SEED SYSTEM SECURITY ASSESSMENT:

DATA MANAGEMENT FOR INDIVIDUAL FARMER INTERVIEWS (TOOL 4.1)

A guide for data managers: How to lead data entry and analysis using the automated Excel workbook

SeedSystem is a collaboration among diverse national and international organizations to improve seed security in vulnerable and high-stress areas of the world.

See seedsystem.org for more.

Questions? Please contact info@seedsystem.org

Contents

1. INTRODUCTION 3

2. OVERVIEW: WHAT THE AUTOMATED EXCEL WORKBOOK DOES 5

3. DATA ENTRY: A STEP-BY-STEP GUIDE 7

4. MANAGING THE DATASETS AND FILES - FIXING MISTAKES and ASSURING QUALITY 19

5. INTERPRETING OUTPUTS 33

6. HOW TO COPY AND PASTE TABLES FOR USE IN REPORTS 45

APPENDIX 1: Worksheets 47

For additional assistance, please email info@seedsystem.org with your query. An expert will reply as soon as possible.

1. INTRODUCTION

This guide will show you how to get results from individual farmer interviews in your SSSA, and get them **quickly**. In a typical SSSA, a field team will spend four to five days in one location, and usually carry out household interviews in the first two or three days. With a small team for data-entry 'in the field', you can see detailed findings (even in table form) from these interviews **immediately** after they are entered, while you are still in the field. This offers great advantages for your SSSA:

- The SSSA team can discuss the findings amongst themselves **while still in the field**
    - The interviews are still fresh in their memories, so the team can help make sense of the findings, also drawing from their own knowledge and work in the region.
- New questions may arise while looking at the findings, **which can be pursued**.
    - Teams looking at findings while still in the field can follow up questions then and there (e.g. modifying questions for any key informants or experts they have yet to interview, or even identifying additional farmers to interview – such as farmers in a particular location, or a particular group, like displaced people).
- Teams can reflect on interview findings (and other data) and start to develop an **Action Plan** while still in the field, and still all together.
- **Findings are ready for presentation** almost immediately.
    - Data tables are formatted and can be copied directly to reports or presentations, which means you can feed back immediately to stakeholders, and focus on developing action plans, project proposals, etc., rather than on formatting tables.

Speed is important, but also rigor. The automated Excel workbook is designed not only to help SSSA teams enter data quickly, and generate and results immediately, but also to help data cleaning to occur quickly, and to limit errors.

This guide explains how to use the automated Excel workbook for data entry and analysis. This guide is important for team leaders in a site as well as for those involved in data-entry and in report writing.

Section 2 gives an overview of how the automated data analysis works, and will be of interest to the entire SSSA team.

Section 3 is a step-by-step guide to data entry, with details on how to enter data, the key codes, how to add new crops, and other key steps in data entry. This section is mainly targeted at the team who enter the data, as well as those managing this team, as it focuses on details such as coding.

Section 4 discusses key steps for managing data-collection and data-entry, and for managing files once data are entered. It addresses key tasks around how to: gather interview forms and check them in the field, convert measures to kg (across crops), check for mistakes in data-entry, add in new crops, filter output displays to only show crops with data, and combine or split different files. This section is for those managing the data-collection and data entry teams, and who are in charge of maintaining the dataset and its quality. (This may be the SSSA team leader, or someone specially charged with managing the dataset).

Section 5 discusses the wide range of outputs from the automatic analysis, and how to interpret them in detail. This will be of particular interest to SSSA team members leading on the Action Plan and reports, but it is also important for the whole team to understand the findings so that they can offer their own insights.

Section 6 discusses how to cut and paste the output tables so they can be used effectively in reports and presentations.

This guide and the automated Excel workbook have been designed with SSSA field teams in mind. The aims are to: assist in entering farmer interview data quickly; limit errors in data entry; help identify and correct any remaining errors quickly; run a wide set of analyses, and help identify important trends from the data gathered.

These processes will work best with a solid management structure and with clearly-defined roles and responsibilities for distinct tasks:

- data entry
- overseeing the quality of data entry and checking for errors
- managing (combining, splitting) files,
- analysis, and
- report-writing.

One person may perform several of these functions. However, it would be a mistake to assume that someone employed for data-entry necessarily has the skills and experience to manage data files or interpret the results. The overall management and analysis of interview data is a major task and should be resourced properly within the SSSA team and partner organizations, with suitably senior personnel managing these processes in the field. For high quality results, discussed immediately among the whole team in a participatory manner, leaders need to be in the field rather than back at the office.

These tools are meant to be discussed among the entire team, while still in the field, for instance by visually **projecting tables from Excel** and facilitating a group discussion of key trends. These tools are meant for **speed,** and to help you be **selective** in which results to discuss with the SSSA team. There are many tables of outputs to look at - exploring all of them in detail is <u>not</u> helpful. It would exhaust any team after a long day in the field, and use up valuable research time. A key role of managers at this stage is to select which results to discuss with the team, and to use the time carefully to gather insights and share understanding of what those results might mean. The time for a more detailed analysis is later, after field work is done, when writing reports.

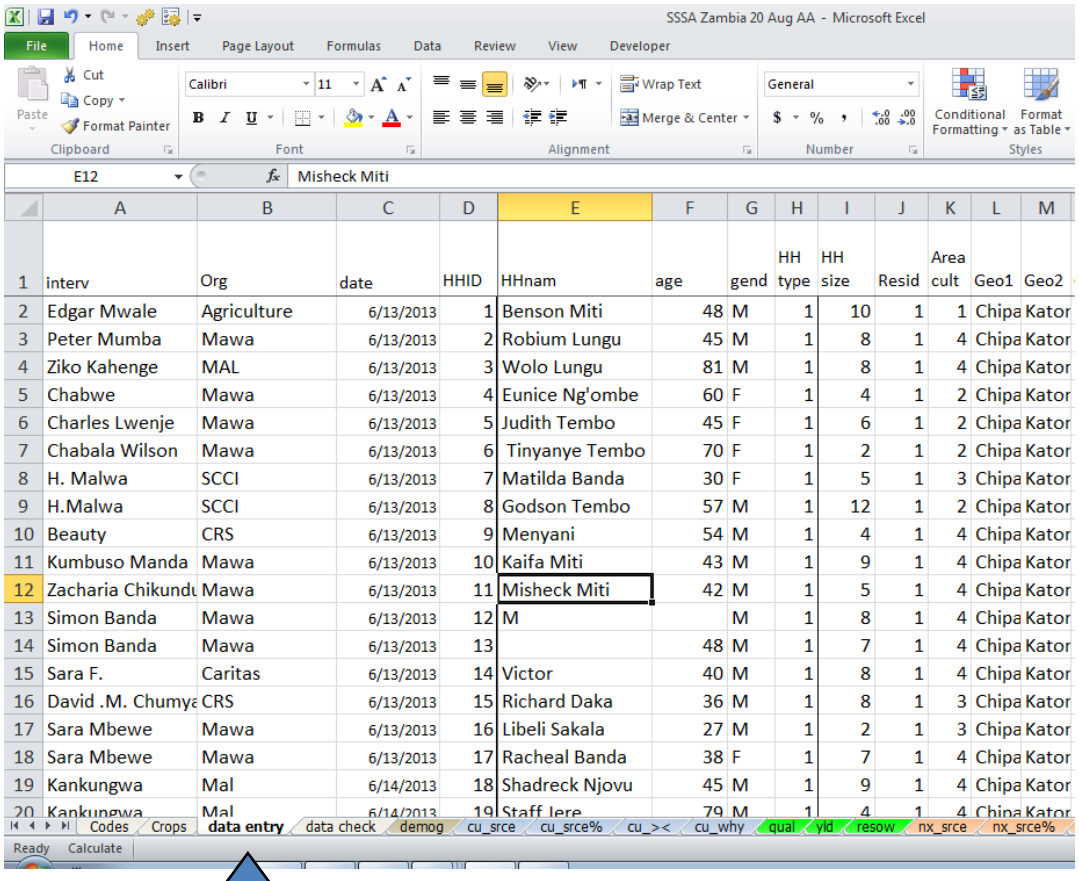
2. OVERVIEW: WHAT THE AUTOMATED EXCEL WORKBOOK DOES

The heart of the tool is a Microsoft Excel workbook file, which is designed to:

- Structure and facilitate entering the data from the interview forms, and make it as quick as possible
- Limit errors during data-entry, and help to identify and correct any mistakes
- Generate instant results for all the questions in the interview, including results disaggregated by gender or land size
- Format output for easy transfer to reports and presentations

The workbook contains **31 worksheets,** which are separate 'pages' in the workbook (these are listed in Appendix 1). Tabs at the bottom of the screen help to navigate among sheets, as shown below (Fig 2.1).

