

Chapter 6: Data Analysis with Excel and SPSS

This chapter will introduce techniques of data analysis. You will learn about Excel and SPSS, two programs widely used by psychologists and others to analyze data. You will learn how to import data from a Web experiment into Excel, how to filter the data, and how to perform simple analyses. You will learn how to import data to SPSS, and how to carry out cross-tabs in SPSS. If you plan graduate work in psychology, then you will eventually learn both of these programs. SPSS and Excel can do many of the same tasks, but each program has advantages for certain tasks. If you are an undergraduate, your instructor may choose to focus on one of these programs and leave the other for a future course. It is well worth learning both programs if you plan to study in graduate school or to join the workforce in a job in which financial records, client records, tables, data, or charts are to be analyzed

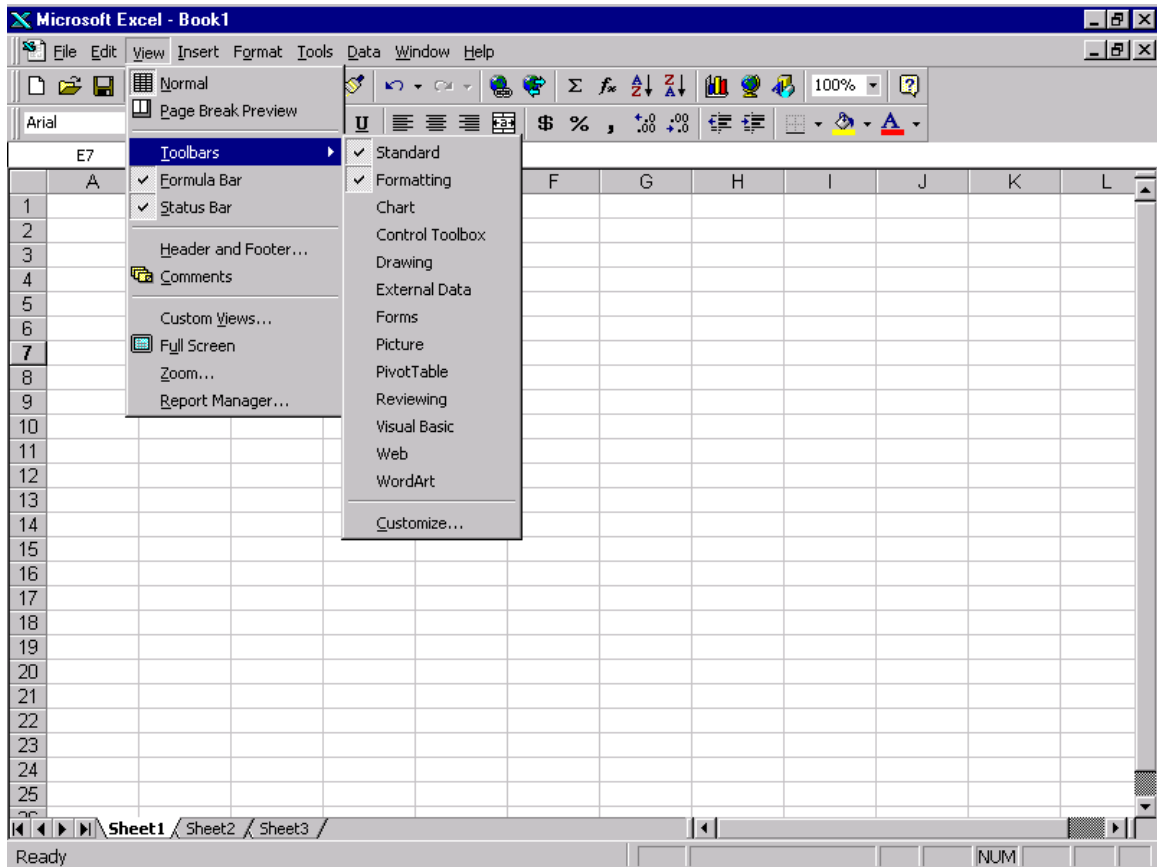
A. Introduction to Excel

Excel is a powerful spreadsheet program that can be used to perform many mathematical and statistical calculations. It also supports graphing of data. Start Excel and examine the window, which will look like that in Figure 6.1.

When you start the program, your toolbars may differ from those displayed in Figure 6.1. That is because there are many different toolbars that may either be displayed or hidden, and they can also be moved around the screen. Pull down the **View** menu, as shown in Figure 6.1, and be sure that the *Formula Bar*, *Status Bar*, *Standard Toolbar*, and *Formatting Toolbar* are all checked. You can experiment by selecting these different toolbars (which toggles them on and off), and examine the effects on the display. The toolbars can also be moved around the screen, which also changes the appearance of the window. To move one of the toolbars, click your arrow pointer on the handle at the left and drag the toolbar around the screen where you want it. Now that you know how to display and move the toolbars, you can make your screen match what is

shown in the book. You can also get help from the **Help** menu or by clicking on the question mark icon on the Standard Toolbar. From help, you can select an "assistant" to offer help and suggestions for using the program. When you are trying new tasks in Excel, it may be of help to keep the assistant on the screen, as the assistant will offer suggestions pertaining to the tasks you are doing. Clicking the assistant will bring up help screens.

Figure 6.1. The **View** menu in Excel can be used to display the Formula Bar, Status Bar, Standard Toolbar, and Formatting Toolbar. If these are not checked, use the **View** menu to select them. You can also get help from the **Help** menu.



Note that the spreadsheet (the grid) is labeled with letters for columns, and numbers for rows. A particular cell is designated by its column letter and its row number. For example, A1 is the first cell in the upper, left-hand corner of the spreadsheet; A2 is the cell directly below A1. B1 is the cell in the next column, directly to the right of A1, and so on. A range of cells is designated by its upper-left and lower-right corner cells, separated by a colon. For example, A1:C5 refers to the 15 cells from A1 to A5, B1 to B5, and C1 to C5.

Suppose you measured the heights and weights of students in a class. You could enter the data for each case (i.e., each person) in a separate row, and use a different column for each variable (the person's name, height, and weight).

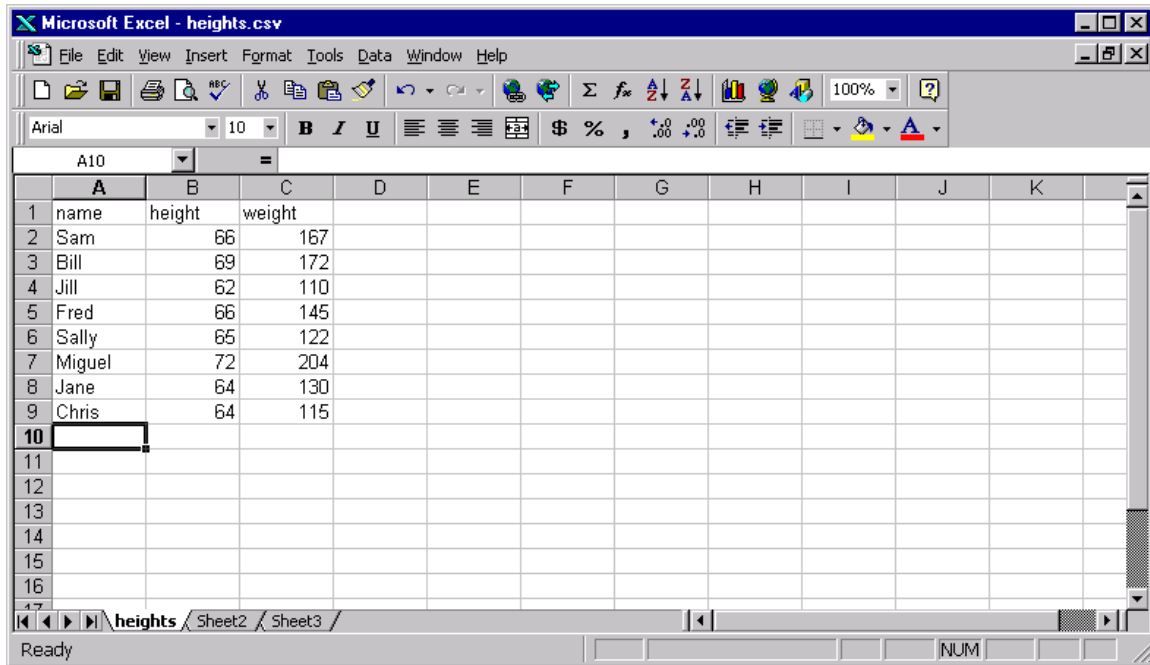
Figure 6.2 shows data obtained from eight students who had their heights measured (in inches) and weights measured (in pounds). Note that the first row has been used to label the variables, *name*, *height*, and *weight*. Type the same values into your spreadsheet. After typing each value, you can push either *Enter* or *Tab*. If you push *Tab*, the datum will be entered, and the selected cell will move to the right. If you push *Enter*, the datum will be entered, and the selected cell will drop to the next row.

