

Importing an Excel Worksheet into SAS

(commands=import_excel.sas)

I. Preparing Excel Data for a Statistics Package

These instructions apply to setting up an Excel file for SAS, SPSS, Stata, etc.

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** (longer variable names are currently allowed in SAS and SPSS)
 - 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!**
- Include only 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 values only 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.
- Include only **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.
- **Use numeric values when feasible**. While character variables are allowed in statistical packages, they are not as flexible as numeric variables, which are preferred.
- 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 Excel file might look like this:

	A	B	C	D	E	F	G	H	I	J
1	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority
2	1	m	2/3/1952	15	3	\$57,000	\$27,000	98	144	0
3	2	m	5/23/1958	16	1	\$40,200	\$18,750	98	36	0
4	3	f	7/26/1929	12	1	\$21,450	\$12,000	98	381	0
5	4	f	4/15/1947	8	1	\$21,900	\$13,200	98	190	0
6	5	m	2/9/1955	15	1	\$45,000	\$21,000	98	138	0
7	6	m	8/22/1958	15	1	\$32,100	\$13,500	98	67	0
8	7	m	4/26/1956	15	1	\$36,000	\$18,750	98	114	0
9	8	f	5/6/1966	12	1	\$21,900	\$9,750	98	0	0
10	9	f	1/23/1946	15	1	\$27,900	\$12,750	98	115	0
11	10	f	2/13/1946	12	1	\$24,000	\$13,500	98	244	0
12	11	f	2/7/1950	16	1	\$30,300	\$16,500	98	143	0
13	12	m	1/11/1966	8	1	\$28,350	\$12,000	98	26	1
14	13	m	7/17/1960	15	1	\$27,750	\$14,250	98	34	1

How to Save the Excel File:

Excel allows you the option of saving a file in several different formats. If you're having problems, Version 4.0 Excel 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**.

Multiple Worksheets:

If you have several worksheets, you can select the worksheet that you wish to import when you bring the data into SAS, but you will need to bring in each sheet individually and then merge them in the statistical package you are using.

The consultants at CSCAR can help you with this.

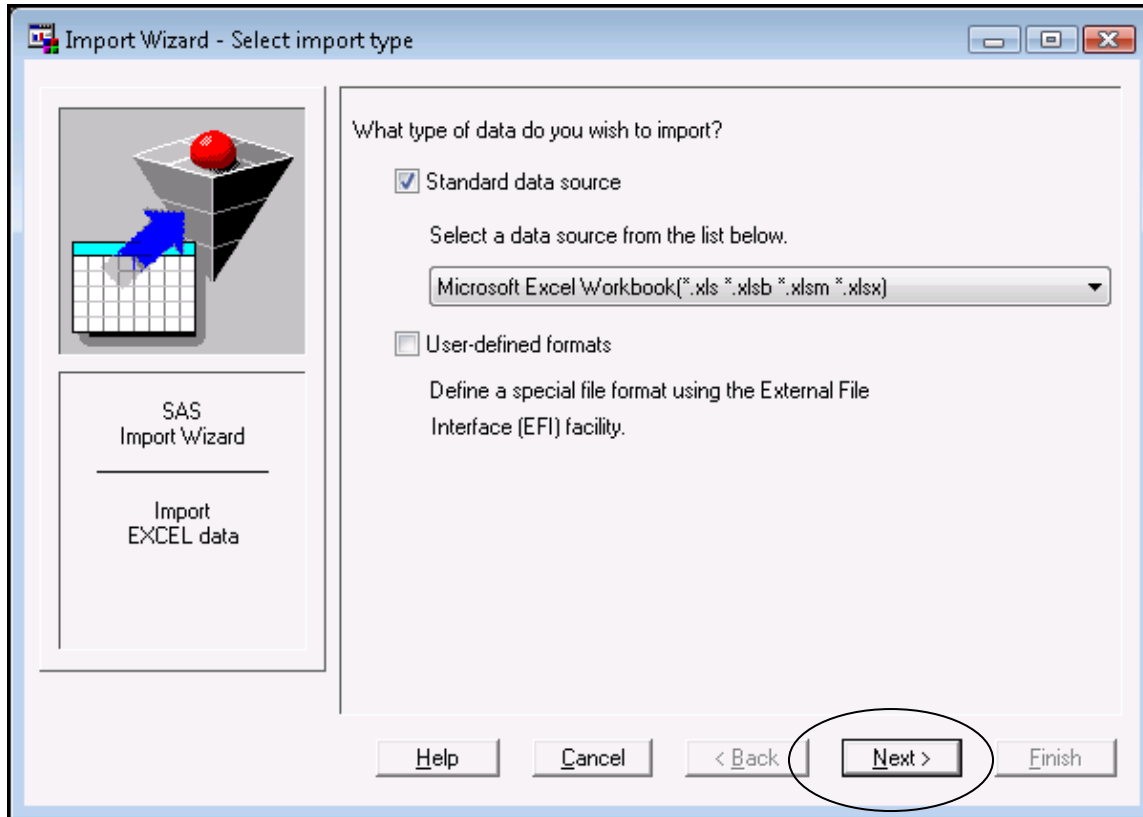
A document very similar to this one is available online at <http://www.umich.edu/~cscar/software/fromexcel.html>

