, that it cannot be confused with a "real" data value.
- Save the spreadsheet with **values only**, not formulas.
- Do not underline text, or use boldface or italics.

An excerpt from an ideal Excel data set might look like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FLIGHT	DATE	TIME	ORIG	DEST	MILES	MAIL	FREIGHT	BOARDING	TRANSFER	NONREV	DEPLANE	CAPACITY
2	182	3/1/1990	8:21	LGA	YYZ	366	458	390	104	16	3	123	178
3	114	3/1/1990	7:10	LGA	LAX	2475	357	390	172	18	6	196	210
4	202	3/1/1990	10:43	LGA	ORD	740	369	244	151	11	5	157	210
5	219	3/1/1990	9:31	LGA	LON	3442	412	334	198	17	7	222	250
6	439	3/1/1990	12:16	LGA	LAX	2475	422	267	167	13	5	185	210
7	387	3/1/1990	11:40	LGA	CPH	3856	423	398	152	8	3	163	250
8	290	3/1/1990	6:56	LGA	WAS	229	327	253	96	16	7	117	180
9	523	3/1/1990	15:19	LGA	ORD	740	476	456	177	20	3	185	210
10	982	3/1/1990	10:28	LGA	DFW	1383	383	355	49	19	2	56	180
11	622	3/1/1990	12:19	LGA	FRA	3857	255	243	207	15	5	227	250
12	821	3/1/1990	14:56	LGA	LON	3442	334	289	205	13	4	222	250
13	872	3/1/1990	13:02	LGA	LAX	2475	316	357	145	13	5	163	210

How to Save the Excel File:

Version 4.0 Worksheets can be read by most statistical packages. To save your Excel file in version 4.0, go to the File menu and choose **Save As...** and then **select Excel 4.0 Worksheet (not Workbook)** as the file type. You will be able to save only one worksheet at a time in Excel 4.0 format. To preserve your original Excel data, use a different name when saving in this special format. To be sure that the file name will be easily recognizable on any system, use a name not longer than eight characters, and add the extension **.xls**. If you have several worksheets, they will need to be saved individually and merged in the statistical package you are using.

The consultants at CSCAR can help you with this.

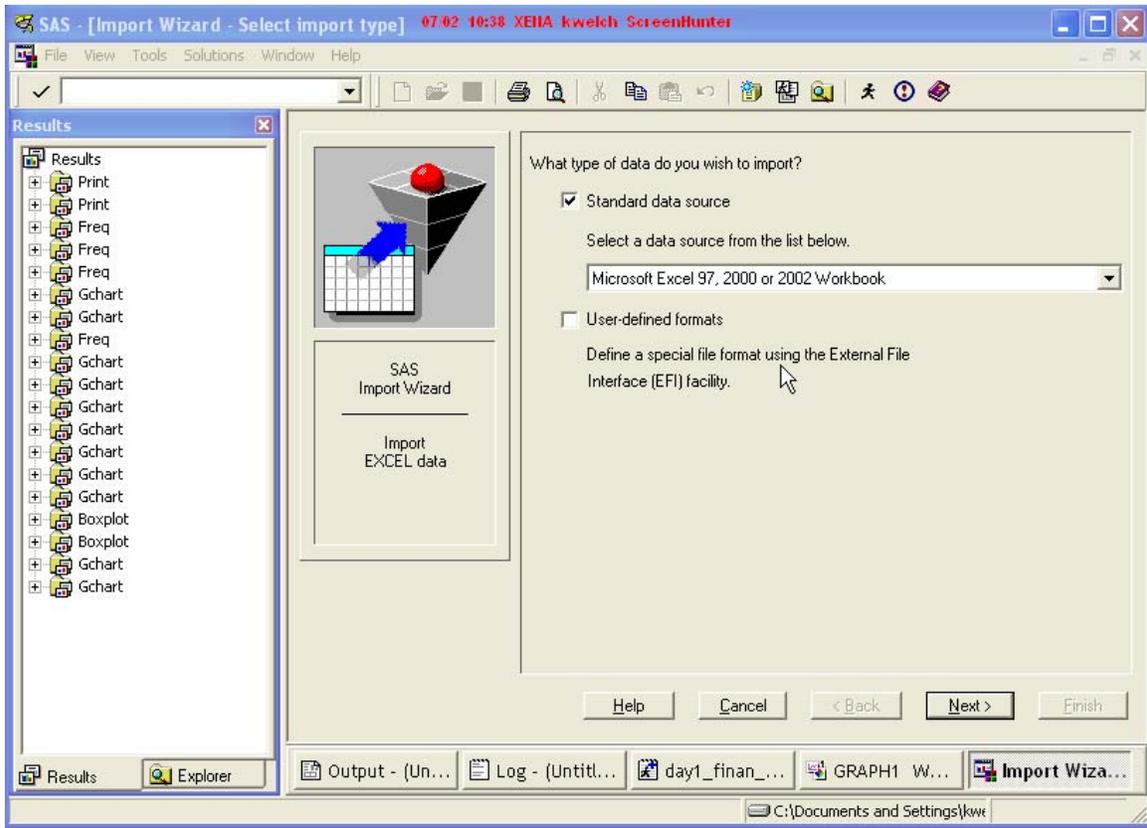
A document very similar to this one is available online at <http://www.umich.edu/~cscar/software/fromexcel.html>