Insert Figure 6.2 about here.

B. Saving Excel Data

To save the data, select *Save As* from the **File** menu. Excel allows you to save your data in many different formats. On the Save dialog, near the bottom, appears the *Save as type* box. Click the drop down arrow, and select *Microsoft Excel Workbook*. Give the file the name *heights.xls* (On a PC, the *.xls* extension will be added for you; on a Mac, you should add it yourself).

Figure 6.2. Entering data in Excel. Note that numerical data are right-justified and alphabetical data (the names) are left-justified. After typing each datum, push either *Enter* or *Tab*. This workbook was saved as *heights.xls*, and this worksheet was saved as *heights.csv*.



The Excel workbook contains three worksheets. You can click back and forth to view the three worksheets by clicking the tabs at the bottom of the window for *Sheet1*, *Sheet2*, and *Sheet3*. For now, only the first sheet has data, and the others are empty.

You can also save (one worksheet at a time) in Comma Separated Value (CSV) format. This format is very useful because CSV files can be imported to a text editor such as NotePad or SimpleText. Files in CSV format can also be easily imported to SPSS and many other programs. When you save data from Excel into CSV format, you should first save your data (give it a different name) as an Excel workbook, because CSV files do not contain all of the information that may be in a workbook. When you save as CSV, the program warns you that you may lose information. If you have already saved it as a workbook, you can go ahead and save as CSV.

To save a worksheet in CSV format, be sure you are looking at the worksheet you want to save (click the tabs at the bottom to switch from sheet to sheet). Then select *Save As* from the **File** menu. Next, choose *CSV (Comma delimited)* from the *Save as type* drop down list. Give the file the name, *heights.csv* to distinguish it from the workbook, *heights.xls*. Once you have saved the file, you can close it in Excel and open it in NotePad or another text editor. In the text editor, you will see that the file is a very simple text file with the variable names in the first line, and each case is on a separate line. In each line, commas separate the values.

C. Download the Data

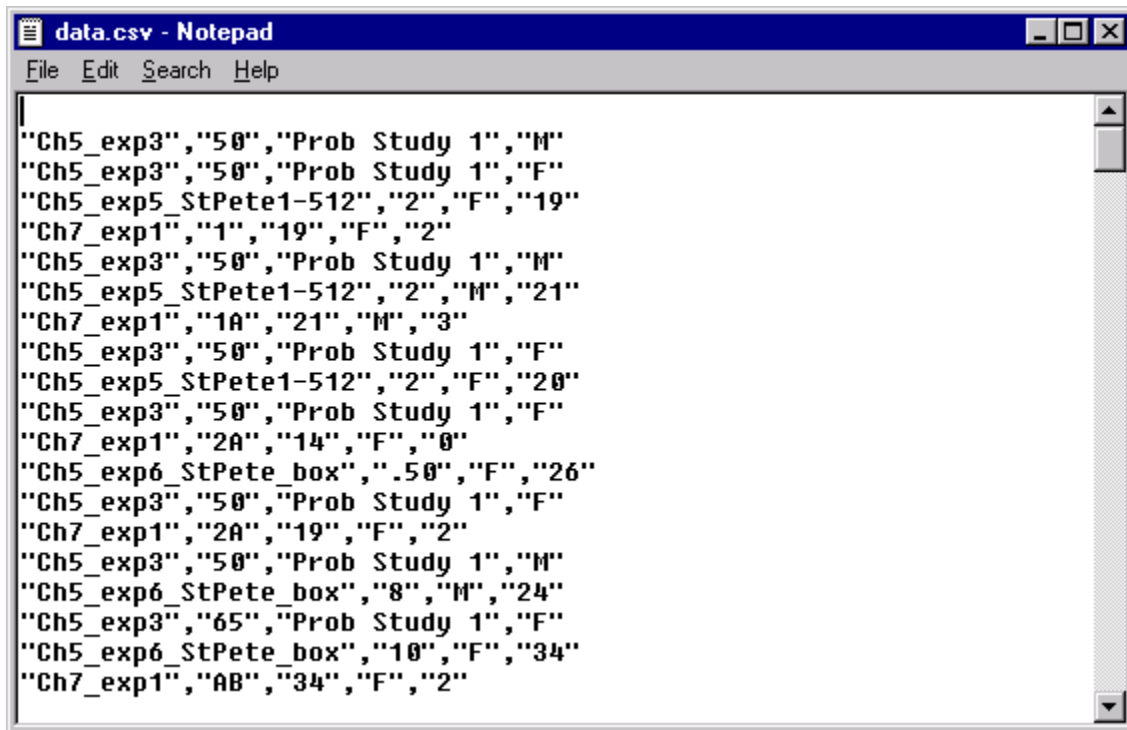
If you use the generic script of Chapter 5, your data will be sent to the file, *data.csv*, on the *psych.fullerton.edu* server. The next task, therefore, will be to download these data and filter them to separate data from different studies.

First, use FTP to download the file, *data.csv*, from the *psych.fullerton.edu* server, using the method you learned in Chapter 2. Next, open the file in a text editor such as NotePad. Remember when using NotePad (and other PC applications), you need to select “All files” (for

Files of type) or the *Open file* dialog will not display files ending in *.csv*. The file, *data.csv*, will change as new experiments are added and as old data are cleaned out from time to time, so the appearance of the file may be different when you open the file from how it appears in Figure 6.4.

Insert Figure 6.4 about here.

Figure 6.4. Appearance of file *data.csv*, containing data from several experiments. Note that each value is in quotes, and commas separate the values. This type of file can be imported to many spreadsheet and statistical programs.



```
"Ch5_exp3","50","Prob Study 1","M"  
"Ch5_exp3","50","Prob Study 1","F"  
"Ch5_exp5_StPete1-512","2","F","19"  
"Ch7_exp1","1","19","F","2"  
"Ch5_exp3","50","Prob Study 1","M"  
"Ch5_exp5_StPete1-512","2","M","21"  
"Ch7_exp1","1A","21","M","3"  
"Ch5_exp3","50","Prob Study 1","F"  
"Ch5_exp5_StPete1-512","2","F","20"  
"Ch5_exp3","50","Prob Study 1","F"  
"Ch7_exp1","2A","14","F","0"  
"Ch5_exp6_StPete_box",".50","F","26"  
"Ch5_exp3","50","Prob Study 1","F"  
"Ch7_exp1","2A","19","F","2"  
"Ch5_exp3","50","Prob Study 1","M"  
"Ch5_exp6_StPete_box","8","M","24"  
"Ch5_exp3","65","Prob Study 1","F"  
"Ch5_exp6_StPete_box","10","F","34"  
"Ch7_exp1","AB","34","F","2"
```


D. Filtering with Excel

As you can see in Figure 6.4, there are data from different experiments in the same file. The next task, therefore, is to filter the data to separate data by experiments. You can do that easily with Excel. Start Excel, and choose *Open* from the **File** menu. You will need to select *All Files (*.*)* from *Files of type* in the *Open* dialog, to see the files displayed as in Figure 6.5.

[Insert Figure 6.5 about here].

The file will appear as in Figure 6.6. The next step is to put variable names in the first row. Name the first column, **Exp**, for experiment, then type **X1** in the next column (in Cell B1). Press the enter key. To make a cell entry bold, click on the cell and then click the **B** on the Formatting Toolbar. Now, click on **X1**, and put your mouse pointer in the lower-right corner of cell B1 (with **X1** in it) until the arrow turns into a "+" (like a plus sign), then drag to the right for 31 columns. That trick is called *AutoFill*, and it allows you to name the variables **X1** to **X31** without having to type them all in. [Insert Figure 6.6 about here].

The next step is to filter the data. Click in cell A1, then select from the **Data** menu *Filters*, then *Autofilters*, as shown in Figure 6.7. Drop down filter arrows will appear in the first row. By clicking on one of these arrows and selecting one of the choices, the data will be filtered to display only those records with the chosen value.

For example, you can look only at the data from *Ch5_exp3* by selecting that from the drop down list in the first column. Recall that in *Ch5_exp3*, the problem was to judge the probability that the other coin is heads if Joe tossed two coins and said "at least one" is heads. The correct answer is 33 and a popular wrong answer is 50. By filtering now on the second column as well as the first, you will see how many people chose 33 as the right answer, or how many chose 50. Realize that the other data are still there; they are just hidden when filtered.

[Insert Figure 6.7 about here].

Figure 6.5. Open file dialog in Excel. Note that *Files of type: All Files (*.*)* has been selected; otherwise, *data.csv* would not be displayed in the window.