Fig. 2.1: Tabs for different worksheets



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	interv	Org	date	HHID	HHnam	age	gend	HH type	HH size	Resid	Area cult	Geo1	Geo2
2	Edgar Mwale	Agriculture	6/13/2013	1	Benson Miti	48	M	1	10	1	1	Chipa	Kator
3	Peter Mumba	Mawa	6/13/2013	2	Robium Lungu	45	M	1	8	1	4	Chipa	Kator
4	Ziko Kahenge	MAL	6/13/2013	3	Wolo Lungu	81	M	1	8	1	4	Chipa	Kator
5	Chabwe	Mawa	6/13/2013	4	Eunice Ng'ombe	60	F	1	4	1	2	Chipa	Kator
6	Charles Lwenje	Mawa	6/13/2013	5	Judith Tembo	45	F	1	6	1	2	Chipa	Kator
7	Chabala Wilson	Mawa	6/13/2013	6	Tinyanye Tembo	70	F	1	2	1	2	Chipa	Kator
8	H. Malwa	SCCI	6/13/2013	7	Matilda Banda	30	F	1	5	1	3	Chipa	Kator
9	H.Malwa	SCCI	6/13/2013	8	Godson Tembo	57	M	1	12	1	2	Chipa	Kator
10	Beauty	CRS	6/13/2013	9	Menyani	54	M	1	4	1	4	Chipa	Kator
11	Kumbuso Manda	Mawa	6/13/2013	10	Kaifa Miti	43	M	1	9	1	4	Chipa	Kator
12	Zacharia Chikundu	Mawa	6/13/2013	11	Misheck Miti	42	M	1	5	1	4	Chipa	Kator
13	Simon Banda	Mawa	6/13/2013	12	M		M	1	8	1	4	Chipa	Kator
14	Simon Banda	Mawa	6/13/2013	13		48	M	1	7	1	4	Chipa	Kator
15	Sara F.	Caritas	6/13/2013	14	Victor	40	M	1	8	1	4	Chipa	Kator
16	David .M. Chumya	CRS	6/13/2013	15	Richard Daka	36	M	1	8	1	3	Chipa	Kator
17	Sara Mbewe	Mawa	6/13/2013	16	Libeli Sakala	27	M	1	2	1	3	Chipa	Kator
18	Sara Mbewe	Mawa	6/13/2013	17	Racheal Banda	38	F	1	7	1	4	Chipa	Kator
19	Kankungwa	Mal	6/14/2013	18	Shadreck Njovu	45	M	1	9	1	4	Chipa	Kator
20	Kankungwa	Mal	6/14/2013	19	Staff Jere	79	M	1	4	1	4	Chipa	Kator

Tabs showing some of the worksheets in file. Currently 'data entry' is showing.

The results include: summary information about: the households interviewed; their seed sources; how (and why) they have changed the amounts they sow (compared with normal amounts); farmers' assessment of seed performance; their input use; how they access new varieties and seed aid; and how much money they spend on seed. All of these findings are broken down by crop, and shown for both the current season and for next season. Finally, the results include simple statistical analyses. Each worksheet groups a particular set of results together, to help you seek out particular findings.

Users make inputs into the worksheet in three main areas: a) inputting the data from the farmer interviews, b) entering seed price data for calculating the money spent on seed, and c) adding the names of new crops to the existing list (if required). These are discussed in turn.

a) **Data entry**. The key sheet for SSSA teams to work with is **'data entry'**, where they can enter in the information from interview forms. Section 3 explains data entry in detail.

Once data are entered into 'data entry', the other sheets (all those that come after 'data entry', to its right) will calculate a wide range of results. These read the data in 'data entry', and use formulas (which are hidden, to avoid being accidentally changed) to generate many different tables of outputs (one example is in Fig 2.2).

Fig 2.2. Example of output table generated automatically from data

1) Gender of HH Head

HH Head	N	%
Male	97	78.2%
Female	27	21.8%
total	124	100.0%

IMPORTANT NOTE: It takes about one minute for the sheets to run all the calculations. For this reason, automatic calculation is SWITCHED OFF; otherwise, there would be big delays every time you add new data to a cell in 'data entry'.

THIS MEANS THAT YOU MUST DIRECT EXCEL TO RUN THE CALCULATIONS BEFORE LOOKING AT THE RESULTS – AND AFTER ANY UPDATES TO THE FILES. This is simply done by pressing F9.

Once all data are entered, you can, if you wish, make calculations run automatically by going to 'calculation options' in the Formulas menu.

The results calculate from **all** the records in 'data entry' (there is space for up to 1,500 records). So if your SSSA has specific sites or regions, you will want to look at those results separately. Hence, each site should have its own file, containing only interview data from that site. All sites can also be combined to look at all data together, if you wish, in a separate file. Finally, you can sub-divide a site to look only at results for a specific group of households, such as Internally Displaced Person (IDPs). Again, this is best done by creating a new file containing only those records. Section 4 explains how to do this.

b) **Seed prices.** The Worksheet 'Money' also has space for you to enter prices for key crops for 'potential seed' sold in local markets, and for seed sold in agro-input shops. These figures are ideally gathered from your own field research, and can be entered for TWO seasons, to show the possible impact of price changes on what farmers spend on seed.

c) **Adding new crops.** The workbook already includes more than 85 crops. However, important crops in your site may be missed, so additional crop names can be added (there is space for up to 200 different crops), to the worksheet 'Crops'.

These actions are explained in detail in Section 4.

KEY LESSON FROM THIS SECTION:

1. **The automated Excel workbook contains 31 worksheets, designed to facilitate data entry, data cleaning, and data analysis.**

3. DATA ENTRY: A STEP-BY-STEP GUIDE

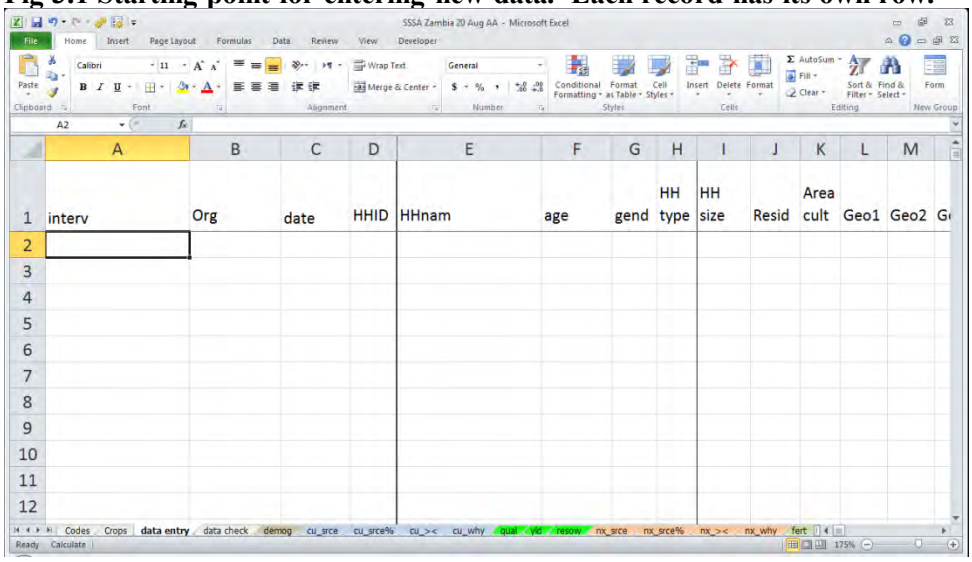
This section is mainly directed at the teams who will enter the data, as well as those managing data and checking for errors. It contains details about codes, and the use of the data-entry form, as well as tips on good practice for minimizing errors.

a) Data entry overview

This takes place in the worksheet 'data entry'. Each individual form is a record, representing one farmer interviewed, and has its own row. Each column represents a variable on the interview form – every possible answer has a column of its own, with 272 spaces to fill in data (see document 4.4).

Data are entered one record at a time (one record per household), starting a new row for each record, in column A (name of interviewer), and then moving across to the right, filling information in that appears on the paper interview form (many spaces may be left blank).

Fig 3.1 Starting point for entering new data. Each record has its own row.



The files should be saved regularly. You should use a standard structure for filenames, so that you can keep track of different updates. This is explained in Section 4, along with how to merge different files together (such as files from different people entering data for the same site).

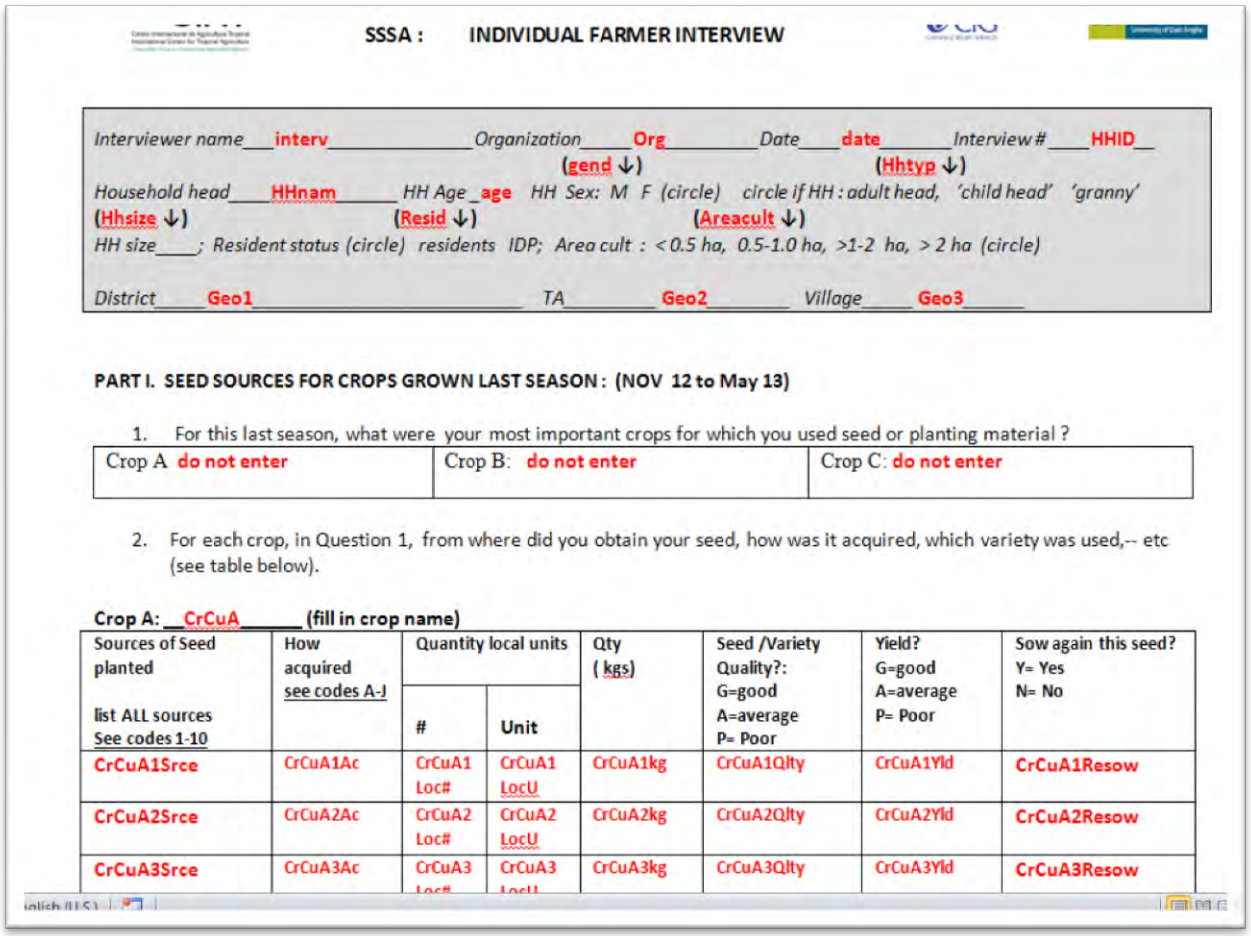
Data entry should ONLY take place in the worksheet called 'data entry'. The other worksheets should not be modified, with only two exceptions - adding new crops to the existing list of crops (see (e) below), and adding in prices of seed / potential seed. These other steps, however, may best be done by managers.

