

# Practical Data Cleaning Cheat-Sheet

## 19 Tips to Make Your Data Cleaning Easy

- |   |   |
|---|---|
| 1 - Record Data on Paper First            | This is where you get to design how you're going to collect and store your data, which gets you to thinking in detail about your study.   |
| 2 - Transfer Your Data to a Spreadsheet   | If you'd have entered your data directly into a spreadsheet without having a paper copy, mistakes made would be permanent.  |
| 3- Enter Your Data on a Single Worksheet  | Enter your data in a single worksheet. The amount of data you're going to collect is unlikely to be a problem for Excel, and if it is, you should probably be collecting your data by more automated means. |
| 4- Use a Unique ID Column                 | It's absolutely crucial that you have the ability to restore the original order, and for this we start by using column A as a Unique ID column.   |
| 5- One Column per Variable                | Don't enter more than one piece of information into a single cell.  |
| 6- Row 1 is the Variable Name             | The standard for pretty much all statistics programs is for the first row to be reserved for the name of the variable, so you should follow this too.   |
| 7- Every Cell Should Have Something In It | If you don't have any data for a particular cell, don't just leave it empty. There's no information in a blank cell.  |
| 8- Keep Great Notes                       | When you put codes in your cells you'll need to keep notes about what they mean. This is perhaps the biggest and best tip I can give you – KEEP GREAT NOTES!!!  |
| 9- Be consistent                          | A lot of the problems with data cleaning are caused because the data is not entered correctly or <i>consistently</i> in the first place.  |
| 10- Don't Guess                           | Data accuracy is really important, so when entering your data don't guess, approximate, round up or down – just enter the value exactly as registered on paper.   |

- |   |  |
|---|--|
| 11- Zero is a Real Number                         | Zeros are real numbers and will be included in any calculations you make on your data, so don't use the number zero as a code to mean 'No data'. |
| 12- Make a Copy                                   | Make a copy of your dataset so if you make any mistakes you'll need to be able to check back and make corrections.                               |
| 13- Clean Your Data in a Separate Worksheet       | Cleaning data is a messy process and you WILL make mistakes so do it in a separate worksheet!  |
| 14- Report Errors Back to the Original Source     | Next time you have to analyse some more data from the same source you'll have a lot less cleaning to do.   |
| 15- Use Excel Functions to do the Hard Work       | Entry errors happened because someone entered data manually, so why use the same method that got you into trouble to get you out of it?          |
| 16- Use Excel Formulae to do the Even Harder Work | I guarantee you won't regret the time investment made to learn the Excel formulae you need.  |
| 17- Keep a Code Sheet                             | For each variable you should list the codes that you've used and explain what they mean.   |
| 18- Identify Your Data Types                      | Take a little time to decide which of your variables are Ratio, and which are Interval, Ordinal or Nominal.                                      |
| 19- Check That Your Data are Sensible             | Real life follows rules, and your data must too.   |