What Type of Excel Files Can You Import?

You can import Excel worksheets, starting with very early versions of Excel (e.g., Excel version 4.0). You can also import individual worksheets from workbooks for later versions of Excel (e.g. Excel 2000), but only one worksheet at a time. The most recent versions of Excel cannot be opened automatically by SAS, and will have to be saved as an earlier version before proceeding.

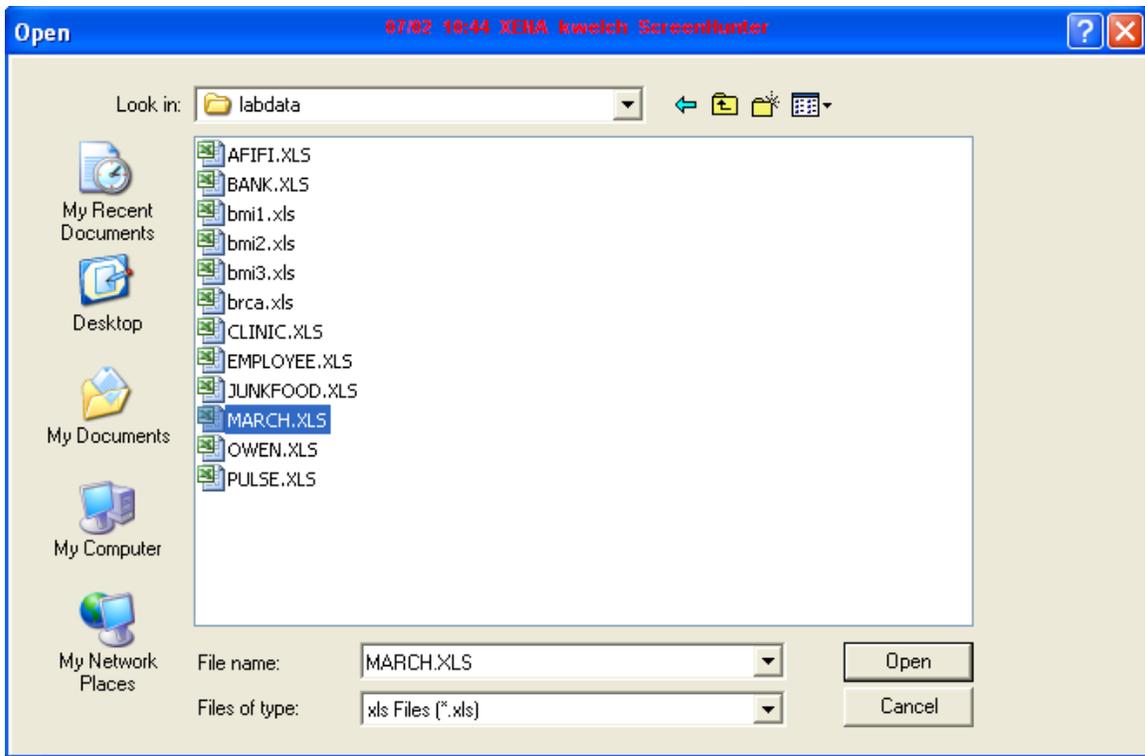
Step-By-Step Instructions:

Go to the File Menu and select Import Data...Select the type of data file that you would like to import from the pull-down menu.