b) Names of variables on 'data entry' worksheet

Each space on the interview form has a unique variable name, which is used for data entry. There are 272 variables in total. In the Excel file, the worksheet '**codes'** (the first tab in the file) lists every variable, using variable names from the 'data entry' sheet listed across the top, in Row 1. Alongside the name is a short description of each variable, followed by CODES (where present). It is helpful to print out copies of the codes for the data-entry teams, so that they are absolutely clear about the correct field names for entering data, and the correct codes (explained in below, and in c).

For clarity, **Document 4.4 shows the variable names on the interview form, in red (**Fig. 3.2 shows how part of this document appears). Printing this out can also help data entry teams be certain they enter data into the correct columns as they consult the paper forms.

Fig 3.2 Document 4.4, showing variable names for data entry in red.



SSSA : INDIVIDUAL FARMER INTERVIEW

*Interviewer name*___interv_____ *Organization*____Org_____ *Date*____date_____ *Interview #*____HHID__
(gend ↓) (Hhtyp ↓)
*Household head*____HHnam_____ *HH Age* _age *HH Sex: M F (circle) circle if HH : adult head, 'child head' 'granny'*
(Hhsize ↓) (Resid ↓) (Areacult ↓)
*HH size*____; *Resident status (circle) residents IDP; Area cult : < 0.5 ha, 0.5-1.0 ha, >1-2 ha, > 2 ha (circle)*

*District*____Geo1_____ *TA*_____Geo2_____ *Village*____Geo3_____

PART I. SEED SOURCES FOR CROPS GROWN LAST SEASON : (NOV 12 to May 13)

1. For this last season, what were your most important crops for which you used seed or planting material ?

Crop A do not enter	Crop B: do not enter	Crop C: do not enter

2. For each crop, in Question 1, from where did you obtain your seed, how was it acquired, which variety was used,-- etc (see table below).

Crop A: __CrCuA_____ (fill in crop name)

Sources of Seed planted list ALL sources See codes 1-10	How acquired see codes A-J	Quantity local units		Qty (kgs)	Seed /Variety Quality?: G=good A=average P= Poor	Yield? G=good A=average P= Poor	Sow again this seed? Y= Yes N= No
		#	Unit				
CrCuA1Srce	CrCuA1Ac	CrCuA1 Loc#	CrCuA1 LocU	CrCuA1kg	CrCuA1Qlty	CrCuA1Yld	CrCuA1Resow
CrCuA2Srce	CrCuA2Ac	CrCuA2 Loc#	CrCuA2 LocU	CrCuA2kg	CrCuA2Qlty	CrCuA2Yld	CrCuA2Resow
CrCuA3Srce	CrCuA3Ac	CrCuA3 Loc#	CrCuA3 LocU	CrCuA3kg	CrCuA3Qlty	CrCuA3Yld	CrCuA3Resow

Going section-by-section in the interview form, the variable names are now discussed, followed by a discussion of the format for data-entry.

OPENING SECTION (Shaded box) - these variables identify when and where the interview happened, and key information to identify the household interviewed. There are some important **codes** in this section, which are described below.

PART I. SEED SOURCES FOR CROPS GROWN IN CURRENT / MOST RECENT SEASON.
This section has three entry tables on the interview form, one for each possible crop (A, B, C), so all variables in this section include **CrCu** (for CRop CUrrent season) followed by A, B, or C. Farmers can identify up to three different SOURCES for each crop, so these sources are identified as 1, 2 or 3. Each one is done in turn: So, after the name of crop A is entered (CrCuA), then all the 8 variables for **SOURCE 1** are entered (CrCuA1Srce to CrCuA1Resow), then followed by the 8 variables for **SOURCE** 2, and then the 8 variables for **SOURCE 3** (if there is data there – if the fields are blank, you skip them).

After filling in any data for these three sources, there are FIVE more variables: **CrCuAkgtot**, at the bottom of the table, then the four variables in the smaller table underneath the big one (**CrCuAActSow** to **CrCUAMLWhy).** Once this is done, you move on to the next crop – in this case, it will be **CrCuB.** On the data-entry worksheet, vertical lines separate groups of variables, to make it easier to tell where a new section starts on the paper interview form, or a new row in a table.

There are often many blanks, so a key challenge is to enter data in the right location. Use the vertical lines on the 'data entry' worksheet to help you find your place.

Fig. 3.3. Details of Table in Section 1, showing the variable names for data-entry.

Crop A: __CrCuA_____ (fill in crop name)

Sources of Seed planted list ALL sources See codes 1-10	How acquired see codes A-J	Quantity local units		Qty (kgs)	Seed /Variety Quality? G=good A=average P= Poor	Yield? G=good A=average P= Poor	Sow again this seed? Y= Yes N= No
		#	Unit				
CrCuA1Srce	CrCuA1Ac	CrCuA1 Loc#	CrCuA1 LocU	CrCuA1kg	CrCuA1Qlty	CrCuA1Yld	CrCuA1Resow
CrCuA2Srce	CrCuA2Ac	CrCuA2 Loc#	CrCuA2 LocU	CrCuA2kg	CrCuA2Qlty	CrCuA2Yld	CrCuA2Resow
CrCuA3Srce	CrCuA3Ac	CrCuA3 Loc#	CrCuA3 LocU	CrCuA3kg	CrCuA3Qlty	CrCuA3Yld	CrCuA3Resow
Total planted for Crop A				CrCuA kgTot			

Follow-up questions Crop A

Total quantity actually planted this last season (see above)	Normally, how much seed do you plant this season?	This season, did you plant M=more; S=same; or L=less than usual?	If different (M or L) explain (see separate code list: 1-34)
CrCuA ActSow	CrCuA NrSow	CrCuA MSL	CrCuA MLWhy

PART II: SEED SOURCES NEXT SEASON - This works in the same way, but the variables start with **CrNx (**for CRop NeXt season), and then A1, A2, A3, etc. This time each source has FIVE variables (**CrNxA1Srce** to **CrNxA1kg**). Again, there are FIVE variables at the end of each table.

PART III: INPUT USE – Questions 4 to 11 have similar formats: the top row is for Current season, so all variables start with **'Cu'**, the bottom row for the Next season, so start with '**Nx**'. As before, you work across one row and then move on to the next row. For these questions, there are THREE variable names for recoding up to THREE crops, and THREE for inputs (fertilizer, pesticides, compost, storage chemical). For question 12. You start with entering the answer for **'LossStoc'** (was there any storage loss – yes or no?), followed by **'LossCr1'** and **'Loss%1',** then on to **'LossCr2'** and **'Loss%2',** and **'LossCr3'** and **'Loss%3' (**if there are data there to enter).

PART IV: ACCESS TO NEW VARIETIES – Question 13 starts with two questions above the table: a Yes/No question ('**Newvar'**), and a total number for varieties ('**newvartot'**). Then, there is space enter up to THREE new varieties, with SIX variables per variety (from **'newvar1srce'**, to '**newvar1sow'**).

PARTS IV and V: ACCESS TO SEED AID – Question 14 has a similar structure, starting with a Yes/No question ('**aidYN?**') and a total number of times aid was received ('**aidtot**'), followed by questions about THREE different times a crop was given in aid (FOUR variables, from '**aid1org'**, to **'aid1yr')**.

c) Entering data – appropriate formats

Moving from one variable to another is easy – using a cursor, or the TAB button to move to the next available cell to the right. Spaces without information should be left BLANK in the spreadsheet.

Some variables have pre-set formats in the data entry spreadsheet, which only allow you to enter in the right format. For example, Yes/No questions are coded **1 for Yes, and 2 for No:** trying to enter something other than a 1 or a 2 will give you an error message (leaving it blank is fine if there is no answer to record).

Some other variables have a drop-down menu, indicated by a small down arrow when you get to the cell (Fig 3.4). Use the mouse to click this arrow, which will open a list of answers for that variable (Fig 3.5).

Use the mouse, or keyboard arrows, to select the right answer. **NOTE: for crops that appear again and again, you don't need to always use the drop-down menu to enter them – once you have entered that crop in the SAME COLUMN immediately above. If you start typing, the cell should complete the correct name** (e.g. in the example below in Fig 3.4, had you started typing 'so' the cell would have completed 'sorghum'. With typing, you will not be allowed to enter a name that is not among the choices available on the drop-down menu.

