

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.**

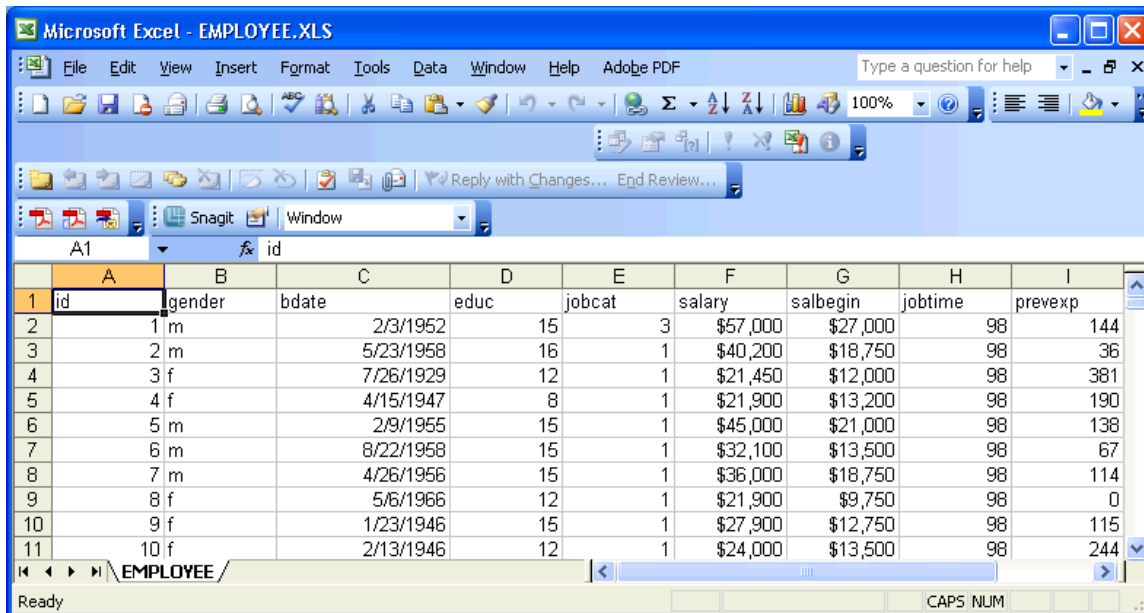
Date values are best entered in three columns: one for month, one for day, one for year. You can change them into date values in your statistics package later.

If you have missing values, you can indicate them with a numeric code, such as -1, 99 or 999, or you can leave the cell blank. Be sure the value you use to indicate missingness 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.

A suitable excerpt from an Excel data sheet might look like this:



The screenshot shows a Microsoft Excel window titled "EMPLOYEE.XLS". The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H	I
1	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp
2	1	m	2/3/1952	15	3	\$57,000	\$27,000	98	144
3	2	m	5/23/1958	16	1	\$40,200	\$18,750	98	36
4	3	f	7/26/1929	12	1	\$21,450	\$12,000	98	381
5	4	f	4/15/1947	8	1	\$21,900	\$13,200	98	190
6	5	m	2/9/1955	15	1	\$45,000	\$21,000	98	138
7	6	m	8/22/1958	15	1	\$32,100	\$13,500	98	67
8	7	m	4/26/1956	15	1	\$36,000	\$18,750	98	114
9	8	f	5/6/1966	12	1	\$21,900	\$9,750	98	0
10	9	f	1/23/1946	15	1	\$27,900	\$12,750	98	115
11	10	f	2/13/1946	12	1	\$24,000	\$13,500	98	244

How to Save the Excel File:

SAS 9.1 and 9.2 cannot read .XLXS files, used by Microsoft Excel 7.0. If you have Excel 7.0, you need to save the file as an **.XLS** file before you can read it into SAS.

SAS can read.XLS files (Excel 5/95/ 97/2000/2002/2003 Workbooks).

If you are experiencing trouble, you can try to save each spreadsheet as an **Excel Version 4.0 Worksheet**. 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, be sure to save the file using a different name.

To recognize the file as an Excel file, SAS will need it to have the **.xls** extension.

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?