What Type of Excel Files Can You Import to SAS?

You can import Excel worksheets, starting with very early versions of Excel (e.g., Excel version 4.0). You can also import individual sheets from workbooks for later versions of Excel (e.g. Excel 2000), but only one sheet at a time. Excel 2007 (.xlsx) files can be opened by the updated versions of SAS 9.2 or SAS 9.3, but if you're using an earlier version of SAS, you will have to save the files as .xls files before proceeding.

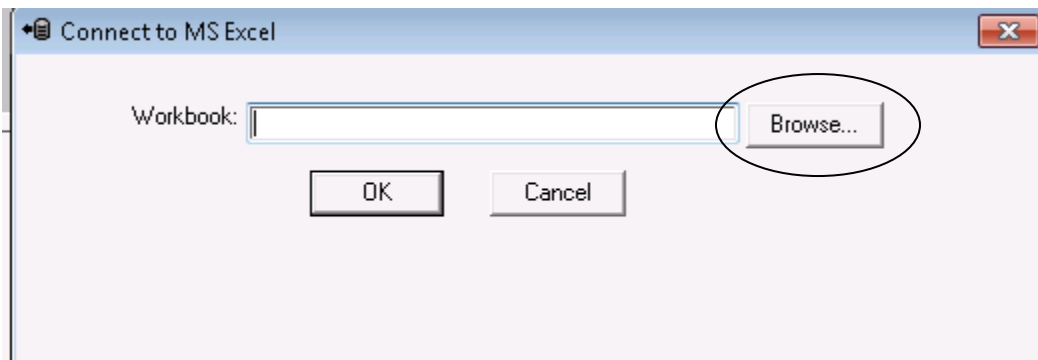
II. Importing the Excel File to SAS Step-By-Step Instructions:

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.

