Using Microsoft Excel to Manage and Analyze Data: Some Tips

Larger, complex data management may require specialized and/or customized database software, and larger or more complex analyses may require additional commercially available statistical software, such as SPSS, but many basic evaluation data collection, management and analysis tasks can be done by hand and/or with readily available software such as Microsoft Excel or Access. To illustrate this see the special Excel file (excelmaster.v3.xls) developed for use with these notes. The file includes multiple worksheets including a master worksheet and examples for you to use.

1. Creating a Data File/Automatic Frequency Calculation

One of the worksheets in the Excel file is a survey EXAMPLE that is described in #2 below, and another is a blank MASTER worksheet that you can use for your own survey analysis. Be sure to make a copy of the file – or at least the MASTER worksheet so that you can use it over and over. The following instructions tell you what's in SURVEY EXAMPLE worksheet and how you could, using 4 easy steps, make a file of your own. You can also just customize the MASTER worksheet and use the file WITHOUT ENTERING ANY FORMULAS. Here's all you have to do.

1) Make a copy of the MASTER worksheet

2) Using your copy, change the names of the variables in all the columns to match your survey. (Variable names can either be question numbers or some short word that helps you remember what the question addressed, e.g., Age, Q5). The master worksheet has spaces for 15 variables – if your survey has more, simply copy all of column P and then paste it in as many cells to the right as you need. As above, change the names of the variables in each of the new columns you have added.

3) Label all the variables. In each of the cells in line 1 your will see a little red triangle. That is an empty comment box. If you scroll your cursor across line one you will see the empty boxes. Place your cursor in the cell you want to label (in line 1) then right click and select from the drop down menu EDIT COMMENT. That puts you inside the comment box and you can type whatever you like to describe the variable – in the example, the whole question or some important phrase are written. You may also want to use some of the data validation options described later in this attachment to make data entry simpler and more accurate.

4) Count up your surveys and see how many cases you have. Each survey is a case and will be entered into one Excel line. (It is a good idea to number each of your surveys too. That is called a case number and it helps you tell one survey from the next and to find them again if there are data entry problems.) The MASTER worksheet is only set up for 20 cases, but you can make it hold many more. Just place your cursor in a line above the calculations (line 21 for example) and enter insert row (use your Microsoft Excel guide book if you don't know how to insert multiple rows). Repeat the inserting process until you have enough rows for all your data.

5) Now fix the labels on the answer choices. Look at the percentages section of this MASTER worksheet (lines 38 – 43). Again you will see the blank comment boxes. Label the answer choices here as you did for the variable names. For example if answer choice 1 = Female, then type FEMALE into the comment box.

PLEASE NOTE: If you need more than 6 answer choices you will need to add some calculation rows. For that you will have to refer to how the calculations are inserted (see Calculating Frequencies Using Microsoft Excel below) or your Excel documentation, or access a TA provider. Most surveys only have up to 6 answer choices and many questions only have two (e.g., YES – NO, MALE – FEMALE) or three.

6) Enter your data into the rows of the spreadsheet. The encoded answers from each respondent's survey should be entered into each row of the spreadsheet. If the answer choices are not numerically encoded on the survey itself, make yourself a survey codebook and then use that for data entry (e.g., Yes = 1, No = 2, Female = 1, Male = 2, Excellent = 3, Okay = 2, Not so Good = 1, etc. – see also the data validation section in this attachment).

The frequencies (counts) and percentages will show automatically below. There are two check sum rows on this spreadsheet too (lines 35 and 45) – the data in each cell in the first check sum row (line 35 in the example) should be equal to the number of respondents. The data in each cell in the 2nd check sum row should always be 100%. If either of these is not true, then you have a bad formula and it will need to be repaired. See the *Calculating Frequencies* section (following) and you will see how all the formulas are constructed, or consult your Excel documentation or a TA provider.

PLEASE NOTE Also:

1) You will NOT be able to simply print out this spreadsheet and use it to report survey results. You will still have to take the percentages and organize them into a table or written report, but you won't have to do any of the calculations by hand.

2) If your survey had any "bail out" answer choices – e.g., *I'm not sure*, or *Didn't do this*, these answers are often removed from the analysis before reporting. Any missing data will automatically be handled, but answer choices that suggest the respondent could not answer are assigned a code and you will have to decide if you want them to be included. If you need help thinking about this – PLEASE REVIEW THE DISAGGREGATING DATA section of this document or contact a TA provider.

3) This survey tool will allow you to conduct some simple cross-tabulations. If you are interested, be sure to read the following section on data disaggregation.

More tips follow ➔➔➔

2. Calculating Frequencies Using Microsoft Excel

Survey Analysis EXAMPLE

The EXAMPLE worksheet in the EXCELMASTER.xls file has 30 cases of "data." Each line or row represents the responses of one individual to a survey. The "survey" had 40 questions and each respondent also reported his/her age and gender. Case numbers were assigned for each respondent. The columns in this spreadsheet represent the questions on the survey. In analytical terms they are known as the variables. The labels for each question (variable) are shown in row 1 of this spreadsheet. If you scroll your cursor along the row you will see comment boxes that show the questions. At the bottom of the spreadsheet there are calculations starting at row 34. These calculations include: the number of respondents answering each question, the frequencies or counts of respondents that chose each answer choice, and the percentages of respondents who selected each answer choice. The percentages are what is typically reported in basic survey analyses. If you scroll across rows 51 through 56 you will see comment boxes again that define each of the answer choices. For example, answer choice 1 for the variable SEX = female. The following explains how the calculations were done.

