

PEN Data cleaning procedure

Step 1: Data submission

Data should be submitted as compressed (.zip or .rar) files, this is because MS Outlook (my email client) blocks MS Access attachments, moreover, the databases will often exceed the attachment limit for most email clients and compressing them will help.

To compress your files, you can use [winzip](#) (shareware), or

- **ZipGenius**: (<http://www.zipgenius.it/eng/index.php>). Allows you to split the zip file (as would be the case if one file was larger than 10MB and you wanted to split it into smaller parts).

- **IZArc**: (<http://www.izarc.org/index.html>).

- **7-Zip**: (<http://www.7-zip.org/>).

Step 2: Data checking

Stata cleaning programs are then run on your data. The result of this process is bug report which lists bugs/problems. In cleaning, we identify

- missing data,
- inconsistencies (totals do not add up, leading questions show you have some products but there is no data for these products), and
- Potential data entry errors (or outliers).

Step 3: Bug fixing

Details on how to read and address the bug report can be found in the document “Reading the bug report”.

Note: Not all bugs can be fixed (for example, not all outliers are data entry errors) so, when re-submitting the “cleaned data”, make a note of what could not be fixed.

Step 4: Re-submission of Data

Having addressed the bugs, the “clean” data are resubmitted for re-checking. Steps 2 and 3 are then repeated until the data are clean (meaning, all bugs have been addressed or can be explained).

Step 5: Clearance, Analysis, Accounts

Once your data have been cleared, they will be archived into the global database. A cleaned copy will be emailed back to you (in Access, Stata, excel or any format you want). It is also at this point that CIFOR will contact you to arrange the payment of the Honorarium.