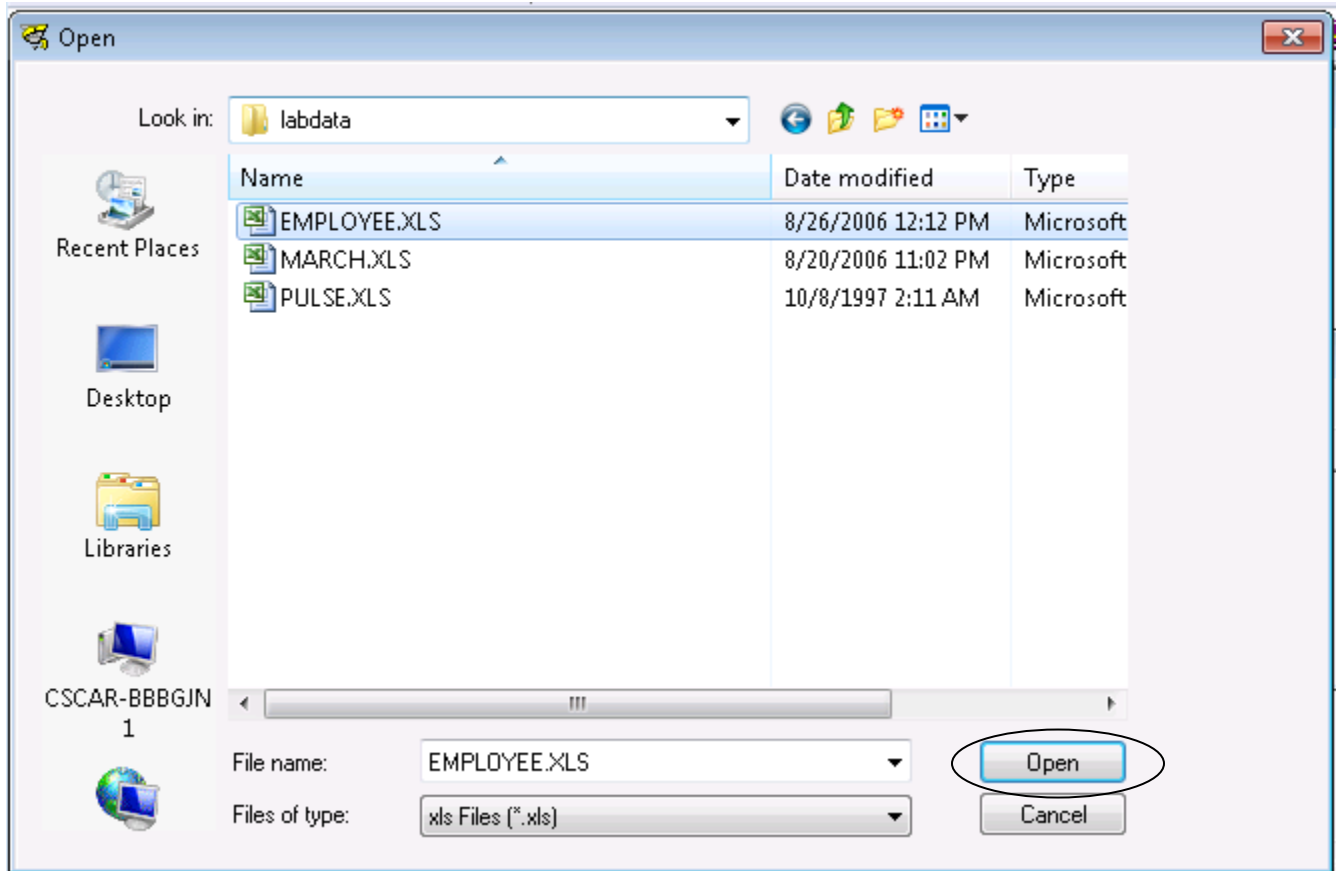


Click on the “Next>” button to proceed.

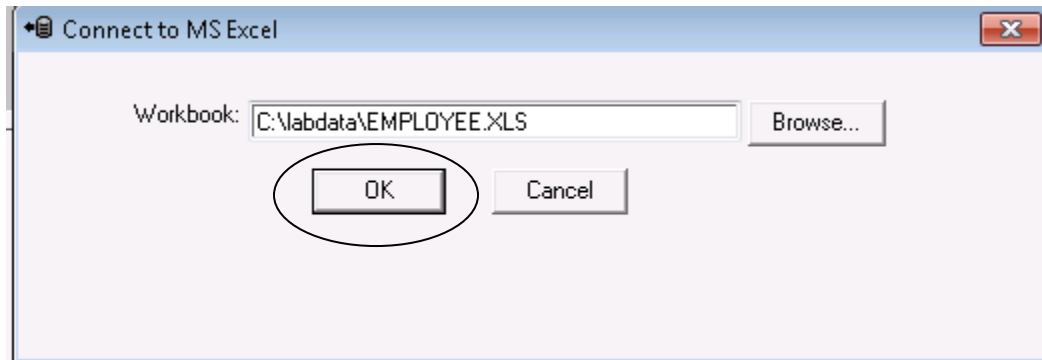
In the dialog box that opens, click on **Browse** to locate the file you wish to import.