Fig 3.4 CrCuA, showing the arrow that indicates that this variable has a drop-down menu

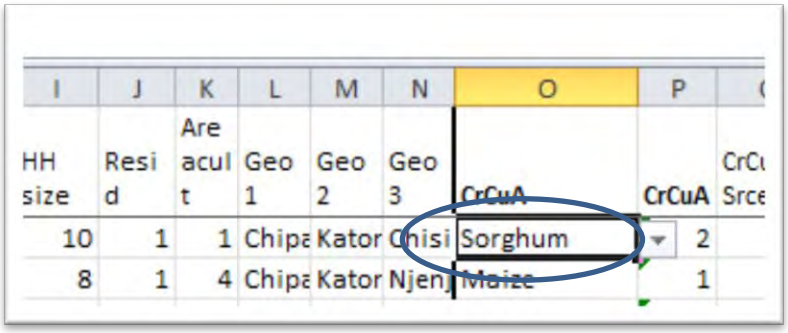
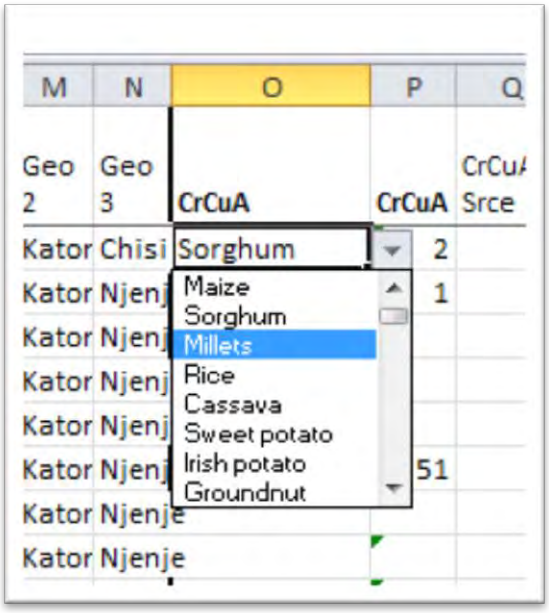


Fig 3.5. When you click on the arrow, a menu of answers opens; select the appropriate answer.



The Worksheet 'Codes' describes all variables, and gives notes on key codes and formatting. Key format notes are summarized below.

OPENING SECTION (Shaded Box)

Interv – text as written on form

Org – text as written

Date – as dd-mm-yy, example: 13-06-17

HHID – a single number. **Each interview form should have a UNIQUE number**. This number is most helpful for later checking against data from the original paper form. The number should be added on immediately as the form is collected. They do not need to be in strict sequence between sites – e.g. Site #1 may have forms 1-62; site #2 from 101-170; site #3 from 201-268. But the numbers must NOT be repeated between forms.

HHNam – name as written

Age – as written

Gend – M or F

HHtype – coded 1 – 3 (see below)

HHSize – as written

Resid – coded 1-2 (see below)

Areacult - coded 1-4 (see below)

Geo1, Geo2, Geo3 – these are normally entered as text, and geographical names chosen at an appropriate level to distinguish the different SSSA sites.

PART I. SEED SOURCES FOR CROPS GROWN IN CURRENT / MOST RECENT SEASON.

CrCuA – here, and everywhere else a crop name is needed, a drop-down list appears. The list contains 85 crops (as of October 2017; the full list can be seen on the worksheet 'crops'). You must move the slider **upwards** until you find the right crop.

CrCuA1Srce – enter the number as written. It should be between 1 and 10

CrCuA1Ac – enter the letter as written. It should be between A and K.

CrCuA1Loc# - enter number as written

CrCuA1LocU – enter text **as** written

CrCuA1kg – enter number as written. This number should be verified by the manager in charge of data analysis, as it may need to be changed. See (e) below, on converting to kg for Vegetative crops.

CrCuA1Qlty – coded 1-3 (see below)

CrCuA1Yld - coded 1-3 (see below)

CrCuA1Resow - coded 1-2 (see below)

Repeat for each row, if data are there

CrCuAkgtot – enter number as written

CrCuAActsow – enter number as written

CrCuANrsow – enter number as written

CrCuAMSL – drop down menu

CrCuAMLWhy – enter number as written (code entered by interviewer, which should be between 1 and 20 if Less, and 30-49 if More)

These formats apply for the rest of Parts I and II (the first three pages)

PART III: INPUT USE

A similar pattern for questions 4-11:

CuFerUse – Yes/No, so coded either **1 (Yes), or 2 (No)**

CuFerWhyN – drop down menu, listing letters A-J

CuFerCr1, CuFerCr2, CuFerCr3 – drop down menus for crops

CuFerNam1, CuFerNam2, CuFerNam3 – enter text as written

NxFerUse – Yes/No, so coded either **1 (Yes), or 2 (No)**

NxFerWhyN – drop down menu, listing letters A-J

NxFerCr1, NxFerCr2, NxFerCr3 – drop down menus for crops

NxFerNam1, NxFerNam2, NxFerNam3 – enter text as written

Question 12:

LossStoc - Yes/No, so coded either **1 (Yes), or 2 (No)**

LossCr1, LossCr2, LossCr3 – drop-down menu for crop name

Loss%1, Loss%2, Loss%3 – number, as written **(NOTE: only allows numbers between 0 and 100, as these are % loss figures.)**

PARTS IV and V: ACCESS TO NEW VARIETIES, AND SEED AID

Newvar1 - Yes/No, so coded either **1 (Yes), or 2 (No). NOTE**, if this is left blank, but other parts of the question are filled in, then **code it 1, as obviously the answer is 'Yes'.**

Newvar1tot – a number (integer, so decimals not allowed.)

Newvar1Srce – drop-down menu for source (numbers 3 – 10 allowed).

Newvar1Ac – drop-down menu for Access (B – J)

Newvar1Cr – drop-down menu for crop

Newvar1nam- enter text as written

Newvar1yr – enter year as written

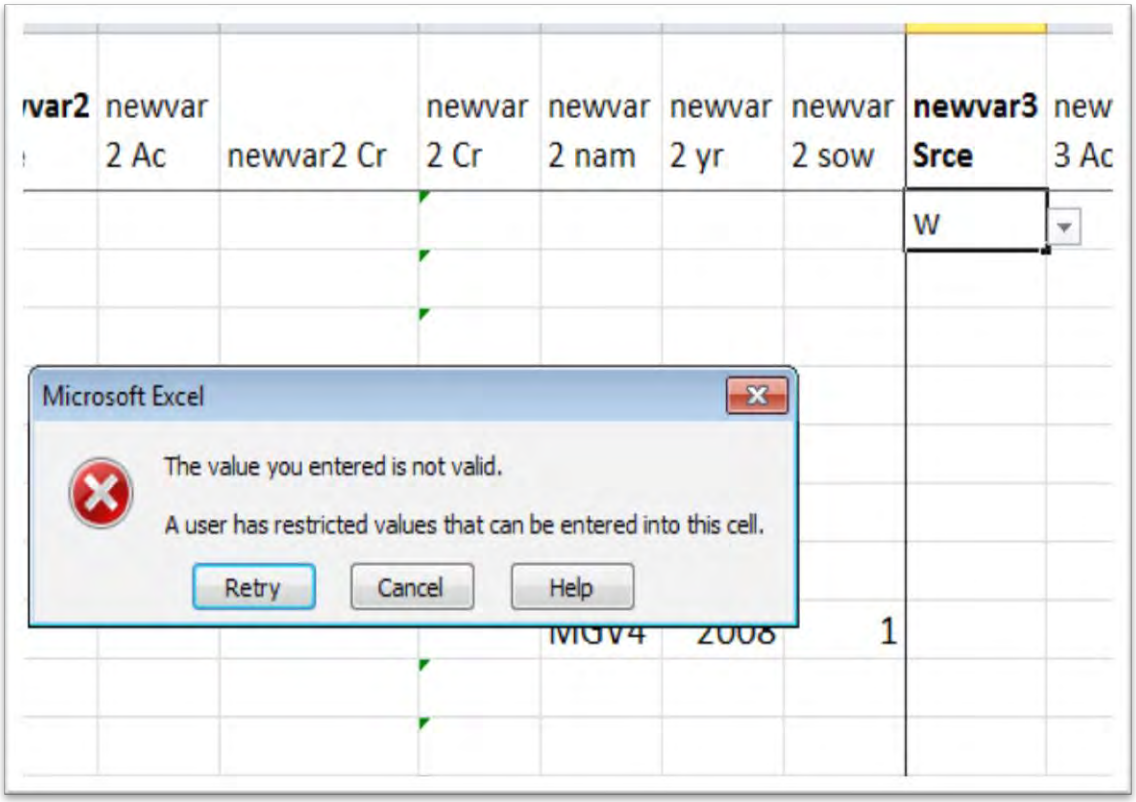
Newvar1sow - Yes/No, so coded either **1 (Yes), or 2 (No).**

Seed Aid follows a similar format.

If you get an error message when entering data (as in Fig 3.6), this means that your entry was not within the range of allowable answers. For instance, for a yes/no question, only 1 or 2 (or blanks) are accepted. Anything else will generate an error message like the one in Fig 3.6 below.

If the incorrect answer comes directly from the form, make a note to discuss with the data manager.

Fig. 3.6. Example of an error message for entering a code outside of the correct range of codes for that variable.



d) **Key codes –** these are the only codes that data-entry people need to look up. All other codes are either written in by interviewers (like seed source), or coded automatically (like crop).