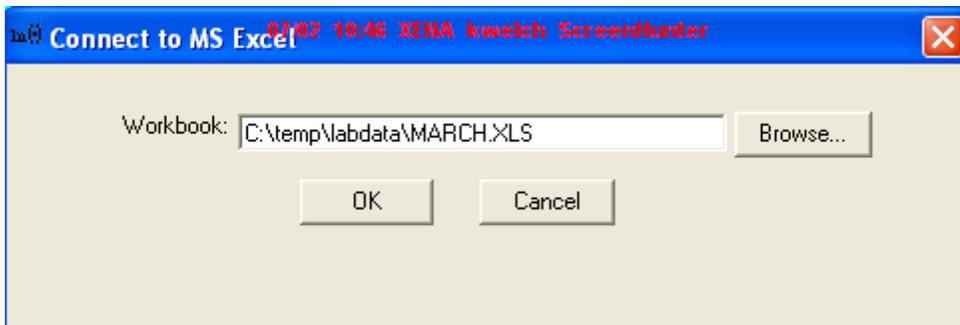


Click on the “Next>” button to proceed.

In the dialog box that opens, browse to the Excel file that you wish to open, and click on the “Open” button.



The filename that you have chosen will appear in the browse dialog box.



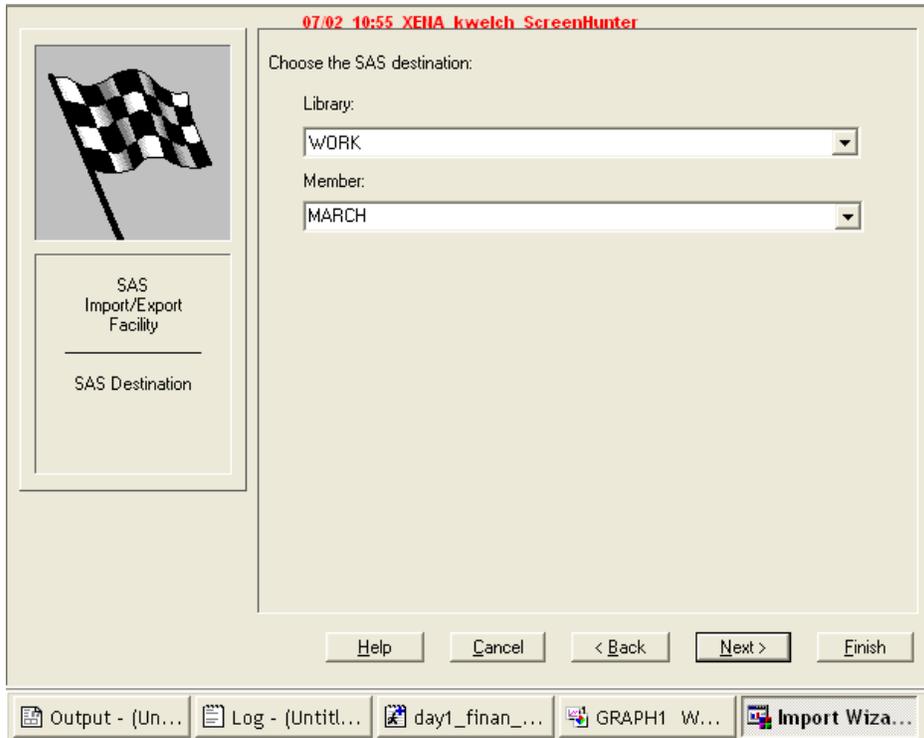
Click on “OK”.

In the next dialog box, you will need to select the table that you want to import from the pulldown list. In this example, we are selecting the table named “march”, which is in fact, the only table in this workbook.



Click on “Next>” to proceed.