Select the Excel file to open and click on the “**Open**” button.

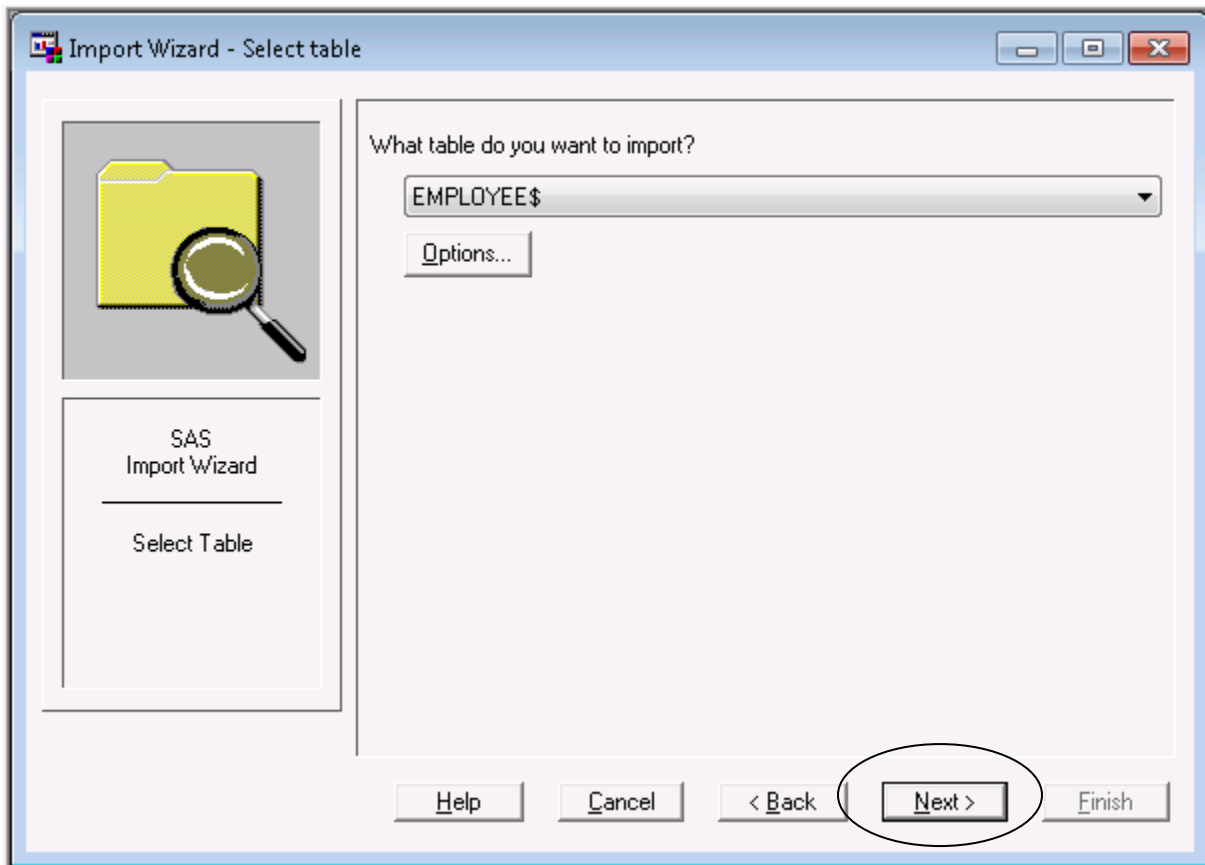


The file name that you have chosen will appear in the browse dialog box.