Table 3.1 Key codes for data entry

Variable	description	CODES
Gend	Gender of the household head	**M=** male**, F=** female
Hhtype	Is household headed by adult, by a child, or by grandparent with young dependent children?	**1**= adult, **2** = child, **3** = "grandparent" (i.e. with dependent children)
Resid	Residence Status	**1**=resident, **2**=IDP
Areacult	Area cultivated	**1**:<0.5 ha, **2**: 0.5 - 1.0 ha, **3:** >1.0-2.0 ha, **4:** >2.0 ha
CrCuA1Qlty	quality of the seed obtained from the source	**1**= Good**, 2=**Average**, 3=** Poor
CrCuA1Yld	Yield of seed obtained from this source	**1**= Good**, 2=**Average**, 3=** Poor
CrCuA1Resow	Will the farmer sow the seed again?	**1**= YES, **2**=NO
CuFerUse NxFerUse CuPest Use, etc…. LossStoc Newvar aidYN?	ALL THESE ARE YES/ NO QUESTIONS, and coded the same way	**1**= YES, **2**=NO

NOTE: all variables ending in 'Qlty' or 'Yld' are coded 1, 2, 3 as above.

Table 3.2 CROPS already established in Excel workbook[1]

Crop	Code		Crop	Code		Crop	Code
Maize	1		Orange	43		Forage	85
Sorghum	2		Apple	44		**Next new crop**	**86**
Millets	3		Pawpaw/papaya	45			
Rice	4		Fruit trees	46			
Cassava	5		Yams	47			
Sweet potato	6		Peas	48			
Irish potato	7		Carrots	49			
Groundnut/peanut	8		Pois lyane	50			
Common beans	9		Rocket	51			
Cowpea	10		Chorol	52			
Pigeonpea	11		Pistasche	53			
Green Grams	12		Bok choi	54			
Banana	13		Cucumber	55			
Sesame	14		Hyacinth bean	56			
Chickpeas	15		Zucchini	57			
Bambara nuts	16		Arrowroot	58			
Pumpkin	17		Long Beans	59			
Soya bean	18		Mung beans	60			
Velvet Beans	19		Black beans	61			
Lima beans	20		Garlic	62			
Sunflower	21		Anise	63			
Tomatoes	22		Wheat	64			
Cabbage	23		Barley	65			
Onions	24		Squash	66			
Leek	25		Faba bean	67			
Okra	26		Canary melon	68			
Eggplant / aubergine	27		Cumin	69			
Pepper/piment	28		Coriander	70			
Amarenthus	29		Black seed	71			
Mustard	30		Lentil	72			
Gombo	31		Turnip	73			
Kale	32		Durum wheat	74			
Bitter leaves	33		Vetch	75			
Spinach	34		Sugarbeet/beet	76			
Green veg	35		Cauliflower	77			
Taro	36		Lettuce	78			
Cotton	37		Thyme	79			
Tobacco	38		Pepper (Spice)	80			
Coffee	39		Teff	81			

[1] The authors of this guide are aware that okra (26) is synonymous with gombo (31); likewise, green gram (12) is synonymous with mung beans (60). The authors will rectify this in the future.

Sugar cane	40		Shallot	82			
Watermelon	41		Field pea	83			
Pineapple	42		Grass pea	84			

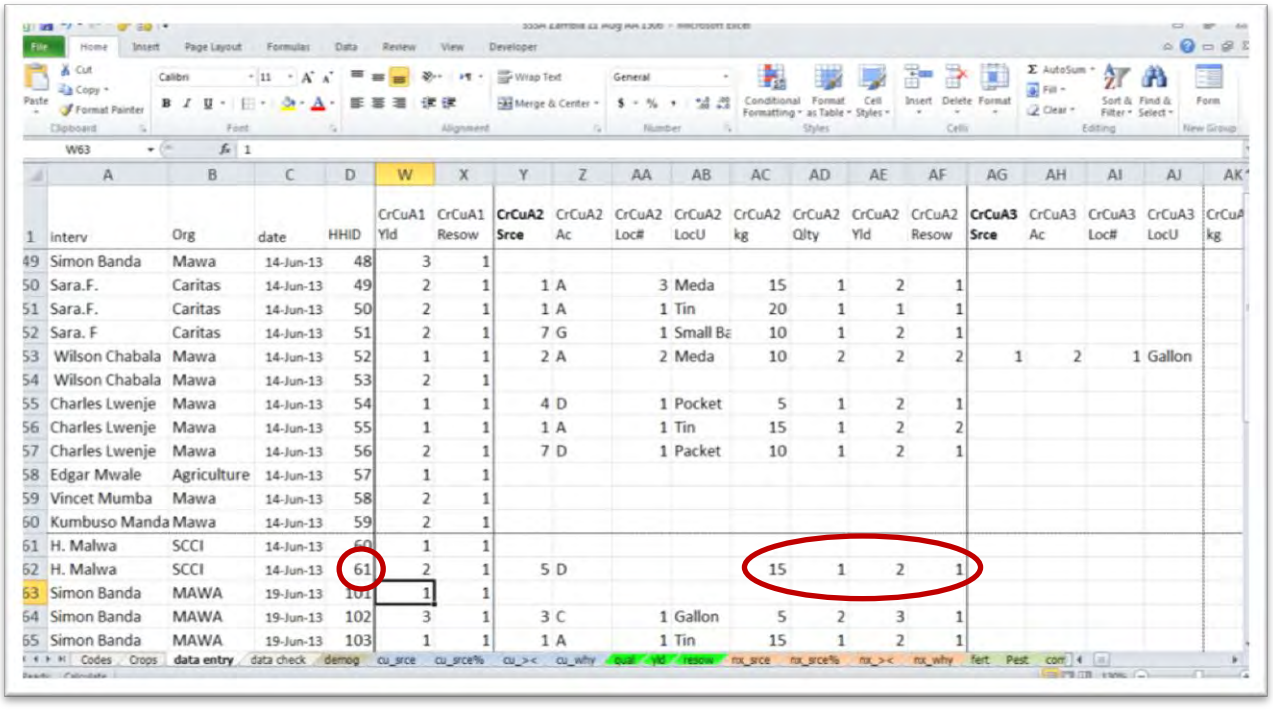
Use this list to check which crops already exist in the list, and where they are on the list in order to more easily find them on the drop-down menu. The most common crops are at the top. If a crop on the interview form is NOT in this list, it will need to be added to the list, with a new code of its own. Note down the new crop, and speak to your data manager for the new code to be added to the file (how to do this is discussed in Section 4).

e) Tips for data-entry and good file management

On the data entry worksheet, the TOP ROW and FIRST FOUR COLUMNS are '**FROZEN'.** This helps data-entry, as i) with **variable names** always in place, you always know which column is which, and ii) the **HHID** (column D) is always visible, helping you identify the record to which the data belongs. In the example below (Fig. 3.7), the circled data comes from record # 61. **HHID is easiest way to check data in the worksheet against what is written on the forms.**

If this needs to be re-set, Place the cursor on **cell E2**, then click on 'Freeze Panes' under the 'View' menu tab, and select 'Freeze Panes' option. The panes may first need to be unfrozen, before resetting the freeze to the right location.

Fig 3.7. Data entry sheet, showing how Row 1 and Columns A-D are frozen. That top row variable names, and first four columns (with record number – HHID – in Column D) will always be visible when you scroll down or to the right.



	A	B	C	D	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1	interv	Org	date	HHID	CrCuA1 Yld	CrCuA1 Resow	CrCuA2 Srce	CrCuA2 Ac	CrCuA2 Loc#	CrCuA2 LocU	CrCuA2 kg	CrCuA2 Qlty	CrCuA2 Yld	CrCuA2 Resow	CrCuA3 Srce	CrCuA3 Ac	CrCuA3 Loc#	CrCuA3 LocU
49	Simon Banda	Mawa	14-Jun-13	48	3	1												
50	Sara.F.	Caritas	14-Jun-13	49	2	1	1	A	3	Meda	15	1	2	1				
51	Sara.F.	Caritas	14-Jun-13	50	2	1	1	A	1	Tin	20	1	1	1				
52	Sara. F	Caritas	14-Jun-13	51	2	1	7	G	1	Small Ba	10	1	2	1				
53	Wilson Chabala	Mawa	14-Jun-13	52	1	1	2	A	2	Meda	10	2	2	2	1	2	1	Gallon
54	Wilson Chabala	Mawa	14-Jun-13	53	2	1												
55	Charles Lwenje	Mawa	14-Jun-13	54	1	1	4	D	1	Pocket	5	1	2	1				
56	Charles Lwenje	Mawa	14-Jun-13	55	1	1	1	A	1	Tin	15	1	2	2				
57	Charles Lwenje	Mawa	14-Jun-13	56	2	1	7	D	1	Packet	10	1	2	1				
58	Edgar Mwale	Agriculture	14-Jun-13	57	1	1												
59	Vincet Mumba	Mawa	14-Jun-13	58	2	1												
60	Kumbuso Manda	Mawa	14-Jun-13	59	2	1												
61	H. Malwa	SCCI	14-Jun-13	60	1	1												
62	H. Malwa	SCCI	14-Jun-13	61	2	1	5	D			15	1	2	1				
63	Simon Banda	MAWA	19-Jun-13	101	1	1												
64	Simon Banda	MAWA	19-Jun-13	102	3	1	3	C	1	Gallon	5	2	3	1				
65	Simon Banda	MAWA	19-Jun-13	103	1	1	1	A	1	Tin	15	1	2	1				

New crops. If a crop is not found in the existing list of crops, then it will need to be added to the list. It is important that EVERYONE, across all the SSSA sites, uses the same codes, so there needs to be some coordination when adding in new codes. This is why it is best left to a manager to do this. Steps for this are discussed below in Section 4. For data entry teams, they should make note of the new crop and discuss with a manager how to proceed.

If there are questions, keep a notebook and raise these with the manager. Normally, if the information is not provided on the interview form, then it should be left <u>blank</u>.

