Tip Sheet

How to Set Up Data Entry Worksheets in Excel and Clean Data in Excel or Online

The following instructions were designed for setting up traditional data entry spreadsheets in Microsoft Excel. Many of the concepts listed below are also relevant for setting up surveys in online survey tools such as Survey Monkey or Survey Gizmo. Special instructions for online surveys are marked with this symbol: 💻.

Helpful terms to know:

- **Variable:** When we enter and analyze survey data, we must break down the survey into a set of variables. Examples of variables include gender, age, and level of satisfaction with a video. Survey questions are often variables, although it is possible for a survey question to be broken out into several different variables.
- **Respondent:** A respondent is the person who filled out (i.e., responded to) the survey. In evaluation research, this is generally a program participant.
- **Response option or, response category:** The possible "answers" a respondent can give to a survey question are called response options. For example, for the variable "Ohio DELTA site," the respondent can choose from the following response options: "Knox County," "Lucas County," or "Warren County."
- **Mutually exclusive:** When a survey question asks the respondent to "choose only one" of the response options, these survey options are said to be "mutually exclusive"—meaning that only one of the responses can be true. In contrast, questions that allow respondents to "select all that apply" must be treated differently when you are setting up variables.

Question type	Example	How to set it up as a variable for data entry in Excel	💻How to set it up as a question type in online surveys
"Choose only one" (mutually exclusive response options)	What school do you currently go to? ❒ Centerburg High School ❒ Danville High School ❒ East Knox High School ❒ Other: _____	Set up as one variable for "school" (one column of data) Note: If lots of kids wrote in names of schools in "other," then you will need to code those and/or create a column for typing in these names.	Survey Monkey: Select "Multiple Choice (Only One Answer), appears as "radio dial" Survey Gizmo: Select "Radio Button (single select)"
"Select all"	What would you do if you were being sexually harassed or bullied at school? ❒ Fight back ❒ Do nothing ❒ Tell my parents ❒ Tell a teacher or principal ❒ Try to keep myself away from them ❒Other: _____	Set each response option up as a separate yes/no variable For example: "Fight": 1=yes, 2=no "Nothing": 1=yes, 2=no Etc.	Survey Monkey: Select "Multiple Choice (Multiple Answer), appears as "check box" Survey Gizmo: Select "Checkboxes (multi-select)"

Step 1: Make a Data Entry Key (a.k.a. "Data Entry Master Sheet")

- Print out a hard copy of the survey
- Label each variable in colored pen
    - These labels will be your "variable names" and will become the column headers in your Excel sheet.
    - Variable names can be short, descriptive words (such as, "school," and "age"), or they can simply be the number of the question on the survey (such as Q1, Q2), Q3a, Q3b). Caution: If you plan to use any of the same survey questions on different surveys of differing lengths, then you should avoid using the "Q1" approach, as your question numbering may change.
    - Is the question a "select all that apply" or a "choose one answer" format? If your question allows for "select all that apply," then you need to set up each response option as its own separate yes/no ("dichotomous") variable.

- Label each response option in colored pen
    - In order to put the responses into the spreadsheet, they need to be turned into numbers. Write a number (1, 2, 3, etc.) next two each response option.
    - Use "0" for "not applicable" options

Step 2: Set up the columns in the Excel spreadsheet

- Open a new Excel spreadsheet and give it a name that indicates the survey and the date (such as SafeDatesPrePostSpring09.xls).
- Label the top of the first column "ID." This will be where you put each survey's ID number.
- Going from left to right along the top row, label the top of each column with names of your variables (from your Data Entry Key sheet).

Step 3: Put ID numbers on surveys

- Write an ID number in the top corner of each survey. Each ID number should be unique (don't re-use the same ID numbers.
- ID numbers are important to have in case you discover later that data were entered incorrectly. You can go back to the hard copy survey to correct data entry errors.
- If you are doing a linked longitudinal design where you need to match up data at the individual level, then it is very important to set up your ID numbers at baseline and to consistently use the same ID numbers in your data files.

Step 4: Give the surveys an "eyeball" check for possible data problems ("pre-cleaning" the data before entry or as you enter it)

Response problems to look for	How to correct them on the paper surveys before entering them into the spreadsheet, or during data entry
Respondent wrote something in for "other specify" that was actually listed in the survey question. For example: What school district is your child enrolled in? ❒ Centerburg ❒ Danville ❒ East Knox ☒ Other: Danville high school 💻 *This problem can happen with online surveys.*	Errors like this need to be "back-coded," meaning you cross out the "other" response and check the appropriate one. For example: What school district is your child enrolled in? ❒ Centerburg ☒ Danville ❒ East Knox ❒ Other: ~~Danville high school~~
Respondent selected two or more responses for a question that was "choose one only." For example: If I saw someone at school being threatened by their partner, I would tell an adult (choose one) ❒ Yes ☒ Maybe ❒ No ☒ Don't know 💻 *This problem can be prevented in online surveys if you choose the "radio dial" response type.*	You can only enter one response. Randomly select which response to enter into the spreadsheet. You can do this with a coin toss. For example, Heads= "maybe" and Tails= "don't know" Coin toss results in heads up Circle and enter "maybe" response and cross out "don't know" response Note: If a large number of respondents did this, then it probably means your survey instructions were not clear enough or that the question needs to be worded differently. In this case, you may need to treat each response option as it's own yes/no (dichotomous) variable. And then revise your survey in the future.
Unclear which number respondent circled or which box they checked. (example: "pregnant chad," sloppy circles) 💻 *This is not a problem in online surveys.*	Ask several different people in your office to see if there is any consensus about which option the respondent intended. If not, use the above coin toss method to select one response to enter.
Respondent skipped a question 💻 *This problem can be prevented in online surveys if you require a question to be answered.*	Leave the cell blank; do not enter anything. This is "missing data."
Respondent answered a question that was not applicable to them (for example, a male checking "not pregnant") 💻 *This problem can be prevented in online surveys by including "not applicable" or by creating a skip pattern.*	Cross the response out and do not enter it. Note: You can generally prevent this problem by including a "not applicable" response option, or by creating a "skip pattern."