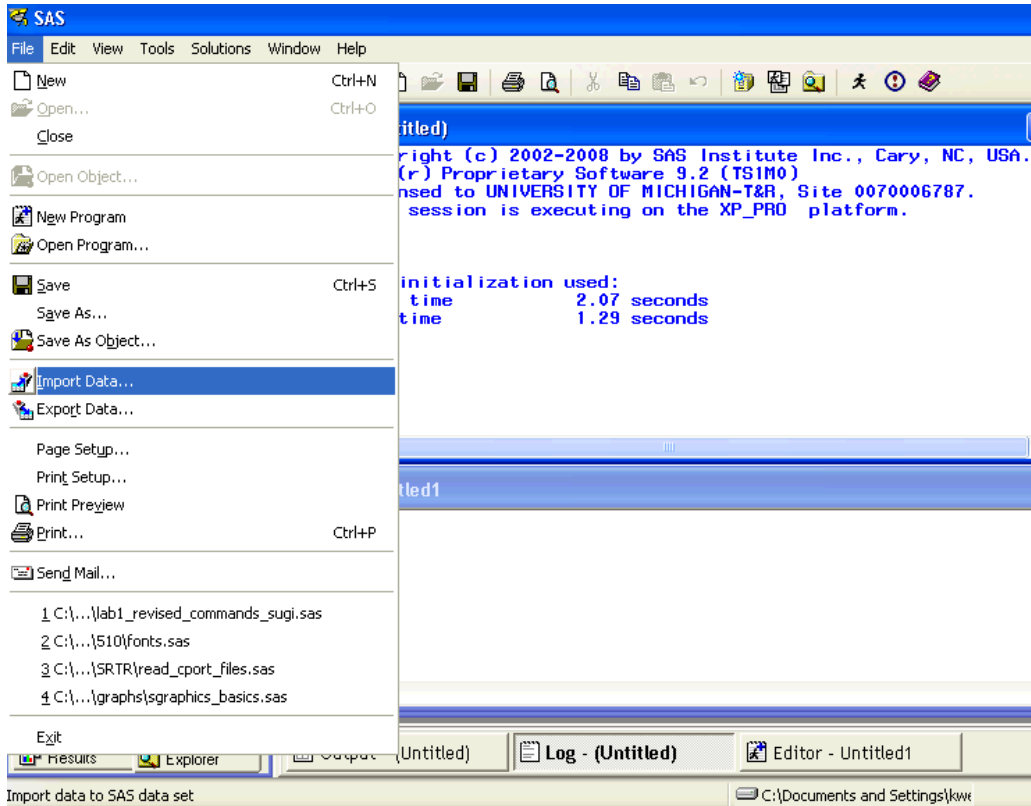
SAS can import .XLS files (Excel 5/95/ 97/2000/2002/2003 Workbooks). If you have an Excel Workbook, you need to import each sheet separately, and then merge them in SAS, if necessary.

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.

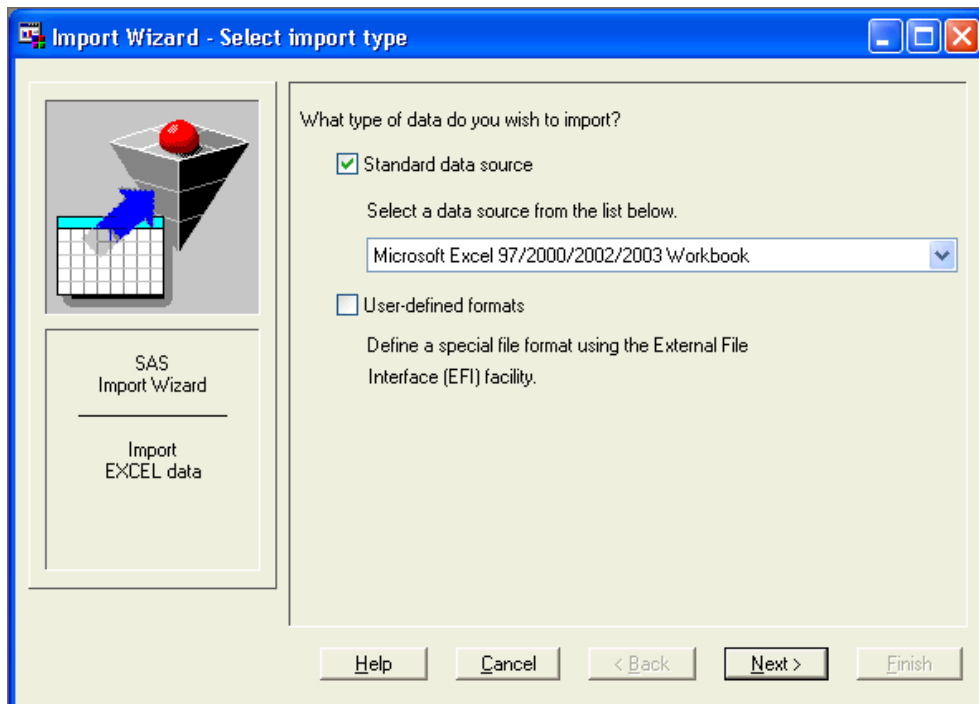
The most recent versions of Excel, **.XLXS files, cannot be opened automatically by SAS**, and will have to be saved as an earlier (**.XLS**) version before proceeding.