Saving the file
i) start from a blank **template** (the original, empty, file) when you first enter data.
ii) IMMEDIATELY save it with a new file name, so that the template remains unchanged.
iii) Save the file using a standard name format for the team, including: **the site**, **data-enterer's name, and date**. This makes it easy to know which files are for which SSSA sites, who entered the data, and which files are the newest ones.

e.g. for SSSA data in Chiro District, entered by 'Aberra' on Aug 22, 2014, the file name:

interview data_Chiro_Aberra_Aug 22 2014.xlsm

Once you have set a file name, you can re-save using 'CTRL-S' or via the menu, as normal. You should re-save often – after every record you enter!

KEY LESSONS FROM THIS SECTION:

1. **The "data entry" sheet of the Excel workbook allows the SSSA team to enter data. Variable/column names correspond to those on the survey form. The data entry team should take time to familiarize itself with the naming conventions for the variables.**
2. **Many of the variables are coded, as described in the tables above. The data entry team should become familiar with the codes as this will allow for faster data entry.**
3. **The data entry team should make use of the "freeze panes" option in Excel to follow a record easily.**
4. **SAVE YOUR WORK OFTEN!**

4. MANAGING THE DATASETS AND FILES - FIXING MISTAKES and ASSURING QUALITY

This section complements Section 3, but is meant for data managers who are responsible for managing the process of collecting, correcting, and entering the data. It discusses tips and key tasks for checking and correcting the paper forms in the field, for managing the spreadsheets and data-entry team, and key areas for managing the files and correctly arriving at outputs.

The steps in this section are essential and should be done first BEFORE looking at any results. For accurate results, you first need to correct mistakes, clean the data file, add in seed prices, and merge files to have a complete sample for each site. If you skip these steps, and look at results before cleaning the data, you may be wasting your time as some results may change once the data are cleaned.

a) Field review of questionnaires.

It is important to review the paper interview forms right away – ideally while still in the field, when interview details are still fresh in the memory. Annotate forms in a different color (such as red), so the data-entry team know which information has been checked (and should be entered).

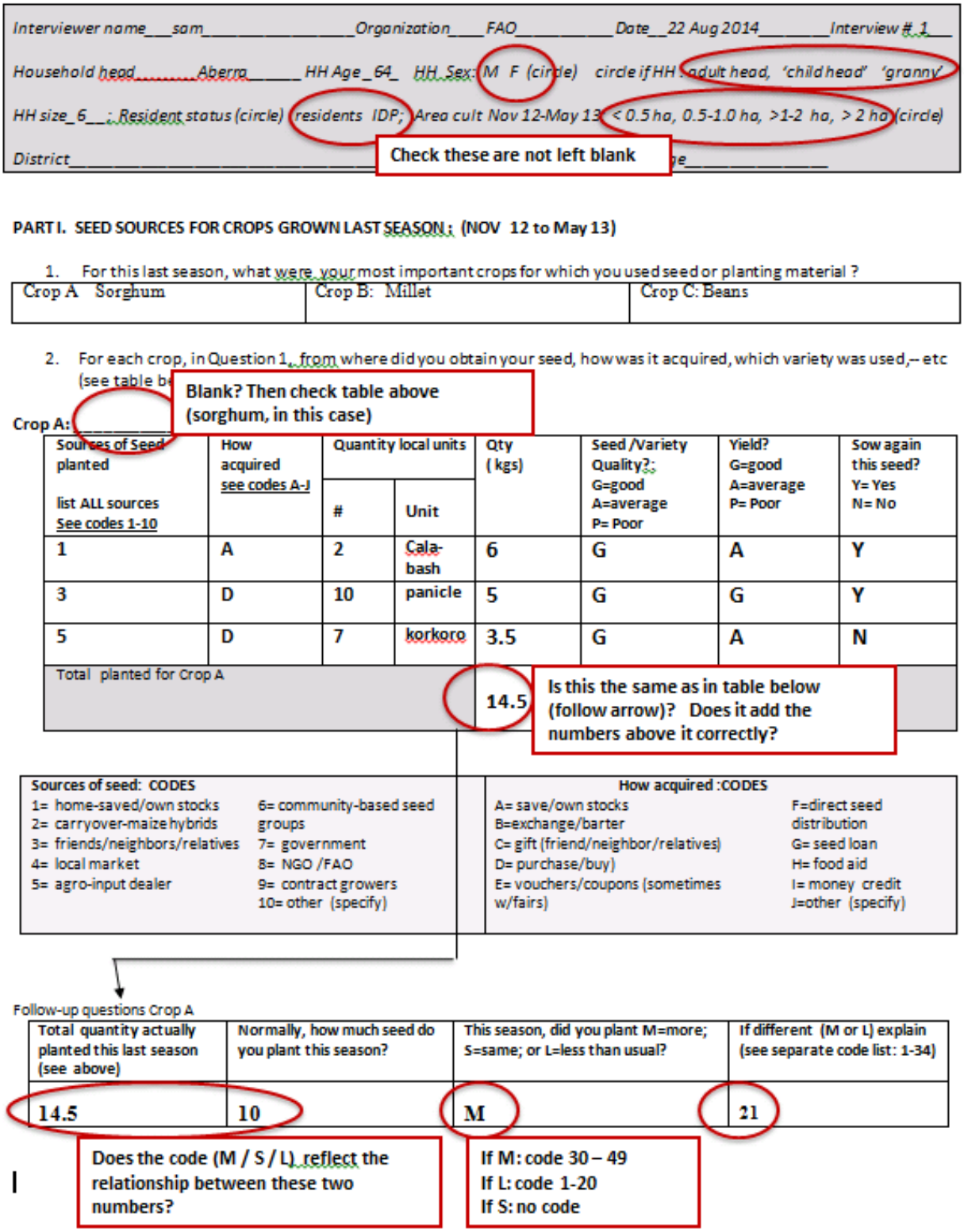
FIRST STEP: Enter a unique ID number for each interview form, in 'Interview #'. This will be HHID. They do not need to be in strict sequence between sites – e.g. Site #1 may have forms 1-62; site #2 from 101-170; site #3 from 201-268. But the numbers must NOT be repeated between forms.

In checking the forms, here are some common errors which should be checked:

- **Legibility** – if anything is hard to read or understand, this is the time to check it!
- **Missing information** –key information, like gender, farm size, or crop names, sometimes does not get filled in. This is the time to fill in these gaps, before the interviewer forgets.
- For **PART I and II Seed source** tables (ALL SIX):
    - Are the crop names filled in at the top of the table? If not, then please fill in from the table.
    - Is the **total kg** entered at the bottom of the table? Is it the same as in the smaller table underneath, for 'total quantity actually planted'?
    - In the smaller table underneath, does the code for 'More / Same / Less' reflect the relationship between the two numbers to its left?
    - Does the code in the fourth box make sense in relation to the 'More / Same / Less' code? - **LESS** → code 1-20; **MORE**→ code 30-49 **SAME** → no code
- For **Part III** – if the farmer said there was NO use of an input, is there a reason given for why not?
- **PARTS IV and V**. Check that YES or NO was circled appropriately. If the table below is BLANK, then it should be NO. If there ARE DATA, then YES should be circled.
- **PART IV** – each variety gets a separate line, even if it came the same time, or from the same source. The table has space for three varieties only, but the farmer may have received more new varieties in the past five years – total figure at the top is for total number of varieties, which can be greater than three.
- **PART V** – each CROP gets a separate line, even if it came at the same time, or from the same source. The table has space for three cases only. Total figure at top of table is number of times seed aid was received, which could be greater than three. Also, a farmer could get several crops from ONE seed aid: so the table could be full, with a total of '1'

- If the interviewer wrote notes, discuss them to make sure key insights and good stories are captured! Even brief notes may spark an interesting story, which should be noted!

Document 4.3 highlights some of the key areas to check when reviewing interview paper forms.



Interviewer name___sam_____Organization____FAO_____Date__22 Aug 2014_____Interview # 1___

Household head_____Aberra_____ HH Age_64_ HH Sex: M F (circle) circle if HH 'adult head, 'child head' 'granny'

HH size_6__; Resident status (circle) residents IDP; Area cult Nov 12-May 13 < 0.5 ha, 0.5-1.0 ha, >1-2 ha, > 2 ha (circle)

District_____

Check these are not left blank

PART I. SEED SOURCES FOR CROPS GROWN LAST SEASON : (NOV 12 to May 13)

1. For this last season, what were your most important crops for which you used seed or planting material ?

Crop A Sorghum	Crop B: Millet	Crop C: Beans

2. For each crop, in Question 1, from where did you obtain your seed, how was it acquired, which variety was used,-- etc (see table b

Blank? Then check table above (sorghum, in this case)

Crop A: _____

Sources of Seed planted list ALL sources See codes 1-10	How acquired see codes A-J	Quantity local units		Qty (kgs)	Seed /Variety Quality?: G=good A=average P= Poor	Yield? G=good A=average P= Poor	Sow again this seed? Y= Yes N= No
		#	Unit				
1	A	2	Cala-bash	6	G	A	Y
3	D	10	panicle	5	G	G	Y
5	D	7	korkoro	3.5	G	A	N
Total planted for Crop A				14.5			

Is this the same as in table below (follow arrow)? Does it add the numbers above it correctly?

Sources of seed: CODES		How acquired :CODES	
1= home-saved/own stocks	6= community-based seed groups	A= save/own stocks	F=direct seed distribution
2= carryover-maize hybrids	7= government	B=exchange/barter	G= seed loan
3= friends/neighbors/relatives	8= NGO /FAO	C= gift (friend/neighbor/relatives)	H= food aid
4= local market	9= contract growers	D= purchase/buy)	I= money credit
5= agro-input dealer	10= other (specify)	E= vouchers/coupons (sometimes w/fairs)	J=other (specify)

Follow-up questions Crop A

Total quantity actually planted this last season (see above)	Normally, how much seed do you plant this season?	This season, did you plant M=more; S=same; or L=less than usual?	If different (M or L) explain (see separate code list: 1-34)
14.5	10	M	21

Does the code (M / S / L) reflect the relationship between these two numbers?