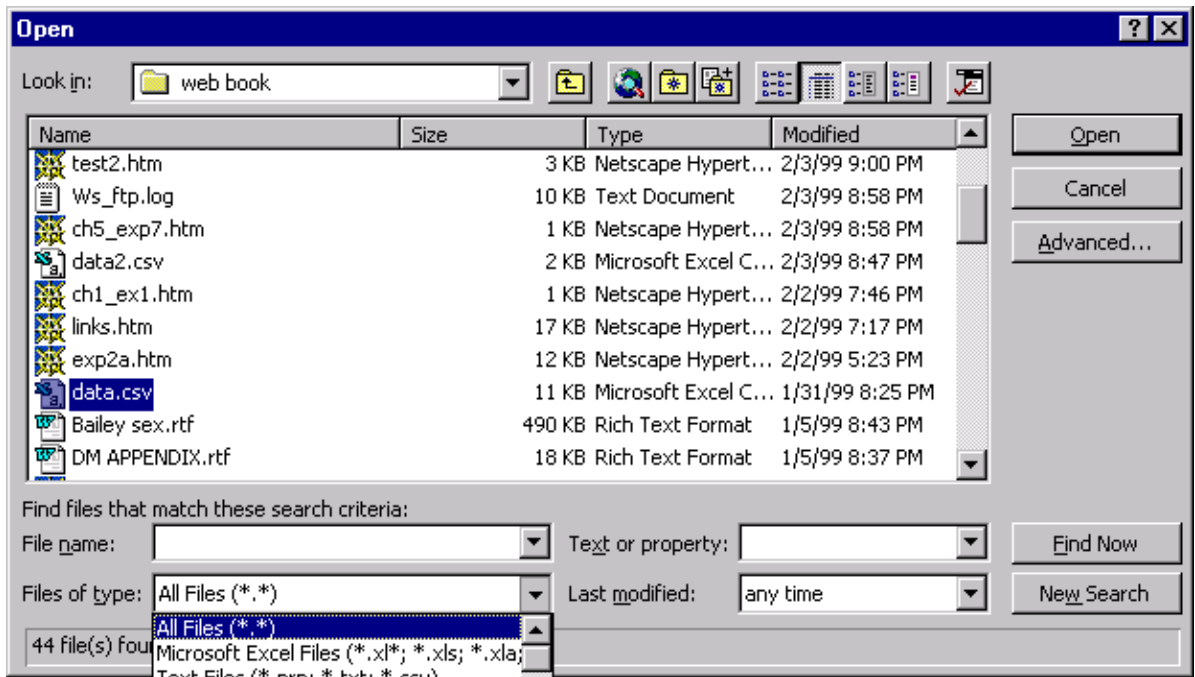


Figure 6.6. Appearance of *data.csv* in Excel. The next step is to enter column labels in the first row.

The screenshot shows the Microsoft Excel interface with a worksheet named 'data.csv'. The data is as follows:

	A	B	C	D	E	F	G	H
1								
2	Ch5_exp3	50	Prob Study	M				
3	Ch5_exp3	50	Prob Study	F				
4	Ch5_exp5	2	F		19			
5	Ch7_exp1	1	19	F		2		
6	Ch5_exp3	50	Prob Study	M				
7	Ch5_exp5	2	M		21			
8	Ch7_exp1	1A	21	M		3		
9	Ch5_exp3	50	Prob Study	F				
10	Ch5_exp5	2	F		20			
11	Ch5_exp3	50	Prob Study	F				
12	Ch7_exp1	2A	14	F		0		
13	Ch5_exp6	0.5	F		26			
14	Ch5_exp3	50	Prob Study	F				
15	Ch7_exp1	2A	19	F		2		
16	Ch5_exp3	50	Prob Study	M				
17	Ch5_exp6	8	M		24			
18	Ch5_exp3	65	Prob Study	F				
19	Ch5_exp6	10	F		34			
20	Ch7_exp1	AB	34	F		2		

E. Filter the Probability Data

The rest of this chapter will be devoted to analysis of data from experiments in Chapter 5.

The next section presents an analysis of data that are included on your CD. You should follow along with and your computer. From Excel, open the file, *clean.csv*, from the folder of data files on the CD, and save it on your hard drive (or zip drive) as an Excel Workbook, using *Save As* from the **File** menu. It will now have the extension *.xls*. Remember where you put it. Notice that this file already has names in the first row (those names will be used in Chapter 8). Now click in Cell A1, and select *Filter, Autofilters* from the **Data** menu, as in Figure 6.7. In the first column, click on the filter arrow and select *Ch5_exp3*, from the drop-down select list. That will bring up the screen shown in Figure 6.8.

Insert Figure 6.8 about here.

Now click in the second row (labeled 52) in column A, and drag the mouse to the right and down to the lower right corner of the data. That will select columns A through D and rows from the second row (row 52) through the end (row 1395)—it will look like ants crawling on the screen, showing what has been selected. Then select *Copy* from the **Edit** menu. Next, from the **Insert** menu, select *Worksheet*, and a new worksheet will appear. Click in cell A2 on the new worksheet, and then select *Paste* from the **Edit** menu. The data for the probability study should now appear in the new worksheet. Now type in the correct column labels for the probability experiment, which are **Exp**, **Answer**, **Title**, and **Sex**. The result is shown in Figure 6.9.

Insert Figure 6.9 about here.

Next, double click on the tab at the bottom of the new worksheet, *Sheet1*, and rename it *Ch5_exp3*. A single click on the tab labeled *clean* will take you back to the sheet that contains all of the data. Repeat this process to select the data for the St. Petersburg study (*Ch5_exp4*, *Ch5_exp5*, and *Ch5_exp6*), and paste each set of selected data into a separate worksheet.

Figure 6.8. Open the file *clean.csv* and save it as an Excel workbook. Then select from the **Data** Menu, *Filter...AutoFilters*. Then filter *exp* (Column A) for *Ch5_exp3*. All records with *Ch5_exp3* in the first column will be displayed, and other rows of data are hidden. Copy the displayed (filtered) data to a fresh worksheet.