Step-By-Step Instructions Using the Import Wizard:

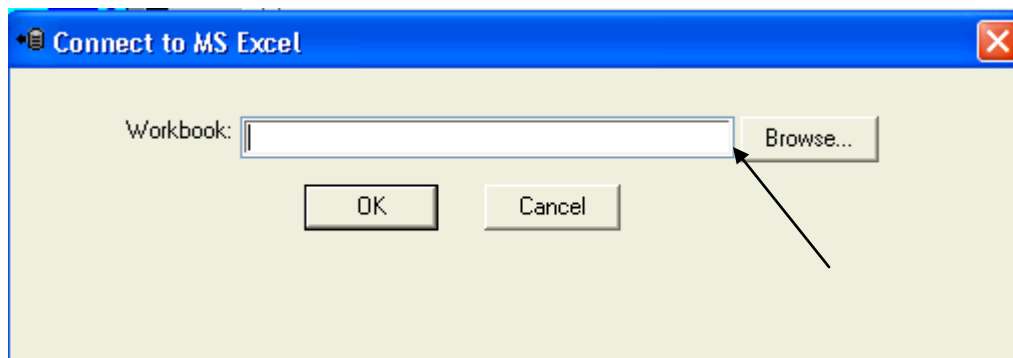
Go to the File Menu and select Import Data...



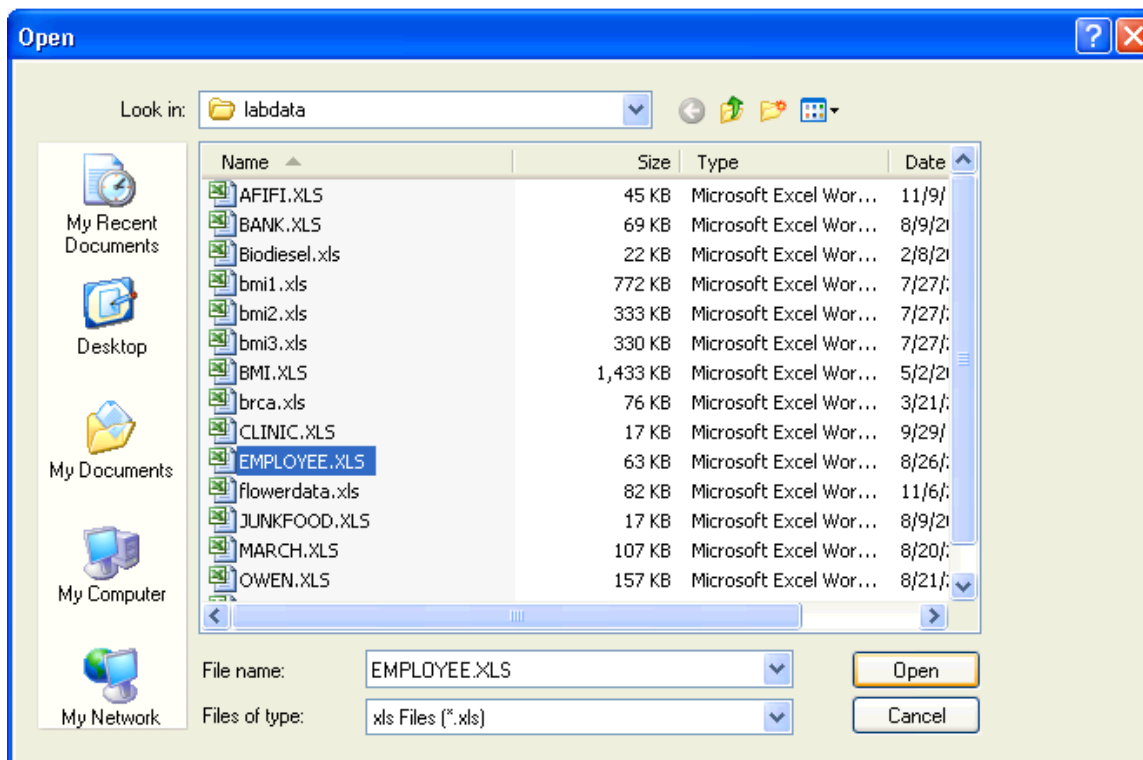
Select the type of data file that you would like to import from the pull-down menu, and click on "Next".