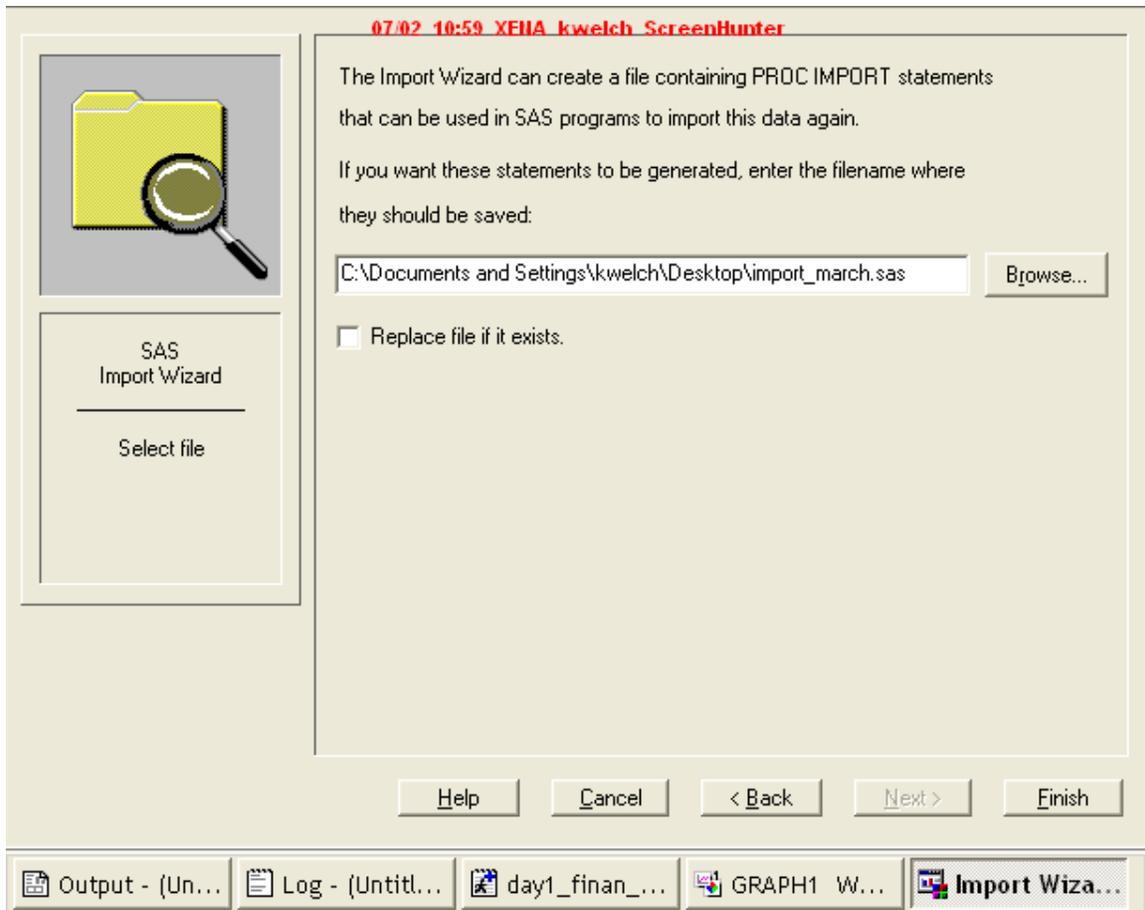
At this point, you will be taken to a dialog box that allows you to save the SAS data set to a library. The default temporary library “WORK” will be automatically filled in for you, but you need to type the data set name. In this case, we are saving the data set as WORK.MARCH.



At this point, you have two choices for what to do.

- If you click on “Finish”, the data set will be saved, and you can proceed to work with it.
- If you click on “Next>”, you will go to the following dialog box, where you will have a chance to save SAS commands to be used to import the data set at a later time.

I usually click on “Next>”, so I can save my commands. This process is shown below:



You can now click on “Finish” to complete importing the data set.

If you saved your commands to import the Excel file, you can bring them into your SAS enhanced editor, by going to File...Open Program... and browsing to the command file that you saved. The command file is shown below:

```
PROC IMPORT OUT= WORK.MARCH
            DATAFILE= "C:\temp\labdata\MARCH.XLS"
            DBMS=EXCEL REPLACE;
SHEET="march";
GETNAMES=YES;
MIXED=NO;
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;
RUN;
```

The data set can be modified by creating a new data step, with additional commands, for example:

```

data march2;
  set march;
  totpassngrs = boarded + transfer + nonrev;
  empty_seats = capacity - totpassngrs;
  totnonpass = mail + freight;
  pctfull = (totpassngrs/capacity)*100;
run;

```

SAS can now be used to run procedures on this data set:

```

proc means data = march2;
run;

```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
DATE	DATE	634	11031.98	8.9801263	11017.00	11047.00
MILES	MILES	635	1615.25	1338.47	229.0000000	3857.00
MAIL	MAIL	634	381.0031546	74.6288128	195.0000000	622.0000000
FREIGHT	FREIGHT	634	333.9511041	98.1122248	21.0000000	631.0000000
BOARDED	BOARDED	633	132.3570300	43.4883098	13.0000000	241.0000000
TRANSFER	TRANSFER	635	14.4062992	5.3362008	0	29.0000000
NONREV	NONREV	635	4.1133858	1.9243731	0	9.0000000
DEPLANE	DEPLANE	635	146.7842520	45.4289656	18.0000000	250.0000000
CAPACITY	CAPACITY	635	205.3795276	27.1585929	178.0000000	250.0000000
totpassngrs		633	150.8878357	43.0930520	31.0000000	250.0000000
empty_seats		633	54.5244866	34.9192529	0	151.0000000
totnonpass		633	715.1927330	124.8981261	341.0000000	1085.00
pctfull		633	73.0774908	17.7696598	17.2222222	100.0000000