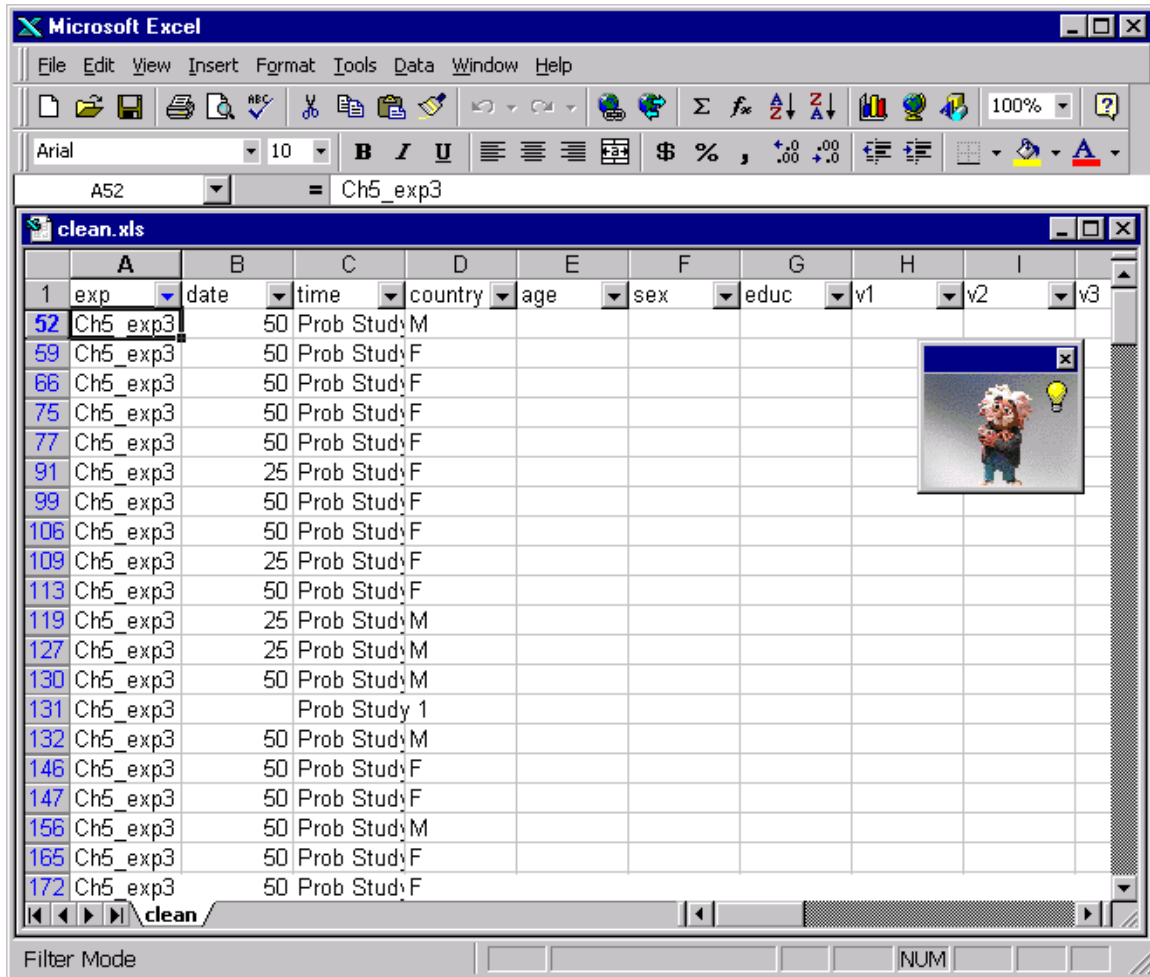
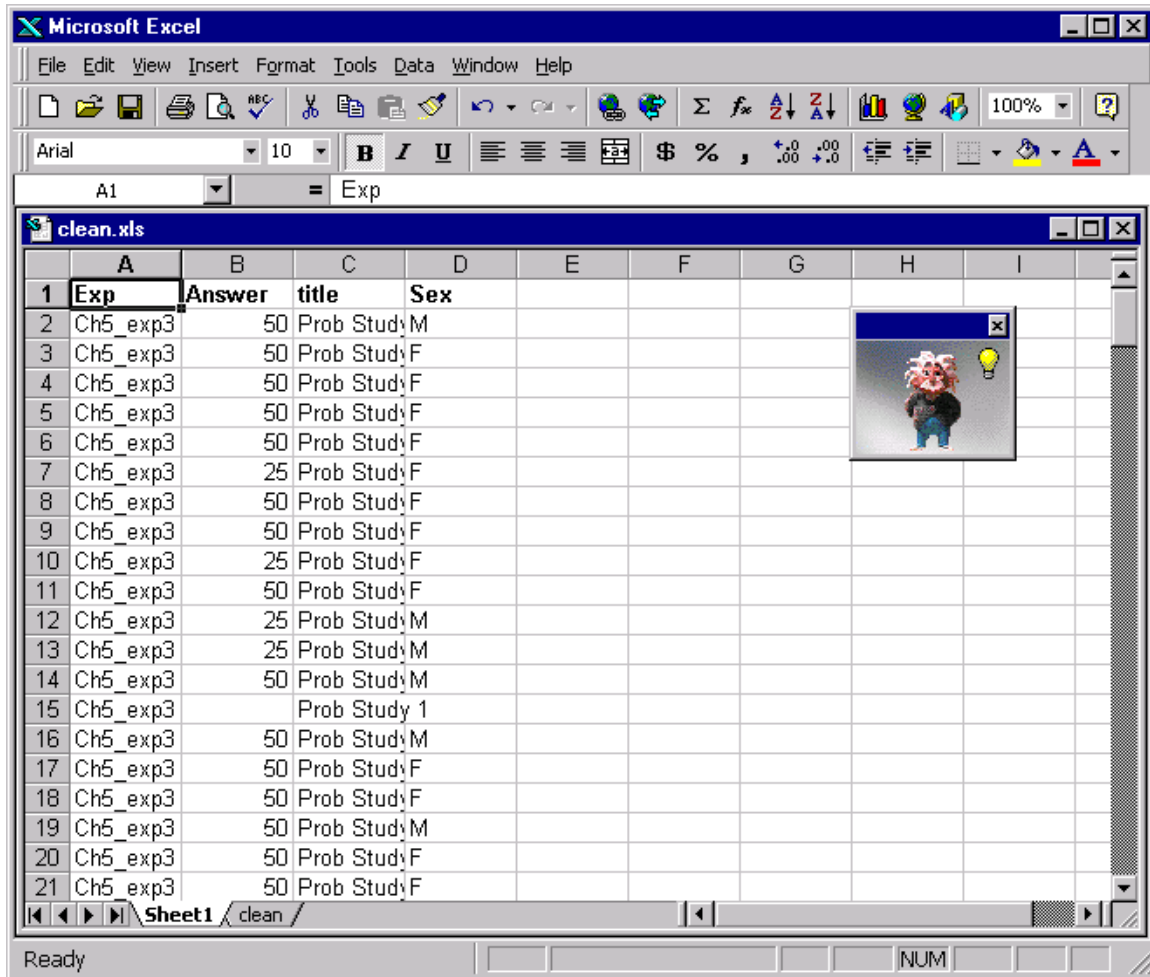


Figure 6.9. The data for the probability experiment have been copied and pasted into a fresh worksheet. Note the tab labeled “*Sheet1*” at the bottom of the worksheet. Double click on the tab and rename it *Ch5_exp3*. Clicking the tab marked “*clean*” will return to the (filtered) sheet of all data. Repeat the process to filter *Ch5_exp4*, *Ch5_exp5*, and *Ch5_exp6* to separate worksheets.



F. A Pivot Table for the Probability Problem

The experiment title (*Prob Study*) in Column C is always the same, so there would be no loss to delete it. You can delete Column C by clicking on the **C** at the top of the column (which will highlight the column) and then choosing *Delete* from the **Edit** menu. Be careful not to delete the whole sheet! If you do make a mistake in Excel, you can undo it (except in certain cases). To undo something, you can select *Undo* from the **Edit** menu (you can also click the undo arrow, which curves to the left, from the toolbar; if that arrow is dimmed, you can not undo). You can undo several steps by selecting *Undo* again and again. That might come in handy someday when you make a big mistake! For now, just undo deleting column C, for practice. If you change your mind again, you can select *Redo* (to undo the undo) from the **Edit** menu. Try it. You can also use the arrow that curves to the right. Double click on *Sheet1* on the tab at the bottom of the screen. Then give it a useful name, like *Ch5_exp3*.

When examining data, you may find that some people did not respond at all. You can either analyze the data with these blanks in the file, or you can go through and delete them row by row, using the same method as just described for the columns. To delete a row, click on the row number (which selects the row), then select *Delete* from **Edit** menu.

The nonresponse data mostly come from people who glance at the page and push the submit button to see what will happen. They get a “thank you,” and then they push the “back” key to come back to actually read the instructions. In most cases, the very next line of data is from the same person. Later, you will learn methods for detecting when this happens that will also help you spot when the same person sends the same data twice.

To analyze these data, you will construct a Pivot Table. First, click in cell A1. Then, from the **Data** menu, select *Pivot Table Report*. A Wizard will pop up with the ungrammatical message, “Where is the data you want to analyze?” (It should say, where *are* the data you wish to

analyze. The word “data,” like the word “people,” is plural; “datum” is singular, as is “person”). If *Microsoft Excel list or database* is selected (which it probably will be), then click *Next*. It will then say (again ungrammatically), “Where is the data you want to use?” It will probably list the correct cell entries (A1:D209, unless you deleted column C, in which case it will say A1:C209). If needed, select the proper data with the mouse. When the “ants” show that the data are properly selected, click *Next* again.

[Insert Figure 6.10 about here.]

In the third step, the screen in Figure 6.10 appears. With the mouse pointer, drag *Sex* from the right to the COLUMN, drag *Answer* from the right to the ROW, and drag *Answer* again from the right to the center of the table, labeled DATA. You may need to double click on *Answer* in the DATA field, and then select *Count* from the menu, as shown in Figure 6.11. When this is done, click *Next*. It will say “Where do you want to put the pivot table?” and you should select *New worksheet*. Click *Finish*, and Excel will then create the Pivot Table shown in Figure 6.12, on a fresh worksheet.

[Insert Figures 6.11 and 6.12 about here.]

Note that there are a grand total of 200 participants who gave an answer. The bottom line shows that there were 112 females, 79 males, and 9 left *Sex* blank (59% of the 191 who gave their gender were female). The column on the right shows that 20 responded 25% and 175 responded 50%. Only one person gave the right answer, 33%, and 4 gave other answers. These results are what one usually finds among students who have not yet taken a course in critical thinking or elementary probability—very few get the correct answer until they have taken one of these classes.

Figure 6.10. Step 3 of the *PivotTable Wizard*. In this panel, you can drag variables to the rows and columns, and drag the dependent variable to the center of the table. If the *DATA* field is labeled as *Sum of Answer* or some other function, double click on it, and select *Count of Answer* from the list, as shown in Figure 6.11.