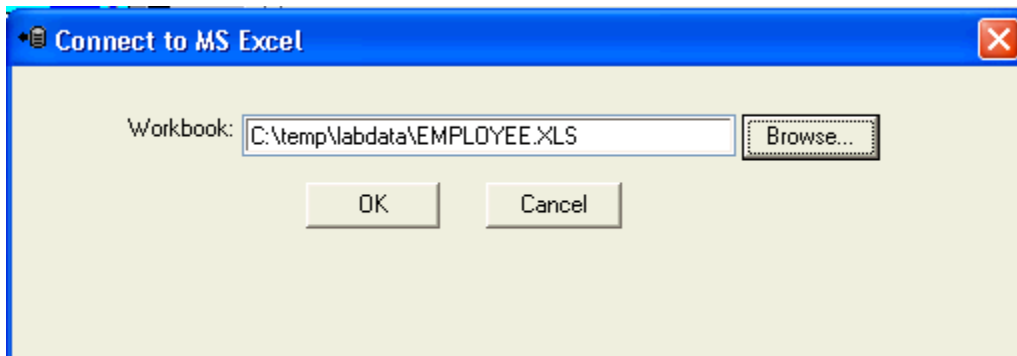
In the dialog box that opens, you can either type the filename that you wish to import or **Browse...** to the location of the file.



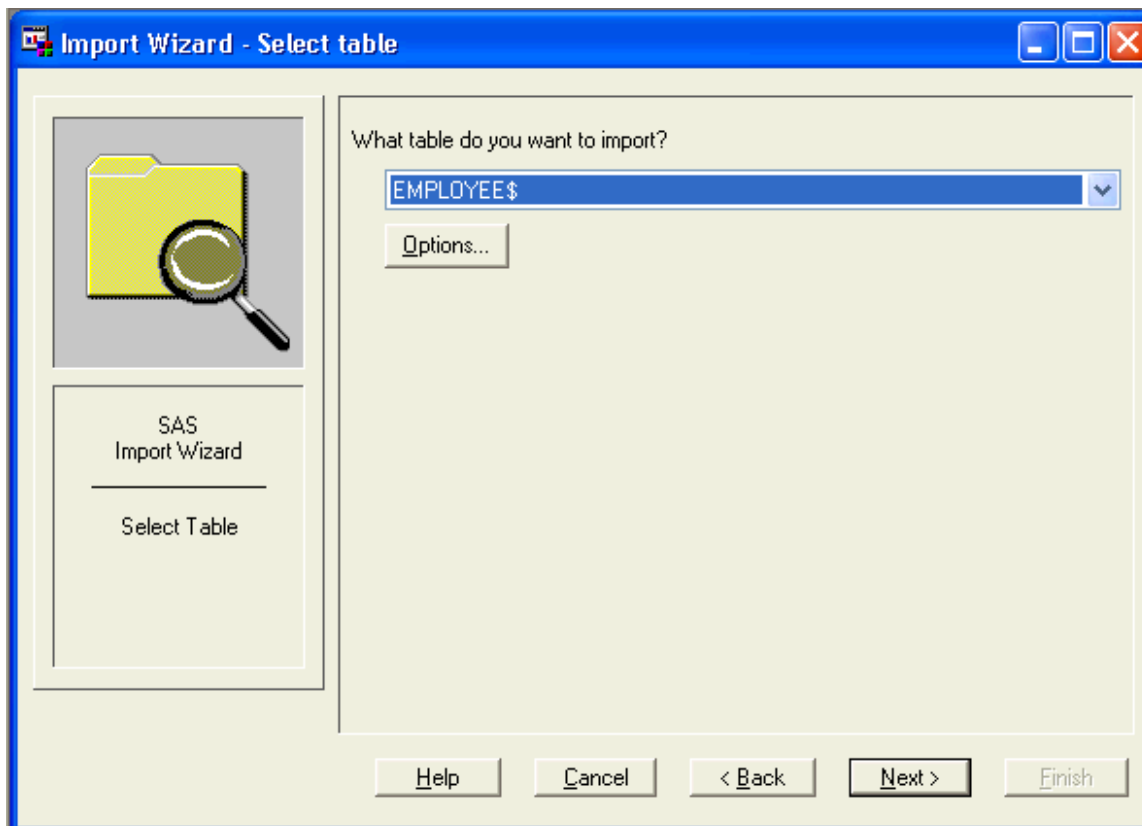
Double-click on the name of the file, and click on **Open**.