If M: code 30 – 49
If L: code 1-20
If S: no code

b) Converting seed quantities to kg – local measures and vegetatively-propagated crops.

Many of the calculations in the spreadsheet refer to quantities of seed. To be able to compare between different farmers and different crops, these quantities need to be recorded using the same units - kilograms (kg). Normally, farmers do not describe seed using kg, but instead use their own units of measurement -containers such as tins, glasses, baskets, or a standard bundle (e.g. 20 cassava stems of 1m long). It is therefore important to convert local units to kilograms, and to convert them in a standard way for the entire SSSA.

Converting vegetatively-propagated crops such as cassava or sweet potato needs extra attention, to go from cuttings, vines, or suckers to kg. Conversion for seed crops (cereals, legumes, etc.,) will be discussed first, before turning to vegetatively-propagated crops.

i) Converting seed crops

An important part of the Launch Workshop for the SSSA, in preparation for field work, is discussing local units of measurement amongst the entire field team, and coming to some agreement about what these terms are and how they can be converted into kg. You need to discuss the most common units of measurement across a range of crops, thinking about the wide range of possible measures farmers use for their crops - especially measures used in local markets. There may need to be a different list of units for each site within a country. Once the list is agreed, each member of the interview team should have a copy, which will help them in filling in the interview forms and converting the units to kg.

Table 4.1 – Example of local measures in Kalemie, DRC, as developed in the Launch Workshop.

Local measure	Français	Swahili of Kalemie, DRC	Equivalent in Kg
Bags	Sacs	Saki / Gunia / Mufuko	50 & 100 kg in shops Maize, ground nuts, etc. - variable
Buckets	Seaux	Mbeketi / Meka	10 kg
		Kimbo	2.5 kg
		Kwaker (Quaker)	1 kg
		Ndochi	0.5 kg
Handful	Poignet	Mukono (chiwango ya Mukono)	0.075 kg
Cups	Gobelets	Kopo marboti	0.5 kg
Basin	Bassin	Beseni	30 kg
Basin	Bassin	Pompa	15 kg
Glass	Verres	Kilahuri (verré)	0.2 kg
Ear (maize)		**Fulin / Fulen**	0.05 kg

NB: groundnut quantities measured WITHOUT SHELLS

Table 4.1 gives an example of a table of measurements from Kalemie, DRC. Obviously, terms and sizes are very locally-specific. It is important to get the full range of terms – for example, note that there are two local terms for a 'basin', and several buckets of different sizes. The kg equivalent should be discussed – and agreed – as a team. It is important for everyone to use the same conversion rate (e.g. one glass = 0.2 kg, or five glasses = 1 kg). However, an 'absolutely precise' conversion rate (e.g. one glass = 0.19 kg, or 0.22 kg) is not necessary, and is not really the point. Instead, seek out conversion rates that are accurate, acceptable to the entire team and which are reasonably round numbers, to allow for easy conversion.

Converting to kg. When interviewing farmers, the local unit (and number of local units) should be recorded on the interview form so that conversions can be verified later, if needed. ONLY THEN should the kg be entered, with the conversion done by the person conducting the interview, and verified later by the data entry manager. So for example, a farmer in DRC talks about 10 glasses (*kilahuri*) of bean seed bought from the local market, 2 buckets (*mekas*) of maize saved on-farm, and

1 ½ basins (*pompas)* of maize seed borrowed from a neighbor – this would be (using Table 4.1): **2 kg** of beans (10 X 0.2 kg); **20 kg** of maize saved (2 X 10 kg); and **22.5 kg** maize borrowed (1.5 X 15 kg).

Horticultural crops may come in small packets –these should be recorded in terms of kg, which means small quantities are likely (e.g. one packet may be 25 grams, or **0.025 kg**).

ii) Converting vegetatively-propagated crops

Units for measuring planting material for vegetatively-propagated crops often refer to a bundle or other way of transporting planting material of these crops off the farm. To make it possible to compare across crops, these need to be converted into '**maize kg equivalents' (and, more specifically, to a standard base of maize kg equivalents)**

On the interview forms, and in the spreadsheet for data entry, instead of 'kg', the quantity of the planting material is entered as <u>number of cuttings/ suckers / vines</u> <u>which will be sowed as individual plants</u>. IN OTHER WORDS, FOR VEGETATIVE CROPS, RECORD SEED MEASUREMENTS AS NUMBER OF PROPAGULES.

Local units of measurement may not be for individual propagules, but rather for bundles or some other means of gathering and transporting a number of propagules at one time. These measures need their own conversions, which again must be discussed and agreed with the SSSA at the launch meeting. Table 4.2 gives an example of such conversion rates for local measures of vegetative crops, in this case from Timor-Leste.

Table 4.2 Example of conversion table for Vegetatively-propagated crops, from Timor-Leste

CROP	Measure (English)	Measure (Tetum)	Number of propagules
Cassava	Stem / Stake	**Lolon**	**5** Kain (cuttings)
	Bundle (cassava)	**Futun**	**100** Kain (cuttings)
	Cutting	**Kain**	Specify number actually planted
Sweet Potato (FEHUK)	Sweet potato vine (long)	**Talin**	**3** Kain (cuttings)
	Bundle	**Futun**	**100** Kain (Cuttings)
	Sack (25 kg)	**Saku**	**400** Kain (Cuttings)
Banana	Banana Sucker	**Hudi oan**	specify number of suckers
Irish Potatoes	usually sold in -saku (sacks) -Balde (buckets)	**Fehuk eropa isin**	 **300** **80**

With cassava, the propagule is a cutting, typically 20-25 cm long, but when farmers talk of quantities of cassava planting material, they will often refer to longer stakes or poles, or to a standard bundle of these stakes. In the table above, a typical stake was about 1m long, so equated to 5 cuttings, while a bundle was typically 20 stakes, or 100 cuttings (20 stakes x 5 cuttings per stake). Similar approaches are used for sweet potatoes – identifying terms of vines and bundles that are sold or exchanged, and identifying how many individual cuttings would be planted from these measures.

This approach to conversion can lead to large numbers of propagules recorded – often in the thousands - but then ten thousand cassava cuttings may be planted in one ha. To illustrate, a farmer may say she borrowed 60 Futun of cassava from a friend, and got another 10 Lolon from an NGO as a gift. In this example, the farmer would have 6,000 cuttings from her friend (60 bundles 'Futun' X 100 cuttings each) – and another 50 cuttings from the NGO (10 long stakes 'Lolon' X 5 cuttings each). These figures would be written down in the 'kg' part of the interview form, with 6,050 as the 'total kg' for these two sources.

Such large numbers are fine for vegetatively-propagated crops, as the spreadsheet converts these figures automatically to 'kg equivalents'. This equivalence is calculated in relation to the recommended sowing rate for maize, which is typically 25 kg per ha. In contrast, a typical 'recommended rate' for cassava is 10,000 cuttings per ha. Thus, in the spreadsheet converts 400 cuttings cassava to 1 kg. **This conversion is done automatically by the spreadsheet so you should enter the number of propagules, as described above. The spreadsheet will make the final conversion.** Thus, Table 4.3 below provides the conversion rates for information only.

Table 4.3 Kg equivalents, calculated automatically in the spreadsheet

CROP	Recommended sowing rate / ha	Number of cuttings for 1 'kg equivalent'
cassava	10 000	400
Sweet potato	12 500	500
Taro	12 500	500
Bananas	1 000	40
Irish Potato	25 000	1 000
Yams	10 000	400

If you add in a NEW VEGETATIVE CROP (i.e. one not on the existing list), you will have to make your own conversions, based on 'kg equivalents', as the spreadsheet will not have 'built-in' conversions already set in the tables.

In these cases, the number of propagules would need to be divided by the amount needed for '1 kg maize equivalent', similar to the examples in Table 4.3. These conversions are best done on the data-entry sheets. For advice on converting a new vegetative crop, do seek advice from info@seedsystem.org

c) Adding in a new crop

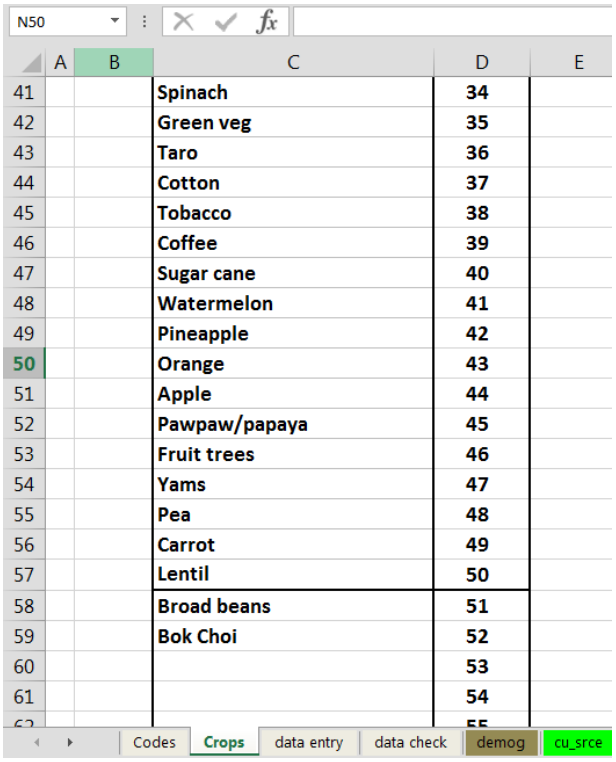
There are more than 85 crops already coded in the system, listed on Table 3.2. However, you may sometimes need to add additional crops from your SSSA, as farmers highlight crops that are important to them and their specific region. Adding a new crop is very straightforward, **but needs to be done by the entire team for data-entry, using the same codes for all added crops.** For this reason, the decision of what codes to give to each new crop should be made by a data manager and communicated to everyone involved in data entry. This is especially important if there is more than one team working at the same time. If there were not coordination across sites, then there would be complications in merging the datasets between sites, as they would be using different crop codes for the added numbers.