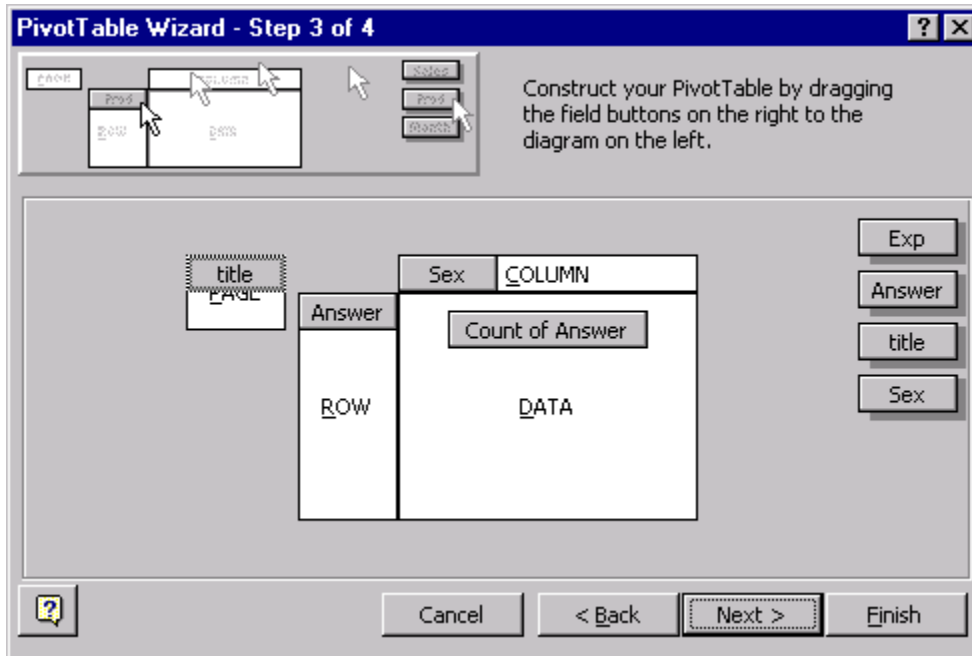


Figure 6.11. Double clicking on the dependent variable in the *DATA* field brings up this dialog box. Select the statistic wanted, in this case, *Count*, and click *OK*.

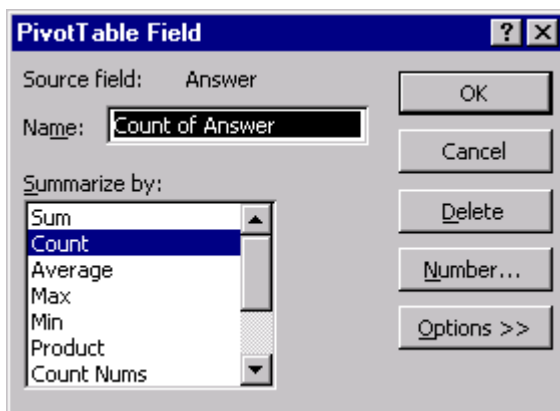


Figure 6.12. PivotTable of Answers by Sex. Note that there are a total of 200 responses, but only 1 person gave the right answer, which is 33.

Microsoft Excel - clean.xls

File Edit View Insert Format Tools Data Window Help

PivotTable

Arial 10 B I U

A3 = Count of Answer

	A	B	C	D	E	F	G
1	title	(All)					
2							
3	Count of Answer	Sex					
4	Answer	F	M	(blank)	Grand Total		
5	2	1			1		
6	12		1		1		
7	25	9	10	1	20		
8	33		1		1		
9	40	1			1		
10	50	100	67	8	175		
11	75	1			1		
12	(blank)						
13	Grand Total	112	79	9	200		
14							
15							

Sheet1 Ch5_exp3 clean /

Ready NUM

Save your Excel file as a workbook with the filename, *clean.xls*, on your floppy or hard drive. Click on the tab for *Ch5_exp3*, and save that worksheet as a CSV file for use later.

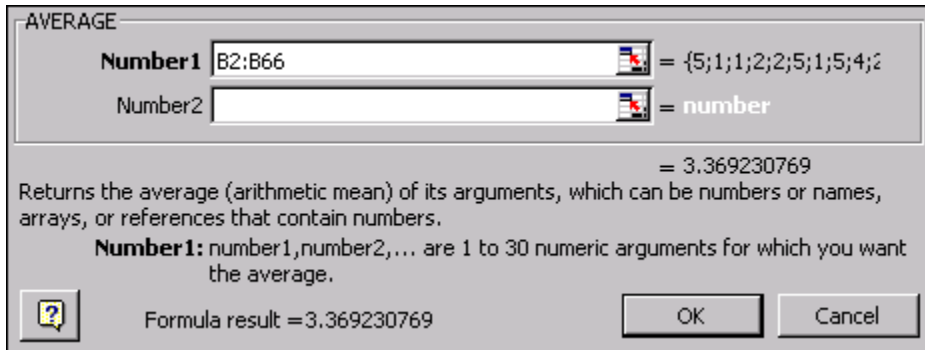
G. Analysis of St. Petersburg Paradox

In Chapter 5, it was hypothesized that different response procedures (different select menus or a text box) might produce different results for the value of the St. Petersburg gamble. To examine the data, there will be three steps. The first step is to use Autofilters to filter the data for each version of the St. Petersburg paradox into separate worksheets. Second, you will find the mean and median value of the St. Petersburg gamble with each procedure. Third, you will learn to conduct a statistical test known as the "t-test" to see if the difference between two procedures is significant.

Click on the tab labeled *clean*, which is the original worksheet with all the data. Using *Autofilter* in the first column, *exp*, select *Ch5_exp4_StPete1-10*. Copy the data, insert a new worksheet, and paste the data. Then go back to the data, click the filter again, and select for *Ch5_exp5_StPete1-512*. Next filter the data again, and copy the results for *Ch5_exp6_StPete_box*, which used a text box that allowed the participant to type an answer in the box provided. Recall that Experiments 4 and 5 used selection lists with different response options.

Next, enter names for the variables in Columns A, B, C, and D in Row 1. Name these variables **Exp**, **Value**, **Sex**, and **Age**. Enter these names on all three worksheets. The first task will be to find the Mean **Value** of the St. Petersburg gamble in each of the three methods. To do this, select the sheet for *Ch5_exp4*. Then click in a cell below the data in Column B (e.g., B68). Click the = sign in the formula line (or select *Function...* from the **Insert** menu), and a list of functions will appear. You may need to click *More functions* to find the function for *Average*. When you select it, the dialog box in Figure 6.13 appears. Insert Figure 6.13 here.

Figure 6.13. Select a cell, then select *function* from the **Insert** menu. From the list of functions, choose *AVERAGE*. In this case, you want to average cells B2 to B66. Note that the *Formula result* at the bottom shows the calculation before you click *OK*. Clicking *OK* inserts the value in the selected cell. Clicking the question mark (lower left) provides more explanation about this step.



The dialog box (Figure 6.13) appears, asking you to specify the range of values for the average. You can either type in the range (B2:B66), or you can use the mouse to select the data you want averaged (which will enter the appropriate range in the box). Check if everything looks ok, then click *OK*. Note that the answer is shown in this dialog box, before it is entered in the cell. When you click *OK*, the value is placed into the selected cell in the worksheet. In the neighboring cell, A68, type “mean=”, to remind yourself that B68 contains the mean.

To find the median, follow the same steps, except choose the *MEDIAN* function from the list of functions. Be careful to enter the same range of values from which to compute the median (do not include the mean as another number among the data). The median and mean are not necessarily equal, and in this case they are different. The mean is found by calculating the sum of the numbers, and dividing the result by the number of cases. The median is the value of the response for which an equal number of cases give smaller and larger answers. The mean of 1, 4, and 10 is 5 (the sum is 15, divided by 3, the mean is 5); the median of the same three numbers is 4, since an equal number (one) is above and below 4. The mean and median calculations for *Ch5_exp4*, where response selection included responses from \$1 to \$10, are shown in Figure 6.14. Follow the same procedures to find the mean and median age of the participants in this condition.

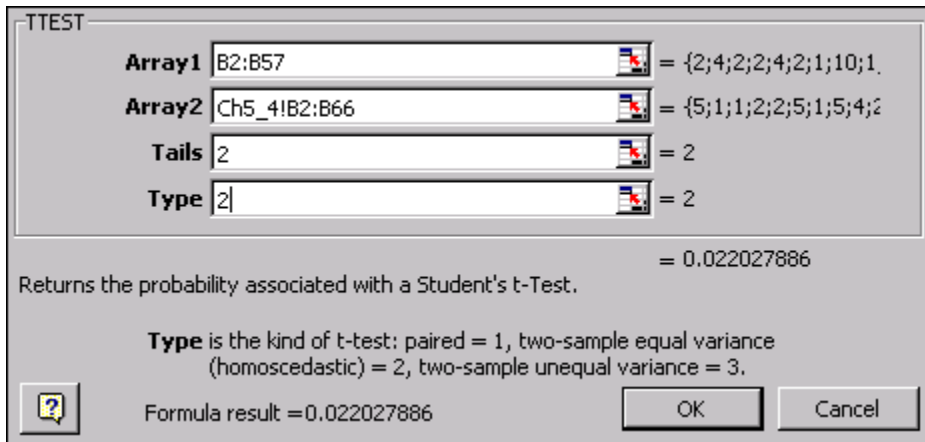
Insert Figure 6.14 about here.