Step 5: Enter the data

- With each row representing a different respondent, type in the numbers that correspond with the response options.
- If you have written responses to open-ended questions, you can enter those into Excel (format cell for "text" and check "wrap text"), or type them up in a Word document, or code them directly on the hard copy surveys.

💻 Instructions for downloading raw data from Survey Monkey into Excel

If you are using an online survey tool, such as Survey Monkey, you will not need to enter your data. However, you may need to download your raw data into Excel in order to do some types of analysis.

- Click "Analyze Results" tab
- Click the "Download Responses" button on the left
- Click the "All Responses Collected" button
- Select: Columns= "Condensed" (The condensed format is easier to use than the expanded format.)
- Select: Cells= "Numerical Value"
- Once Survey Monkey has prepared the file (this takes a few minutes), click the "Download file" button
- Open the file in Excel
- In Excel, select the first two rows, go to "Format Cells," select "Alignment" and check "Wrap Text." This will make the survey questions/variable names readable.
- In Excel, the numbers represent the response options in the same numerical order they were presented in the survey.

Once you have your data in Excel, you can analyze it in Excel, or upload it into StatCrunch.com, or export it to SPSS, SAS, or MyStat.

Step 6: Clean the data (systematically check for inconsistencies and errors)

- You can do "eyeball" checks for data errors by scanning the Excel columns. It often helps to "sort" the data by specific columns in order to easily spot incorrect values.
- When you run your initial analysis of the data (basic frequencies or percentages for each question), look for inconsistencies, odd patterns, or surprising findings. These may indicate systematic or one-time data entry errors.

Common data entry errors to look for	How to spot and correct them
Numbers typed in that are not on the Data Entry Key Master Sheet (incorrect values). For example, if the question is "gender" and the possible responses are "male"=1 and female=2, but a "5" is typed into the cell for that question.	You can spot this problem through a visual check or initial analysis. Look at the survey ID number and go back to the original survey to get the correct response. Type in the correct response.
Numbers transposed (shifted over or flipped digits)	You can spot this problem through in initial analysis. Look for odd patterns.
Missing data entered as "0"	Do a visual check to see if there are any "0"s for variables that do not have "0" as an option.
Reversal Items (positive direction/negative direction items) not entered correctly (see example below)	You can spot this problem through initial analysis. Look for odd patterns.

Example of "reversal item" attitude items

If you survey includes a set of items that express positive (desirable) and negative (less desirable) attitude statements, be very careful in setting up the variables and entering the data. In the example below, the value "5" represents the most desirable response; for the first and third statements, 5= "strongly *disagree*," and for the second statement, 5= "strongly *agree*." When you are entering data from these types of questions form hard copy surveys into Excel, it is helpful to make a clear plastic overlay sheet with the response values written on it so that your values (such as numbering 1-5) are consistent.

10. How much do you agree or disagree with these statements?

e.	If a girl wears a short skirt or tight jeans, she is asking to be sexually harassed.	**Strongly Disagree 5**	Disagree 4	Not sure 3	Agree 2	Strongly Agree 1
f.	*Writing dirty things about someone on a bathroom wall at school is sexual harassment.*	*Strongly Disagree 1*	*Disagree 2*	*Not sure 3*	*Agree 4*	***Strongly Agree 5***
g.	If a girl says she is being sexually harassed and they boy says he is just joking, then it's not really harassment.	**Strongly Disagree 5**	Disagree 4	Not sure 3	Agree 2	Strongly Agree 1

Note: The italics, bold, and response option numbers are added here as they would be on a Data Enter Master Sheet. The survey itself that respondents get should not indicate which items are positive or negative.

💻 A note about cleaning data when using online survey tools

Most of the potential data errors discussed above are avoided when you use an online survey tool, with the following two exceptions:

- Respondent writing in something for "other specify" that was actually listed in the survey question. These errors will still need to be back-coded.
- Respondent answered a question that was not applicable to them. You can generally prevent this problem by including a "not applicable" response option, or by creating a "skip pattern." Survey Monkey and Survey Gizmo allow you to create skip logic if you subscribe above the basic level.

To do data cleaning in Survey Monkey:

- Click the "Analyze Results" tab
- Click the "Browse Responses" button
- View each respondent's entries and click the "Edit Response" or "Delete" buttons to make changes

If you pilot-test your online survey (which is advisable), be sure to delete your pilot respondents from the final data file if they were not really part of your intended respondent group.

Step 7: Now that your data are squeaky clean, you are ready to jump in and analyze your data!