How to Use Excel for Data Entry

Research Computing Support

Excel is a very popular tool for entering and manipulating data. This document shows you how to enter data that you can easily open in statistics packages such as SPSS or SAS. Excel has some statistical analysis capabilities but they often provide incorrect answers and we do not recommend using them. For a comprehensive list of these limitations, see http://www.forecastingprinciples.com/paperpdf/McCullough.pdf. You can easily add accurate analysis methods to Excel by installing "R and Friends" available at: http://rcom.univie.ac.at/.

We support many options for automated data entry; including darken-the-circle scan forms, web surveys and data acquisition directly from scientific instruments. If you have questions regarding data input or analysis, contact UT's Office of Information Technology Helpdesk at 865-974-9900. We will do our best to get back to you within 24 hours. Most UT Knoxville students, faculty and staff may receive up to 10 free hours of statistical computing assistance each semester. See http://oit.utk.edu/research for details. We also offer training in data analysis each semester. See http://oit.utk.edu/training/ for details.

Basic Rules of Data Structure
- All your data should be in a single spreadsheet of a single file.
- Enter variable names in the first row of the spreadsheet.
- Use variable names that are no longer than 8 characters, beginning with a letter.
- Variable names cannot contain spaces, but may use the underscore character.
- No other text rows such as titles should be in the spreadsheet.
- No blank rows should appear in the data.
- Always include an ID variable on your original data collection form and in the spreadsheet to help you find the case again if you need to correct errors. You may need to sort the data later, so the row number in Excel would then apply to a different subject or sampling unit.
- If you have multiple groups, put them in the same spreadsheet along with a variable that indicates group membership (see Gender example below).
- Avoid using alphabetic characters for values. For example to enter political party, enter 1 instead of Democrat, 2 instead of Republican and 3 instead of Other.
- If your group has only two levels, coding them 0 and 1 makes some analyses much easier to do.
- For missing values, leave the cell *blank*. Although SPSS and SAS use a period to represent a missing value, if you actually type a period in Excel, stat packages will read the column as character data so you won't, for example, be able to calculate even the mean of a column.
- You can enter dates with slashes (6/13/2003) and times with colons (12:15 AM).
- For text analysis, you can enter up to 32K of text, about 8 pages in a single cell. However, if you cut & paste if from elsewhere, remove carriage returns first so as they will cause it to jump to a new cell.

A data structure that's easy to analyze:

ID	Gender	Salary
1	0	32000
2	1	23000
3	0	37000
4	1	54000
5	1	48500

Same data, but not easy to analyze:

Data for Female Subjects	
ID	Salary
1	32000
3	37000
Data for Male Subjects	
ID	Salary
2	23000
4	54000
5	48500

Data Entry Tips

- Save your data frequently and make backup copies and store them in separate buildings. Don't risk losing all your hard work in a fire or theft!

- Avoid using Excel to sort your data. It's too easy to sort one column independent of the others, which essentially destroys your data! Statistics packages can sort data and they understand the importance of keeping all the values in each row locked together.

- If you need to enter a pattern of consecutive values such as an ID number with values such as 1,2,3 or 1001,1002,1003, enter the first two, select them and drag the box in the lower right corner as far as you wish. Excel will see the pattern of the first two entries and extend it as far as you drag your selection. This works for days of the week and dates too. You can create your own lists in Options>Lists, if you use a certain pattern often.

- To help prevent typos, you can set minimum and maximum values, or create a list of valid values. Select a column or set of similar columns, then choose Validation from the Data menu. To set minimum and maximum values, choose Allow: Whole Numbers or Decimals and then fill in the values in the Minimum and Maximum boxes. To create a list of valid values, choose Allow: List and then fill in the numeric or character values separated by commas in the Source box.

- The gold standard for data accuracy is the dual entry method. With this method you actually enter all the data twice. Only this method can catch errors that are within the normal range, but still wrong. Excel can show you where the values differ. Enter the data first in Sheet1. Then enter it again using the exact same layout in Sheet2. Finally, go to Sheet3 in cell A1 and enter this formula:
  =IF(sheet1!A1=sheet2!A1,1,0)
  This means that if the value in Sheet1 cell A1 is equal to the value in Sheet2 cell A1, then Sheet3 A1 will display a 1 to indicate a match and 0 to indicate bad data. To extend this formula to all the cells, select cell A1 in Sheet3 and drag the box in the lower right corner until the cell stretches to cover all the space you used for your data in Sheet1. Then check to see where the zeros are in sheet 3. Those will be your typos. You then check to see which entry was wrong, Sheet1 or Sheet2. Make corrections until Sheet3 is full of ones, indicating no errors. When you read the data into a statistics package, you will only need to read the data in Sheet1.