Select Cells A68:B69, increase their font sizes, and click the **B** to make them bold. Next, click the drop-down selection arrow next to the paint can (Formatting Toolbar) and choose a yellow fill color, to highlight the mean and median. Now click on the tab at the bottom of the screen to switch to *Ch5_exp5*. Find the mean and median of this condition. You will find that the median is still \$2, but the mean on this sheet is \$21.08, which is considerably higher than the mean for *Ch5_exp4*, which was \$3.37! In both cases, about half of the subjects responded that the St. Petersburg gamble is worth \$2 or less. Because the arithmetic average is computed by summing up all of the numbers and dividing by the number of cases, it can be “pulled” toward a few large numbers. In this case, there are a few large numbers in Experiment 5. Examine the data, and you will see that a few people indeed said the gamble was worth \$512. Those cases pull the mean upwards.

Do the same for *Ch5_exp6*, and you will find that in the condition where the participants typed in their responses that the median is also \$2, but the mean is \$8.06.

To find out if means are significantly different in two conditions, we can calculate a *t*-statistic. To calculate a *t*-test between two conditions, click in a cell (where the *p*-value for the *t*-test will be inserted); next click the “=” sign in the edit line (or choose *Function* from the **Insert** menu), and select *t-test* from the list of functions. The dialog box of Figure 6.15 appears. Then select B2:B57 for *Array1*, or type in B2:B57 in the space provided. Click in the box for *Array2*, then click the tab at the bottom of the screen to select *Ch5_exp4*. Then select the data from that worksheet (B2:B66). Next click in the *Tails* box (Figure 6.15), and enter 2 for the number of tails, and then enter 2 in the *Type* box, to indicate that we have two independent groups with equal variances. Click *OK* and the computed value of *p* will be inserted in the cell selected.

Figure 6.15. The dialog box in Excel for a *t*-test. This function returns the probability level, rather than the value of *t* itself. Another function in Excel will compute the *t* value itself from the probability.



The value of the calculated p represents the probability that if one randomly selected two samples of data from the same population, that the difference in sample means would be as large (or larger) than the difference observed. This is the probability of the observed difference in means given the null hypothesis. Note that $p = .02$. Because this value is less than $\alpha = .05$, the significance level, one can reject the hypothesis that the response format had no effect in favor of the theory that the numbers come from different distributions. Presumably, the difference in response procedures caused a significant difference in means.

Although the means differ significantly, the medians are the same in all three versions of the experiment. Most people are willing to pay very little for this gamble, even though it has infinite expected value. Although the EV is infinite, most people say they would not pay more than \$4 for the chance to play the St. Petersburg gamble. These judges, like those in the days of Bernoulli, apparently do not judge a gamble by its expected value.

H. Introduction to SPSS

The Statistical Package for the Social Sciences (SPSS) is a popular program for statistical analysis, tables, and plotting of data. Whereas Excel is more widely used than SPSS in the business world, SPSS is more widely used in the academic world of psychology.

Start SPSS. You may get a dialog that asks what you want to do. If you are given the option of running the tutorial, run it and learn about SPSS. You can also run the tutorial from the **Help** menu. When you are done with the tutorial, close the tutorial screen and examine the data editor. (You will also get to the data editor if you close the box that offers you the tutorial, or if someone else has modified your lab copy to skip that dialog).

The grid in SPSS looks similar to Excel's, and it is, except for a very important difference. Variable names in SPSS should *not* be entered into the grid (as they are in Excel), but instead, you enter their names by double-clicking on the VAR at the top of each column. Unlike Excel,

variable names are limited to 8 characters, and must not include spaces. When you double-click on VAR, the *Define Variable* dialog box appears. (You can also get to this dialog for a variable by clicking in its column, and choosing *Define Variable* from the **Data** menu.)

To enter the heights data in SPSS, you will need three variables, the person's name, height, and weight. A person's *name* is a String variable, a variable consisting of letters and numbers. To change the name of the variable in the *Define Variable* dialog, type in the variable name (in this case, *name*) over *Var001*, the default. To change the variable from numeric to String, click on *Type*. A new dialog appears. Click the radio button next to *String*. You could also change the number of possible characters. In this case, all of the names are 8 characters or less, so if it shows 8, you can click *Continue*. Then click *OK*.

For the second variable, double-click on VAR at the top of the second column, and change the name (from *Var002*) to *height*. Change the *Type* to *Numeric* (8 columns with 2 decimals would do), click *Continue*, and change the measurement to *Scale*.

A nominal scale is a variable such as a list of names; an ordinal scale measures order, but does not retain interval information; a scale retains numerical information beyond the ordinal level. Height and weight are good examples of scales. An example of an ordinal scale would be a list of who took first, second, third, fourth, etc. in a race. If you knew the time it took each contestant to complete the race, however, you would have more than ordinal information because the times give information about the differences, or intervals, between the racers. The times also contain the ordinal information, since the person with the lowest time was first, the next lowest time was second, etc.

Do the same for the third column, as for height, but give it the name, *weight*. Next, enter data in the same manner as in Excel. As in Excel, *Enter* will drop the selected cell to the next row, and *Tab* will cause the selected cell to move to the next column. Save the data by selecting

Save As from the **File** menu. Save it as an SPSS data file, with the name *heights.sav*. On PC, the extension *.sav* will be added for you; on the Mac, you can add the extension.

I. Importing CSV Data to SPSS

Start SPSS, and view the data editor. From the **File** menu, select *Read Text Data*. Open *Ch5_exp3.csv*. Remember, you need to select *All Files(*.*)* in the open dialog box to see the file. When you select the file to open, the Data Import Wizard appears. The first screen is shown in Figure 6.16. The data are not in preformatted columns, so check “No.” The data can be previewed in the mini-window. Click *Next*. The next steps are shown in Figures 6.17, 6.18, 6.19, 6.20, and 6.21; follow the steps suggested in the figure captions. When you press *Finish*, you may see a list of warnings in the Output window. From the **Window** menu, select the *data editor*.

Insert Figures 6.16—6.21 about here.

Figure 6.16. Text Import wizard in SPSS. The data are not in a predefined format, in which each item is in the same number of columns, so check No. A preview of the file is shown in the mini-window. Click *Next*.