The filename that you have chosen will appear in the *Connect to MS Excel* dialog box. Click on **OK**.

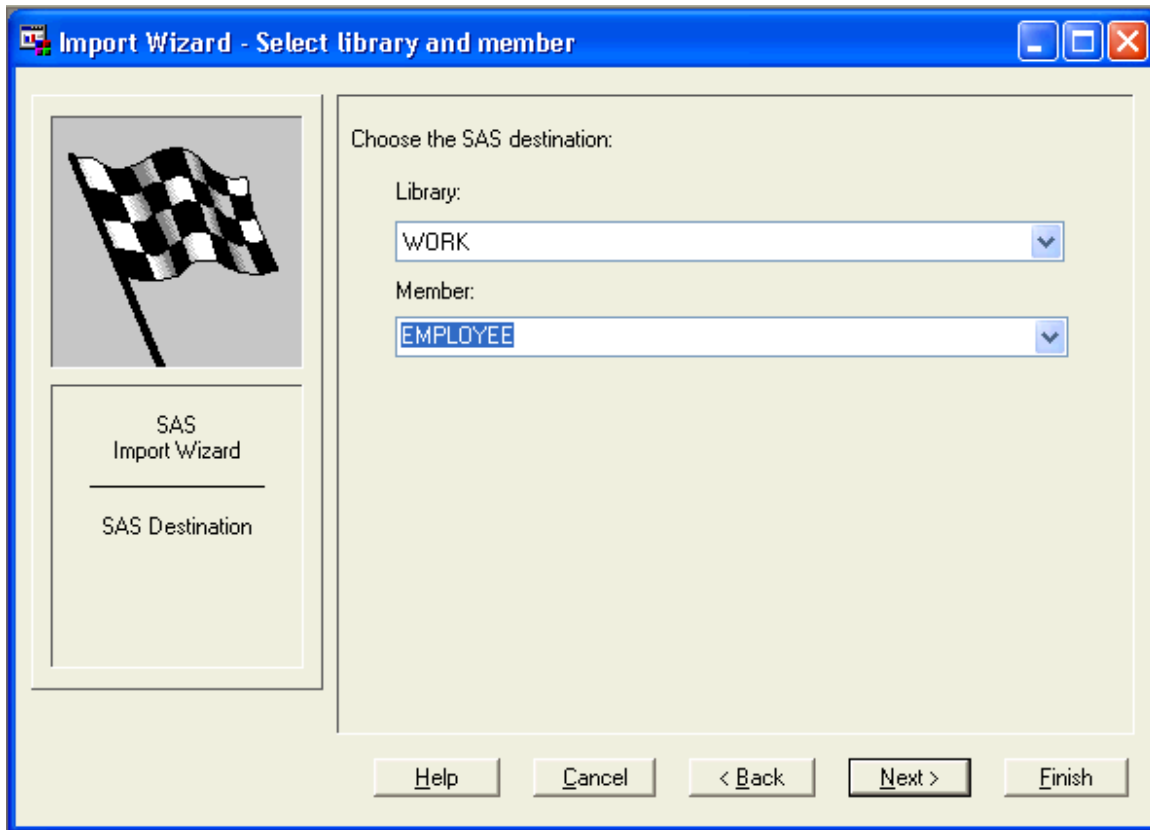


In the next dialog box, **select the table** (worksheet) that you want to import from the pulldown list. In this example, we are selecting the table named "EMPLOYEE", which is in fact, the only table in this workbook. Click on **Next >**.