STEP 1: DETERMINE THE DENOMINATOR. Scroll to the line after all the entered data (line 34 in the example file). Insert, function and then select COUNT and hit enter. A dialog box will pop up asking you to identify the range you would like to count (it will default to the range of all cells above the cell you have chosen as your starting place). Hit OK.

- The result of the function you have inserted is the denominator for that column
- Copy this cell and paste it into the cells at the right for each variable column (ex. line 34, columns C – AQ).

STEP 2: CALCULATE THE FREQUENCIES using the COUNT IF Function.

a. Insert, function, COUNT IF, hit OK
b. Next you will have to identify the range in the dialog box, and then the criteria. TYPE IN RANGE (e.g., B2:B31), then TAB to Criteria, ENTER 1, then tab to OK, finally hit ENTER.
c. Copy this cell and past it into the cells at the right for each variable column
d. For all the other frequency calculations (lines 39 – 43 in the example), repeat steps b and c. Use the same column range B2:B31, but change the criteria to 2 then 3 then 4, then 5, then 6.

- At this point you have calculated all the frequencies for the different responses (i.e., the counts of all those who selected answer choice 1, choice 2 and so on). These frequencies show in rows 38 – 43.

STEP 3: CONVERT THE FREQUENCIES TO % (WHERE APPROPRIATE) Move your cursor down again until it is one row below the frequency calculations (b51 in the example).

a. Type in the formula to divide each frequency cell by the denominator and hit ENTER (e.g., = (B38/B34). Format the cell (see c below).
b. Type in the conversion formula for the rest of the frequency cells (from B51 through B56 in the example; i.e., B38/B34, B39/B34, B40/B34 B41/B34, B42/B34, B42/B34, B43/B34)
c. Format the results as percentages. Highlight the converted results (from B51 through B56 in the example), HIT FORMAT on tool bar, select CELLS, then scroll to PERCENTAGE and click. Adjust the decimal places to 0, 1 or 2 depending on your need for precision (0 decimal places is usually sufficient for the survey data we work with). Then hit OK. Copy this cell and past it into the cells at the right for each variable column.

STEP 4: LABEL THE CELLS: Place your cursor in the cell with the percentage, hit INSERT, THEN COMMENT. Below the default comment type in what the cell represents.

More tips follow ➔➔➔

3. Using Excel to Store Survey and Administrative Data (Validating Data)

Scroll through the worksheets in the special file until you see the one entitled Admin Data. This is a mock data-base that was developed to store administrative data i.e., background and other information about participants in a program. Once again, the rows represent cases (participants), and the columns are the variables (administrative data). The specific box where data is input is called a cell. Where the survey database variables captured responses to a questionnaire, this database stores relevant information about participants and their involvement in a program. Just like the survey database, however, the information has been encoded. For example, if a participant's age at the time the data was collected was 19, then he/she would have a "3" in column E. You will notice as you scroll through these cells that variables and code choices are shown to the right or left of the cell that needs to be filled in. If you try filling in a code that is invalid in a selected column, (e.g., insert a 4 in a cell where the choices are only 1, 2 or 3), a tone will sound and an error message appears -- *The value you entered is not valid. A user has restricted values that can be entered into this cell.* This prevents out-of-range data entry (it does not eliminate all errors since a 2 could be entered when the true code should have been 3, but it does eliminate totally invalid code use and allows you to restrict values for information). As you can see, much information can be stored in a database like this – **and it will work for both survey data and administrative data**. For those variables that are encoded, the same "calculator" can be appended to this database to calculate frequencies (for example, the proportion of participants who are certain ages or genders). Other analyses, such as calculating averages for the amount of fees paid, can be calculated with the data that are not encoded (remember calculating the frequencies of non-encoded data is a no-no). If you'd like to set up a database like this one, the steps are simple:

- Determine which data elements (columns) will be stored and what codes are needed.
- To restrict a cell put your cursor in it and click **Data** (from the toolbar) and then **validation** from the drop down list.
- Now you will see 3 tabs: *settings, input message,* and *error alert*. Pick *settings*. On the drop down list under validation criteria allow: select what's meaningful (for encoded data it will always be *whole number*). Following that you will be able to clarify the range of your data (*between, not equal to,* etc.) and specify and minimum and maximum value. Next select the input message to clarify acceptable input for the data category. Finally, click on the error alert screen and then click the box that says *Show error alert after invalid data is entered*. Then type in an error message, for example: STOP, you must use these codes 1 = YES, 2 = NO. After you have completed the last validation step you can return to the main database, copy the cell and paste it all the way down the column (and across rows as needed – if you have several data items which use the same codes). BE SURE TO SAVE YOUR WORK FREQUENTLY.