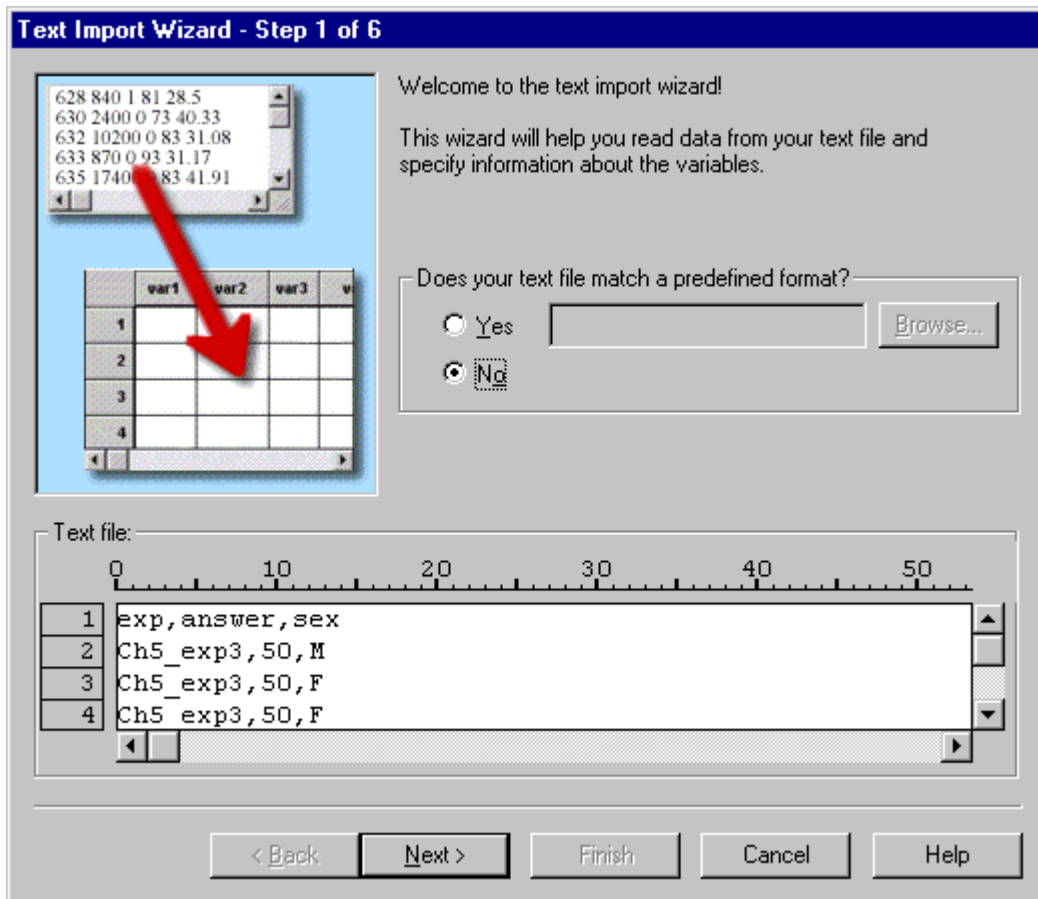


Figure 6.17. Step 2 of the Text Import Wizard. Click to indicate that the data are delimited (they are delimited by commas), and *Yes*, variable names are in the first row. Click *Next*.

Text Import Wizard - Step 2 of 6

How are your variables arranged?

Delimited - Variables are delimited by a specific character (i.e., comma, tab).

Fixed width - Variables are aligned in fixed width columns.

Are variable names included at the top of your file?

Yes

No

Text file:

1	exp, answer, sex
2	Ch5_exp3, 50, M
3	Ch5_exp3, 50, F
4	Ch5_exp3, 50, F

< Back **Next >** Finish Cancel Help

Figure 6.18. Step 3 of the Data Import Wizard. In this case, data begin in line 2, each line represents a case, and all of the cases should be imported. Click *Next*.

Text Import Wizard - Delimited Step 3 of 6

The first case of data begins on which line number?

How are your cases represented?

Each line represents a case

A specific number of variables represents a case:

How many cases do you want to import?

All of the cases

The first cases.

A random percentage of the cases (approximate): %

Data preview

0 10 20 30 40 50

1	Ch5_exp3, 50, M
2	Ch5_exp3, 50, F
3	Ch5_exp3, 50, F

< Back Next > Finish Cancel Help

Figure 6.19. In Step 4, indicate that a comma is used to separate values, and examine the preview. If Space or Tab is checked, be sure to remove the checks. Click *Next*.

Text Import Wizard - Delimited Step 4 of 6

Which delimiters appear between variables?

Tab Space
 Comma Semicolon
 Other:

Data preview

exp	answer	sex
Ch5_exp3	50	M
Ch5_exp3	50	F
Ch5_exp3	50	F
Ch5_exp3	50	F
Ch5_exp3	50	F

< Back Next > Finish Cancel Help

Figure 6.20. In Step 5, click on each variable, and specify its name and data type. The *answer* is numeric, and *sex* is a string variable. When you have clicked on each variable and checked its name and type, click *Next*.

Text Import Wizard - Step 5 of 6

Specifications for variable(s) selected in the data preview

Variable name: Original Name: answer

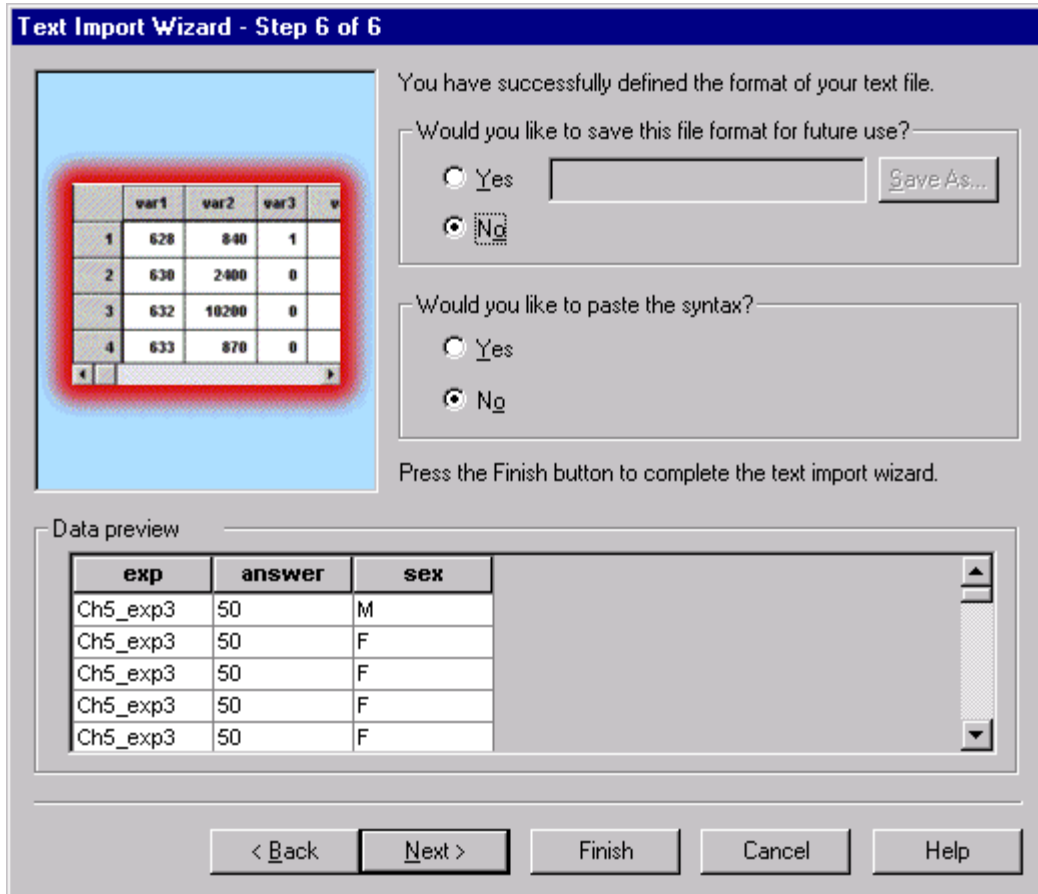
Data format:

Data preview

exp	answer	sex
Ch5_exp3	50	M
Ch5_exp3	50	F
Ch5_exp3	50	F
Ch5_exp3	50	F
Ch5_exp3	50	F

< Back Next > Finish Cancel Help

Figure 6.21. In the sixth step you can save the file format and syntax. In this case, just press *Finish*.



J. Crosstabs in SPSS

In general, to do a statistical analysis in SPSS, you will select the type of analysis from the **Analyze** menu. If you have an earlier version of SPSS than 9.0 (or if you have SPSS for Mac), this menu is labeled the **Statistics** menu. To do crosstabulations, from the **Analyze** menu select *Descriptive Statistics*, then *Crosstabs...*, as shown in Figure 6.22. Then follow the steps illustrated in Figure 6.23. The result is shown in Figure 6.24, after enlargement of the fonts. Note that this same table was created by Excel using the technique of Pivot Tables, as shown in Figure 6.12.

Insert Figures 6.22—6.24 about here.

On the CD are data files that contain additional cases for the St. Petersburg experiment. These are labeled *StPeteAll.csv*, and *StPeteAll.sav*. Open *StPeteAll.sav* in SPSS. To compute a t-test by SPSS, from the **Analyze** menu, select *Compare Means*, and then select *Independent Samples T test*. Use a group variable (the experiment number, labeled group, is in Column 5 of the data) to define the groups. The output in *StPeteAll.spo* gives the results of t-tests between each pair of groups by SPSS. You can compare your results with this output.

Figure 6.22. To do Crosstabs, select *Descriptive Statistics* from the **Analyze** menu, then *Crosstabs...* from the submenu.