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 the **Member** box. In this case, we are saving the data set in the **WORK** library, with the member name **EMPLOYEE**. SAS will call this dataset **WORK.EMPLOYEE**, or simply **EMPLOYEE**.

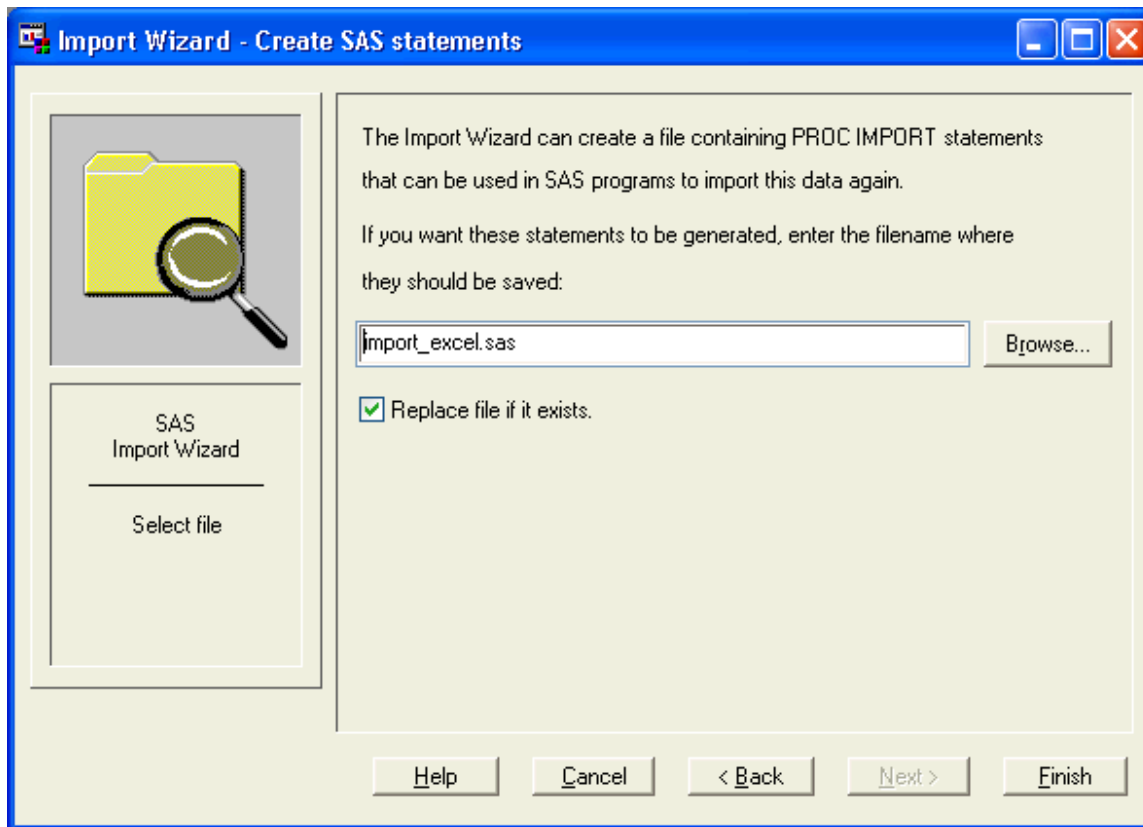


At this point, you have two choices for what to do.