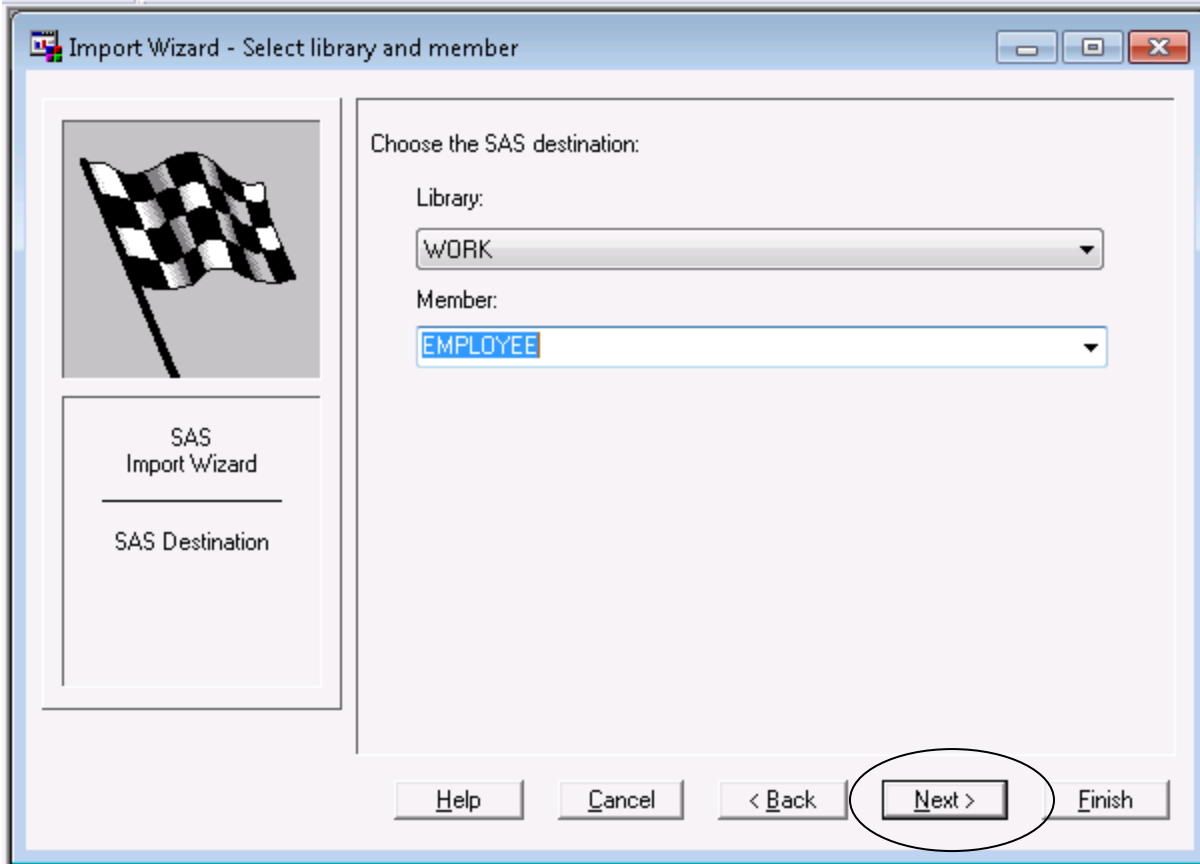
Click on “**OK**”.

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



Click on “**Next>**” to proceed.