More tips follow ➔➔➔

4. Disaggregating Data

Disaggregating data is an important analytical process. Most of the time, participants in a program are not all the same and there is a need to look at the outcomes or responses of groups of participants separately or comparatively. For example, survey responses are typically disaggregated by gender, participant experience with the program and/or other relevant background characteristics (e.g., race/ethnicity, incoming achievement, etc.). **You can use Excel to calculate frequencies for disaggregated data.** Look again at the SVY EXAMPLE database. You will see that some of the respondents are female (SEX CODE = 1) and some are male (SEX CODE = 2). To look at the responses of only females or only males, simply make a copy of the whole database, sort the file and remove the responses you are not interested in. Follow these steps.

- **Copy the worksheet with all data**. Place your cursor in the far left corner of the worksheet (the only cell that does not have a row or column label) and click on it. This will highlight the entire worksheet. Then select EDIT and COPY.
- **Insert a new worksheet into the file**. While the worksheet is in memory, select INSERT from the toolbar and then click WORKSHEET. Once in the new worksheet, place your cursor in the first cell and then select EDIT and PASTE. Now you will have a copy of the original worksheet with all the data.
- **Sort the data by the relevant variable**. Highlight rows 2 through 31 (i.e., click and drag on those cells) and then select DATA from the toolbar. Now select SORT. In the drop down menu identify the variable you want to sort by – SEX in our example – by finding it in the drop down list or just typing it into the sort by box. This will organize all the data for females in the first rows of the database, and all the data for males in the next rows of the database. Anyone without a gender (SEX) code will be sorted to the bottom of the file. Data sorting is very valuable – don't forget to highlight all the rows in the file though so whole cases are sorted together.
- **Eliminate the rows with unwanted data**. Once you have sorted your file, highlight the rows with data you don't want – in our example the data associated with males or participants with no SEX codes. Now all that is left are the answers of the female participants.
- **Repeat the above process as needed**. If you wanted to now look at the data for males, for example, make another copy of the master, insert a new worksheet in the file and copy it in, sort the new copy by SEX, highlight and remove all data except that associated with males.

In the file you will see that there are example worksheets with the responses of females only and males only. These worksheets were constructed using the steps described above. TRY IT!

Disaggregating data can also be accomplished by using the Pivot Table function of Excel, but it is somewhat more difficult than the steps suggested above. If you are comfortable using Excel and its help function, be sure to try that option, just follow the drop down menus.

5. Using Excel to Develop an Evaluation (or program) Budget

To estimate how much labor and other expenses will be needed for an evaluation project and to determine overall costs, use Excel to calculate a proposed level of effort (LOE) chart and budget. Included in the attached Excel file is an LOE/budget example for a project spanning two phases (EVAL BUDGET Phase 1, EVAL BUDGET phase 2). Scroll through these. For more details on how to construct the LOE/budget, consult the *Participatory Evaluation Essentials Guidebook* (available online at www.brunerfoundation.org). Use the Excel master worksheet to view the projections and calculations and feel free to copy and customize it as needed – the basic categories and calculations are all there. **Please note that the budgets for Phases 1 and 2 are linked. This is another very valuable Excel Tool.** If there are any changes to one budget, the 2nd worksheet is automatically updated. You can link worksheets by doing the following: in the cell where you want to bring in data from another worksheet simply type ='worksheet name'!cell address. At the bottom of the PHASE 2 budget there are a series of grand totals. Please note the linked cells (B42, B46, B51).

6. Using Excel and Word Together to Construct Tables and Figures

The Microsoft Office software is set up to work together and you can produce a wide variety of communication enhancing images for your reports by combining Excel and Word. Below are a few quick suggestions regarding this usage of Excel. PLEASE CONSULT YOUR GUIDEBOOK FOR ADDITIONAL DETAILS AND EXAMPLES.

1. Format the table in Excel minimally and input your data. Use Excel to do as many calculations as you need. ** Note the value to constructing a table in Excel is that it can do more calculations, more easily, and it can convert the data into graphs fairly easily. IF YOU DON'T NEED TO DO ANY SOPHISTICATED CALCULATIONS, AND YOU ARE NOT PLANNING TO GRAPH YOUR DATA AND INSERT A FIGURE, JUST USE THE TABLE FUNCTION IN WORD.
2. Once your table is completed in Excel, simply copy it and paste it directly into your Word document at the point where you want the table to show. The table has now become a Word table and you can use Word to Modify it as you like (adjust fonts etc., sort data, improve labels)
3. If you want to also insert a graph, or to do that instead of inserting a table, switch back to Excel.
4. Construct the graph you desire in Excel using the Chart Wizard. Be sure to make the graph look in Excel, exactly as you want it to look in your document. (Also remember that if you are printing in black and white, you must choose colors with enough contrast to be seen in your document.)

Once again, simply block and copy the graph in Excel and then paste it into your Word document. Once the graph is in Word, you can click on the image and there are some limited editing functions within Word to further improve the appearance of your Graph.