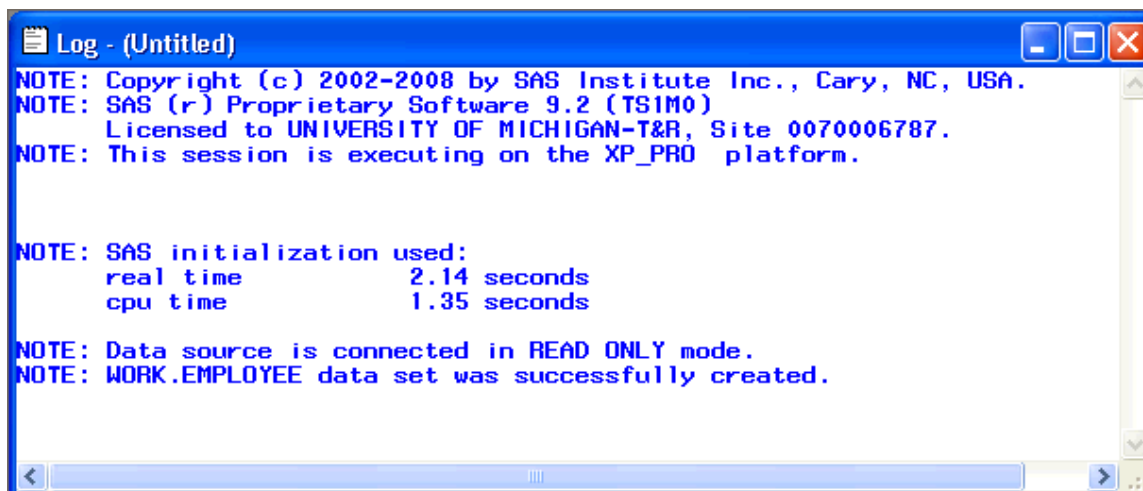
- If you click on **Finish**, the data set will be saved, and you can proceed to work with it.
- If you click on **Next>**, you will go to another dialog box, where you will have a chance to save SAS commands to be used to import the data set at a later time.

I usually click on **Next>**, so I can save my commands for later use. This process is shown below.

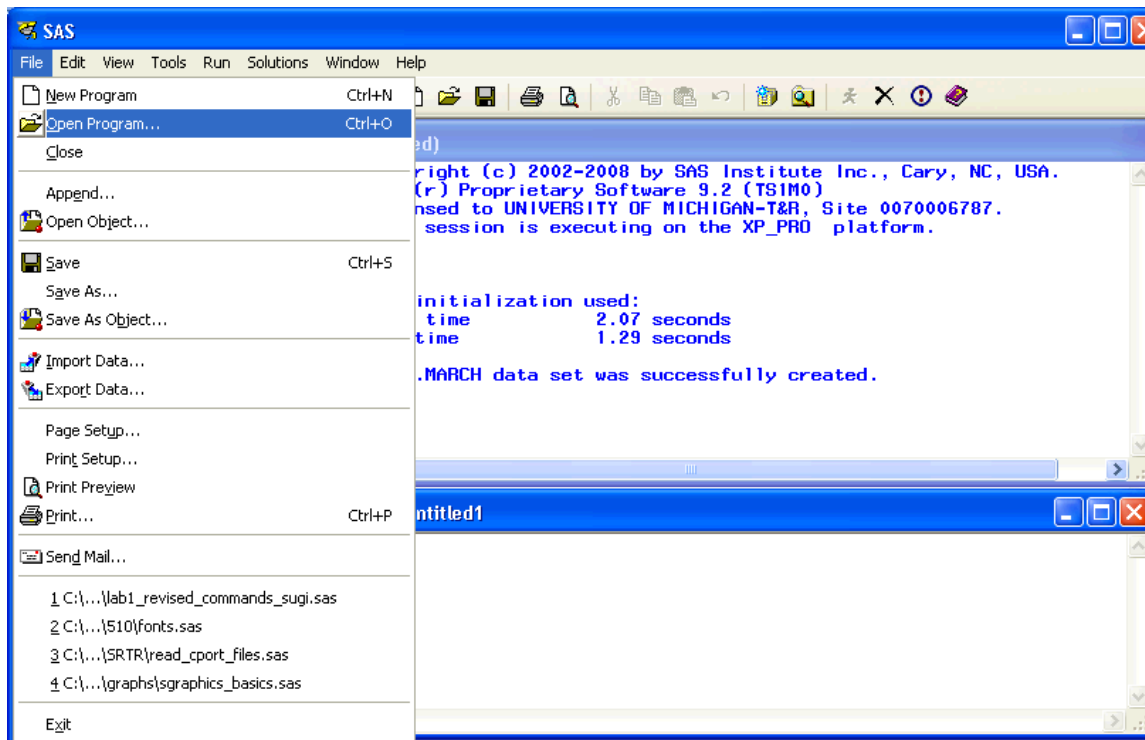
Browse to the location where you wish to save your SAS commands (e.g., import_excel.sas). Click on **Finish** to complete importing the dataset and saving your commands.