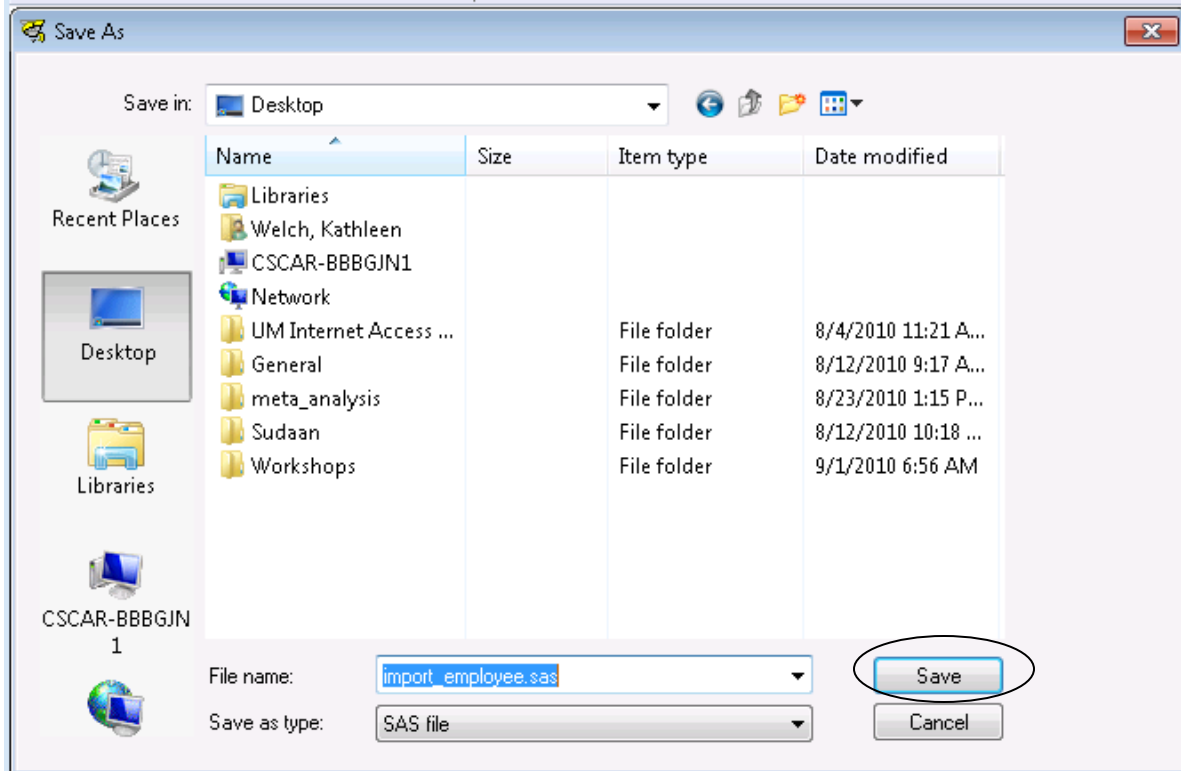
You will be taken to a dialog box that allows you to save the SAS data set to a library. The default temporary library “WORK” will be automatically filled in for you, but you need to type the data set (Member) name. In this case, we are saving the data set as WORK.EMPLOYEE.



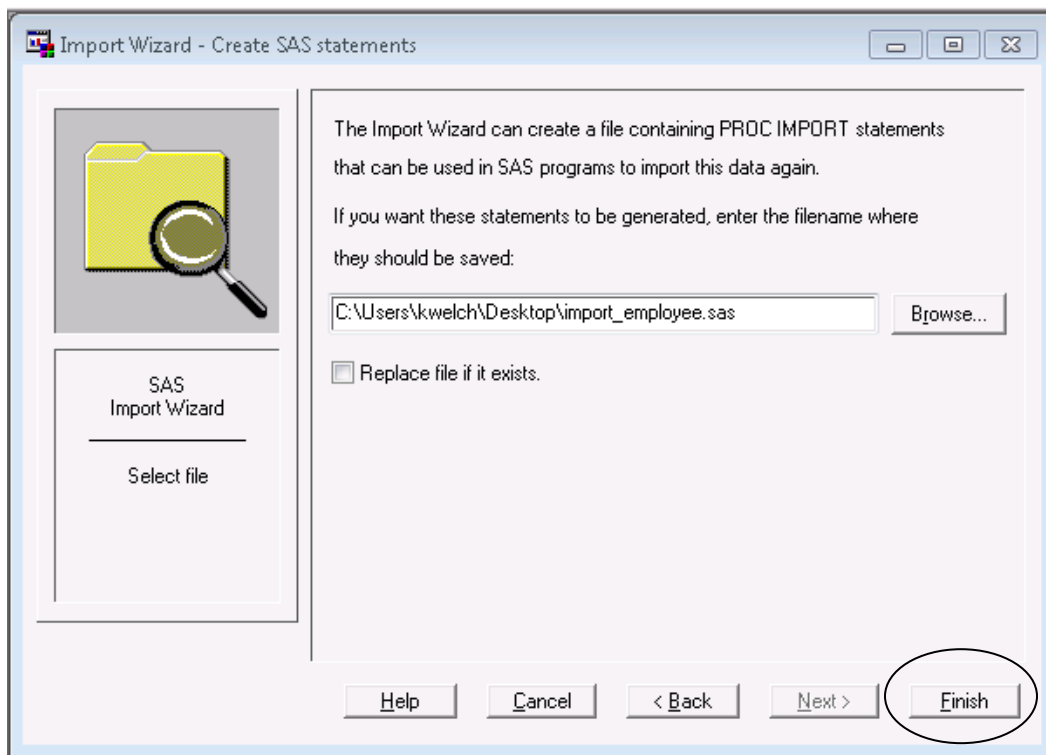
At this point, you have two choices:

- If you click on “Finish”, the data set will be saved, and you can proceed to work with it.
- If you click on “Next>”, as illustrated here, you will go to the following dialog box, where you will have a chance to save the SAS commands that were used to import the dataset. You can use these commands later to re-import the data..

I usually click on “Next>”, so I can save my commands. This process is shown below. Browse to a location where you wish to save your SAS commands and give them a name, as in the example below (the commands were saved on the desktop as “import_employee.sas”).



Click on “Save” and you will then see the dialog box below.

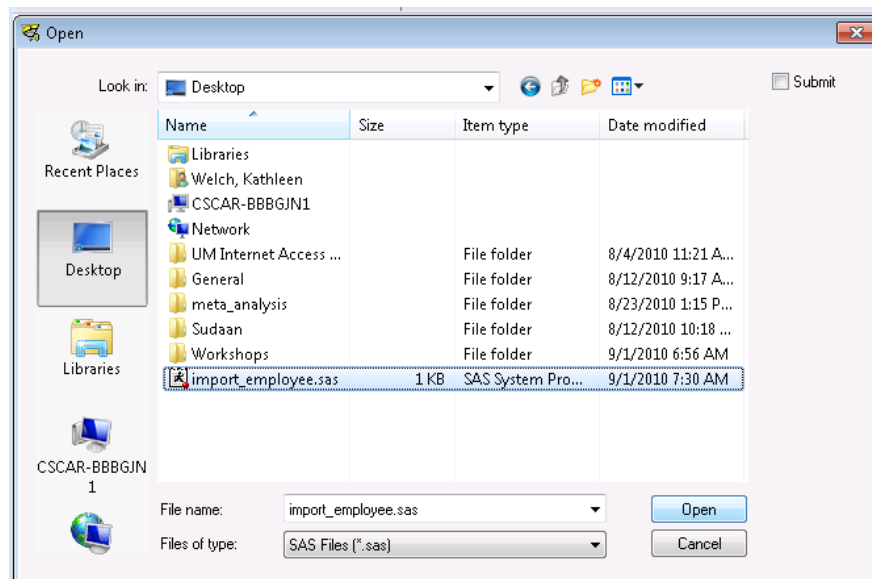


You can now click on “Finish” to complete importing the data set. Check the SAS Log. You should see the following message in the log:

NOTE: WORK.EMPLOYEE data set was successfully created.

Using SAS commands to import Excel Files:

If you saved your commands in the previous step, you can now bring them into your SAS enhanced editor, by going to File...Open Program... and browsing to the command file that you saved. Alternatively, you can type these commands by hand and submit them to SAS.



The command file is shown below:

```
PROC IMPORT OUT= WORK.EMPLOYEE  
             DATAFILE= "C:\labdata\EMPLOYEE.XLS"  
             DBMS=EXCEL REPLACE;  
RANGE="EMPLOYEE$";  
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 employee2;
  set employee;
  saldiff = salary-salbegin;
  if 0<= educ < 12 then edcat = 1;
  if educ = 12 then edcat = 2;
  if educ > 12 then edcat = 3;
run;
```

SAS can now be used to run procedures on this new data set, for example, Proc Means, as shown below:

```
proc means data=employee2;
run;
```

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
saldiff		474	17403.48	10814.62	5550.00	76240.00
edcat		474	2.3755274	0.6775720	1.0000000	3.0000000