- When looking for data errors, it can be very helpful to display only a subset of values. To do this, select all the columns you wish to scan for errors. Choose Filter from the Data menu and then choose Autofilter. A downward-pointing triangle will appear at the top of each column selected. Clicking it displays a list of the values contained in that column and the words (All) and (Custom). If you have entered values that are supposed to be, for example, between 1 and 5 and you see 6 on this list, choosing it will show you only those rows in which you made that error. Then you can fix them and choose (All) from the drop-down menu. The (Custom) selection will allow you to use simple logic to find, for example, all rows with values greater than 5. When you are finished, choose Autofilter from the Data->Filter menu.

Steps for Reading Excel Data Into SPSS

1. In SPSS, choose *File: Open: Data*.
2. Change the "*Files of file type*" box to "Excel (*.xls)"
3. Select the spreadsheet name as you would in Excel
4. When the *Opening Excel Data Source* box appears, check the box for *Read variable names from the first row of data*, then click OK.
5. When the data appears in the SPSS data editor spreadsheet, Choose *File: Save as* and leave the *Save as type* box to *SPSS (*.sav)*.
6. Enter the name of the file without the .sav extension and then click *Save* to save the file in SPSS format
7. Next time open the .sav version, you won't need to convert the file again.
8. If you create variable or value labels in the SPSS file and then need to read your data from Excel again you can copy them into the new file. First, make sure you use the same variable names. Next, after opening the file in SPSS, use Copy Data Properties from the Data menu. Simply name the SPSS file that has properties (such as labels) that you want to copy, check off the things you want to copy and click *OK*.

Steps for Reading Excel Data Into SAS

The process of importing data into SAS is quick but saving the data permanently as SAS file is complex. Therefore, we recommend that you import the data each time you need it. If you are an advanced SAS programmer familiar with SAS data libraries, it will probably be obvious how to avoid this repetition.

1. In SAS, choose *File: Import Data*. The Import Wizard will appear.
2. Make sure that the *Standard data source* box is checked and that the *Select a data source from the list below* is set to the version of Microsoft Excel that you used to create the file. Then click Next.
3. In the Select File box, browse to find the file and click Next.
4. In the *Choose the SAS destination* box, leave the Library box set to WORK and enter TEMP as the *Member name*. Then click Finish.
5. If you click Next instead of Finish in the step above, SAS will say, *The Import Wizard can create a file containing PROC IMPORT statements that can be used in SAS programs to import this data again. If you want these statements to be generated, enter the filename where they should be saved*. SAS programmers will appreciate this feature but we recommend beginners avoid this step by clicking on Finish.
6. The data can now be used by any SAS program. For example, submitting:
   PROC MEANS; RUN;
   should calculate means and other basic statistics using your data.

Steps for Reading Excel Data Into R

R is a free, open-source statistics package available from http://www.r-project.org/. The instructions for reading from the clipboard or tab delimited files are below. More details and examples about reading various file types including SAS and SPSS are in: http://oit.utk.edu/scc/RforSAS&SPSSusers.pdf .

1. To import Excel data into R via the clipboard, in Excel select only the variable names and data (don't select any extra blank rows or columns) and choose *Copy* from the *Edit* menu. In R, you can create a data frame called mydata by entering the commands:
```
mydata <- read.delim("clipboard")
mydata #this just shows you the file.
```
Note that you do *not* choose *Paste* from the *Edit* menu as you would with most other programs.
2. To import from a file, first save the file in tab delimited form. In Excel, choose *Save As* from the *File* menu. In the *Save As* box, go to the *Save As Type* drop-down list and choose Text (tab delimited)(*.txt). Enter a filename like mydata and click *Save*.
3. Read the tab-delimited file in R with the commands:
```
mydata<-read.table("c:\\mydata.txt",header=TRUE,row.names="ID")
mydata
```
R uses a single backslash "\" as a special character. Using two in a row "\\" in file paths will work instead. The `header` argument tells R to read the variable names from the first record. The `row.names` argument tells it that your ID variable is named "ID". If your ID variable is named something else, even if it is a lower case "id", fill that name in there. If you don't have an ID variable, remove the `row.names` argument and the preceding comma.

This example assumes the file is on your C: drive and is not in a folder. If you instead saved the file to your desktop, you would have to use a file path like:
C:\\Documents and Settings\\myname\\desktop\\mydata.txt
or in My Documents:
C:\\Documents and Settings\\myname\\My Documents\\mydata.txt
4. To save the R workspace, enter the command below (fill in appropriate file path)
```
save.image(file="c:\\mydata.Rdata")
```
5. To read the R workspace the next time you start R, enter the command
```
load(file="c:\\mydata.Rdata")
```