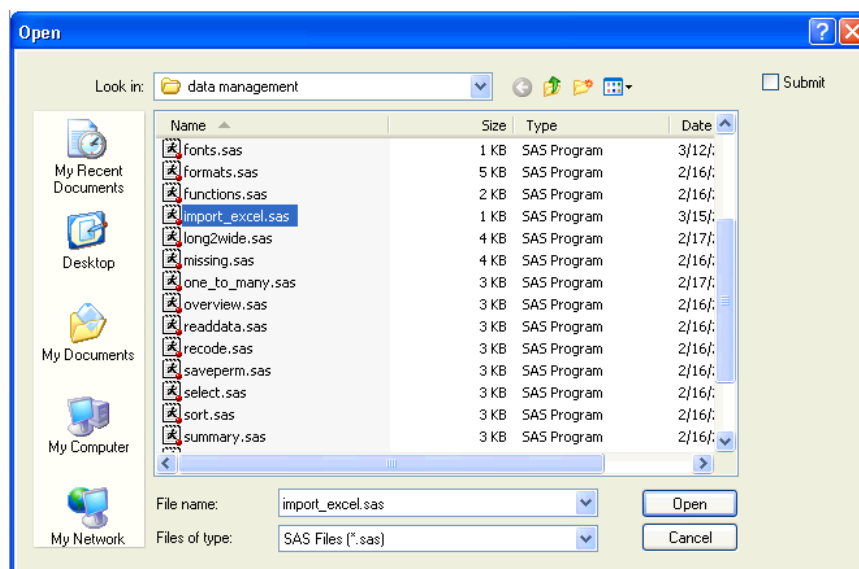
When you check the SAS LOG, you will see a message stating that the data set was successfully created.



If you saved your commands to import the Excel file, you can open them in your SAS enhanced editor, by going to File...Open Program... and browsing to the command file that you saved.



Browse to the file you saved, and select it. Click on **Open**, or simply **double-click** on the filename.



The command file will open in your Program Editor Window.

```
PROC IMPORT OUT= WORK.EMPLOYEE
```

```

        DATAFILE= "C:\temp\labdata\EMPLOYEE.XLS"
        DBMS=EXCEL REPLACE;
RANGE="EMPLOYEE$";
GETNAMES=YES;
MIXED=YES;
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;
RUN;

```

From the Program Editor Window, you can modify your SAS commands, and save them to be used later. The data set can be modified by creating a new data step, with additional commands, for example:

```

data employee2;
  set employee;
  jobtime_yrs = jobtime/12;
  prevexp_yrs = prevexp/12;

  if educ not=. then do;
    if educ < 12 then edcat=1;
    if educ = 12 then edcat=2;
    if educ > 12 then edcat=3;
  end;
run;

```

SAS procedures can be run using the new or the original dataset:

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
id	id	474	237.5000000	136.9762753	1.0000000	474.0000000
bdate	bdate	473	-1179.56	4302.33	-11282.00	4058.00
educ	educ	474	13.4915612	2.8848464	8.0000000	21.0000000
jobcat	jobcat	474	1.4113924	0.7732014	1.0000000	3.0000000
salary	salary	474	34419.57	17075.66	15750.00	135000.00
salbegin	salbegin	474	17016.09	7870.64	9000.00	79980.00
jobtime	jobtime	474	81.1097046	10.0609449	63.0000000	98.0000000
prevexp	prevexp	474	95.8607595	104.5862361	0	476.0000000
minority	minority	474	0.2194093	0.4142836	0	1.0000000
jobtime_yrs		474	6.7591421	0.8384121	5.2500000	8.1666667
prevexp_yrs		474	7.9883966	8.7155197	0	39.6666667
edcat		474	2.3755274	0.6775720	1.0000000	3.0000000