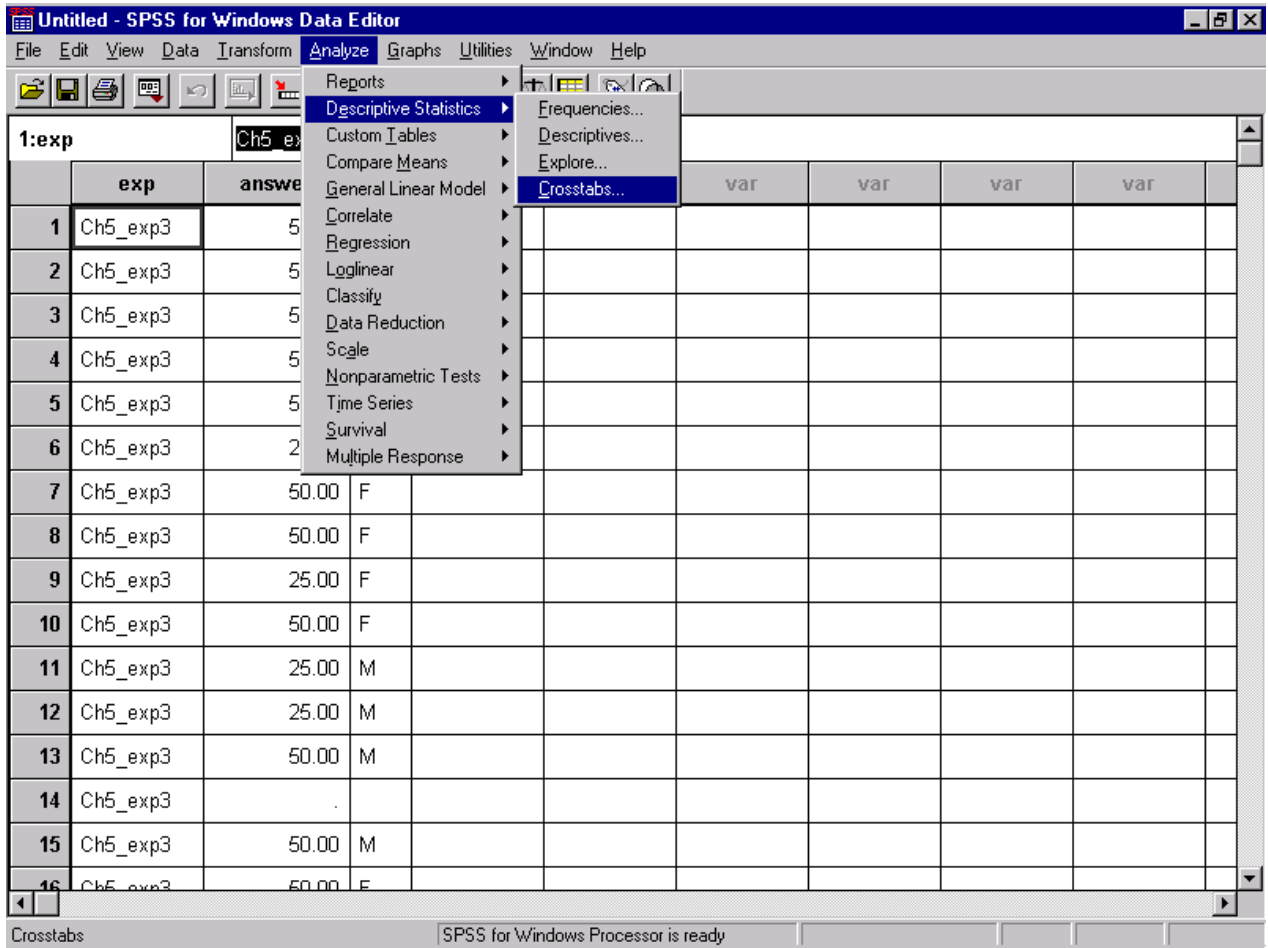


Figure 6.23. To choose *answer* as Rows, click on its name in the left display to select, then click on the arrow to send it to *Row(s)*. Then select *sex* for columns, by the same method, clicking the arrow to send it into the box for *Column(s)*, as shown. To undo an error, click on a variable and the arrow will send it back to the list of variables. Click on the buttons for *Statistics...*, *Cells...*, and *Format...* for other options. For now, press *OK*, to produce the results in Figure 6.24.

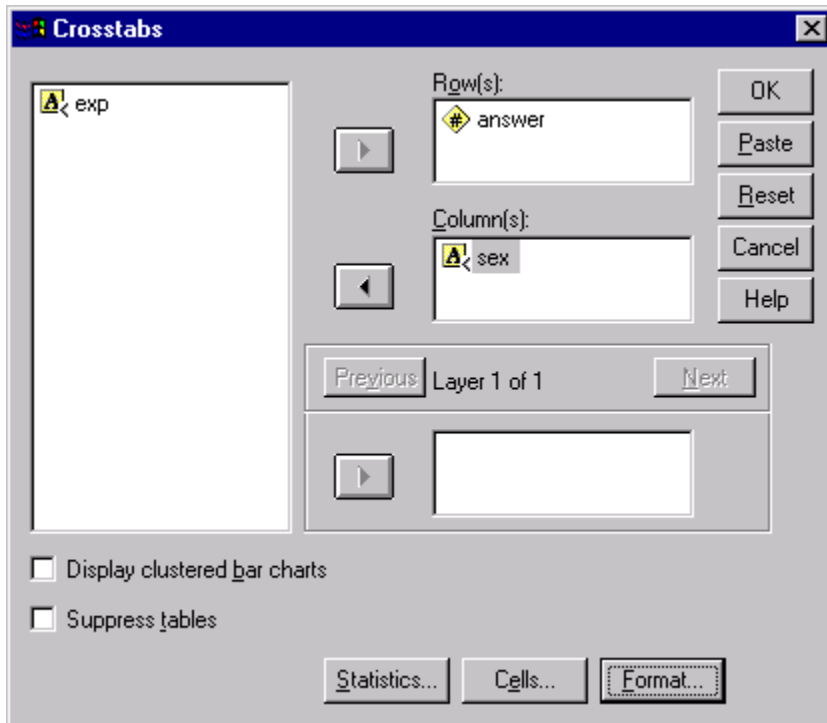


Figure 6.24. The ANSWER*SEX Crosstabulation table, produced by SPSS, provides the same results as the Pivot Table of Figure 6.12, produced by Excel. By double clicking on the table, fonts can be enlarged and formats changed.

ANSWER * SEX Crosstabulation

Count		SEX		Total
		F	M	
ANSWER	2.00		1	1
	12.00		1	1
	25.00	1	9	10
	33.00		1	1
	40.00		1	1
	50.00	8	100	175
	75.00		1	1
Total		9	112	200

SPSS for Windows Processor is ready H: 236 , W: 317 pt.

K. Summary

You should know how to open a data file, such as *data.csv*, in Excel, how to save it as an Excel workbook, and how to save a worksheet as a CSV file. You should know how to use *AutoFilters* to filter data from one experiment and how to insert them on a fresh worksheet. You should know how to construct a PivotTable report, how to find the means and medians, and how to construct a t-test of significance between two sets of data.

You should also know how to create a file of data in SPSS, open a CSV file in SPSS, and how to construct crosstabulations that are the equivalent of the Pivot Table of Excel.

L. Exercises

1. In SPSS, import the file *Ch5_exp3.csv*. Create a crosstabulation showing *sex* by *Answer* without referring to the book. Do the same in Excel, also without looking in the book. If you need to peek at the book, do so, then start over without the book and see if you can do it on your own.
2. On the CD is included a file, *StPeteAll.csv*, that contains additional data on the value of the St. Petersburg paradox. In this file are data from all three experiments, *Ch5_exp4*, *Ch5_exp5*, and *Ch5_exp6*, with the experiment listed in the first column. Import these data to Excel and perform the same analyses that were illustrated in the chapter. Hint: you can to Cut and Paste the data for different experiments to arrange them in the same way as was done in the chapter. However, that is not really necessary. You can check your results against the results in *StPeteAll.xls*, which is also on the CD.
3. Import the file, *StPeteAll.csv*, to SPSS. Analyze the St. Petersburg data in SPSS. To compute a t-test in SPSS, define a new variable called *group*. Assign this new variable the value 6 for *Ch5_exp6* (box), assign the value 4 for *Ch5_exp4* (1-10), and assign the value 5 for *Ch5_exp5* (1-512). Now, from the **Analyze** menu, select *Compare Means, then Independent-samples T*

test. In the dialog box, insert *Value* (the judgment of value of the gamble) as the *Test variable*, then insert *group* as the grouping variable. Click the button to define groups, and then select specified values. Let the values be 4 and 6 for groups 4 and 6, respectively. That will compare the text box with the 1-10 scale. You can repeat the process to get each of three comparisons of means. Explain why the biggest mean difference is not necessarily the "most significant" according to the p-value.

4. In SPSS, create a crosstabulation of *group* (columns) by *value* (rows). This crosstabulation will give you an idea of how the numbers for the different groups were observed.
5. Create the crosstabulation of *experiment* by *value* with Excel, using Pivot Tables. Use *StPeteAll.csv*. Use *exp* as columns, *Value* as Rows, and *Count of Value* as the Data.
6. In Excel, also create a Pivot Table with the Rows for *sex*, Columns for *exp*, and put the mean (Average) *Value* in the data field.