So, for instance, two teams working in different sites (A and B) at the same time: The team in Site A adds 'broad beans' (*Vicia faba*, also known as faba bean or field bean) while the team in Site B adds 'bok choi' (Chinese cabbage) as a key crop. If both call their new crop as code #86, then it will not be possible to combine them. **So it is essential that data-entry teams coordinate around new crop codes so that there is a common set of codes across the whole country.** So, in this example, broad beans may be 86, bok choi 87.

How to add in new crop codes: In the Excel spreadsheet, go to the 'Crops' worksheet (Fig 4.1). You will see that there are already crops listed alongside code numbers. These are protected, and cannot be changed. However, there is space below these crops, with code numbers going up to 200, giving you plenty of space to add in new crops if needed. Simply go to the first blank space, and type in the crop name as you want it to be displayed. This will then display on the drop down list on the 'data entry' sheet, added to the bottom of the existing list. **<u>This needs to be done on every computer used to enter data, and then the data file saved.</u>**

The data manager should keep track of crop changes across all sites as a 'master list'.

Fig. 4.1 Image from the 'Crops' worksheet, showing two new crops added alongside codes 51 and 52.



d) Naming files for data entry

It is best to save several copies of the blank Excel template file, and always use that to create a blank data-entry workbook work for each member of the data-entry team. This is to avoid any accidental changes to the display format.

Each person entering data should save files regularly, and use a standard filename system, so that you can keep track of: **the site**, **data-enterer's name, and date**. This makes it easy to know which files are for which SSSA sites, who entered the data, and which files are the newest ones.

e.g. for SSSA data in Chiro District, entered by 'Aberra' on 22 Aug 2014, the file name: interview data_Chiro_Aberra_22 Aug 2014.xls

e) Checking for errors in data-entry, and fixing them

Minor mistakes are always possible, due to errors in data entry, or mistakes on the interview form not identified earlier. Checking for mistakes is an essential step in managing data-entry teams, and ensuring that the dataset is free from mistakes ('clean'). A special worksheet within the Excel spreadsheet helps managers check for mistakes quickly, and correct them.

If there is more than one person entering data, then each individual's file should be checked separately, in the field, working alongside the person who entered the data.

Errors should be corrected BEFORE merging files across a site, and BEFORE looking at results.

Key steps:

- Start with the file containing the data entered by the team-member, and the paper forms, for consultation
- Save the spreadsheet that is being checked.
- Hit F9, to run the calculations. Do not touch the computer until this has completely run (progress is reported at the bottom right of the screen).
- Now go to the file, and select the worksheet '**data check'.** This has several sections: each will be discussed in turn.

i) **Crop kg totals – Columns C to Y**

Every crop table on the interview forms has space for entering up to three 'kg' figures (e.g. **CrCuA1kg, CrCuA2kg, CrCuA3kg,** for sources 1, 2 and 3 of crop A in the Current Season). The SUM of these three figures **should equal** the figure entered as the total kg for that crop (e.g. **CrCuAkgtot**). The figure for "amount actually sowed" (**CrCuAActsow**) should also be the same. This part of the sheet checks for this.

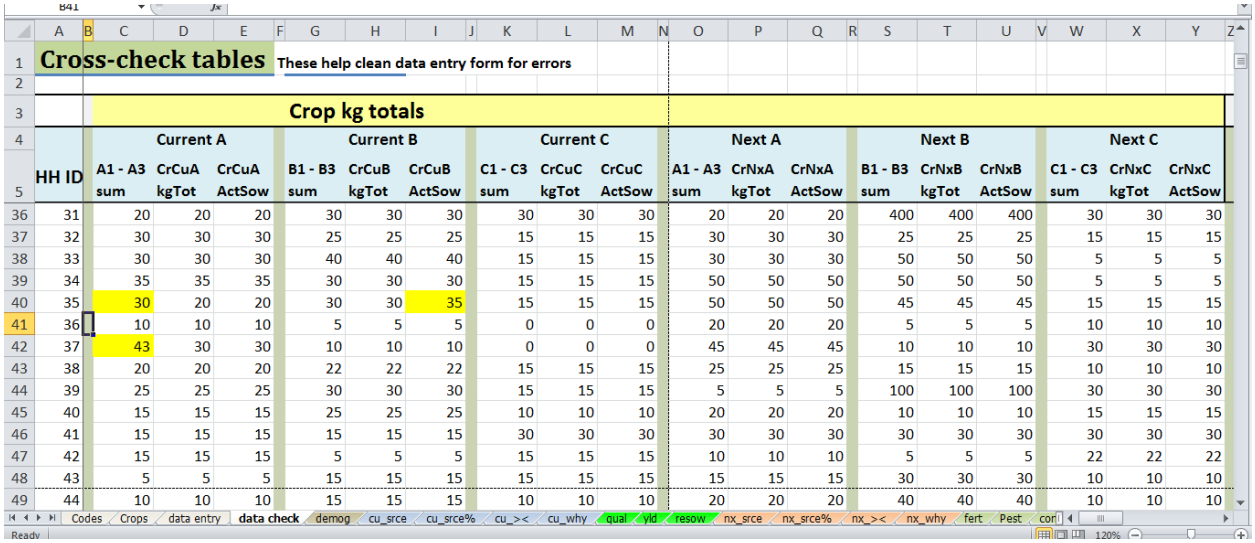
For each of the six crops (Current Crops A, B, C; Next Crops A, B, C), the section provides the key figures in three columns: the **sum** of the three individual kg figures; the figure entered as **'kg tot'**, and the figure entered as **'Actsow'** for that crop.

THE THREE FIGURES SHOULD ALL BE THE SAME. WHERE THERE IS A DIFFERENCE, IT WILL BE HIGHLIGHTED IN YELLOW. In Fig 4.2, below, there are errors for 'Current Crop A' for records 35 and 37. In both cases, the sum of A1-A3 was different from the total. These need to be checked against the interview forms. In this case, you should go to form HHid 35, look at the table for 'Current Crop A', and compare it against

what was entered in the data entry worksheet (**you check against HHid 35 in this case, NOT ROW 35, which would be for another record)**.

It may be that there was a simple typing error when entering the data. However, it may be that the mistake is on the form itself. If so, the data-entry manager must take a decision on how to resolve this mistake. All changes must be made in the **'data entry'** worksheet, and saved before proceeding to the next error.

Fig 4.2 Image from 'data check' worksheet, showing how errors for 'crop kg totals' are highlighted.



Cross-check tables — These help clean data entry form for errors

Crop kg totals

		Current A			Current B			Current C			Next A			Next B			Next C		
	HH ID	A1 - A3 sum	CrCuA kgTot	CrCuA ActSow	B1 - B3 sum	CrCuB kgTot	CrCuB ActSow	C1 - C3 sum	CrCuC kgTot	CrCuC ActSow	A1 - A3 sum	CrNxA kgTot	CrNxA ActSow	B1 - B3 sum	CrNxB kgTot	CrNxB ActSow	C1 - C3 sum	CrNxC kgTot	CrNxC ActSow
36	31	20	20	20	30	30	30	30	30	30	20	20	20	400	400	400	30	30	30
37	32	30	30	30	25	25	25	15	15	15	30	30	30	25	25	25	15	15	15
38	33	30	30	30	40	40	40	15	15	15	30	30	30	50	50	50	5	5	5
39	34	35	35	35	30	30	30	15	15	15	50	50	50	50	50	50	5	5	5
40	35	30	20	20	30	30	35	15	15	15	50	50	50	45	45	45	15	15	15
41	36	10	10	10	5	5	5	0	0	0	20	20	20	5	5	5	10	10	10
42	37	43	30	30	10	10	10	0	0	0	45	45	45	10	10	10	30	30	30
43	38	20	20	20	22	22	22	15	15	15	25	25	25	15	15	15	10	10	10
44	39	25	25	25	30	30	30	15	15	15	5	5	5	100	100	100	30	30	30
45	40	15	15	15	25	25	25	10	10	10	20	20	20	10	10	10	15	15	15
46	41	15	15	15	15	15	15	30	30	30	30	30	30	30	30	30	30	30	30
47	42	15	15	15	5	5	5	15	15	15	10	10	10	5	5	5	22	22	22
48	43	5	5	5	15	15	15	15	15	15	15	15	15	30	30	30	10	10	10
49	44	10	10	10	15	15	15	10	10	10	20	20	20	40	40	40	10	10	10

Checking for errors here is best done crop by crop. Starting at the top (Row 6), and scroll down looking at Current Crop A until you reach the bottom of the data (all columns will have zeroes only). Once this is done, go back to the top, and check Current B, then Current C, then Next A, etc.

Once you have completed the corrections, SAVE the file (CTRL-S). This should re-run the formula calculations. If the corrections are done properly, there should be no more highlighting in the 'Crop kg totals' section. If so, you can move onto the next section for error checking.

ii) More / Less / Same and codes (Columns AA to AQ)

This section looks at the answer 'was the amount you actually sowed (or are planning to sow) **More / Same/ or Less** than normal for that crop?', and the reason given for '**Why More or Less** for that crop?'. If planting more than normal, farmers should give a reason for planting more than normal (coded 30-49), and if less than normal, a reason for planting less (coded 1-20).

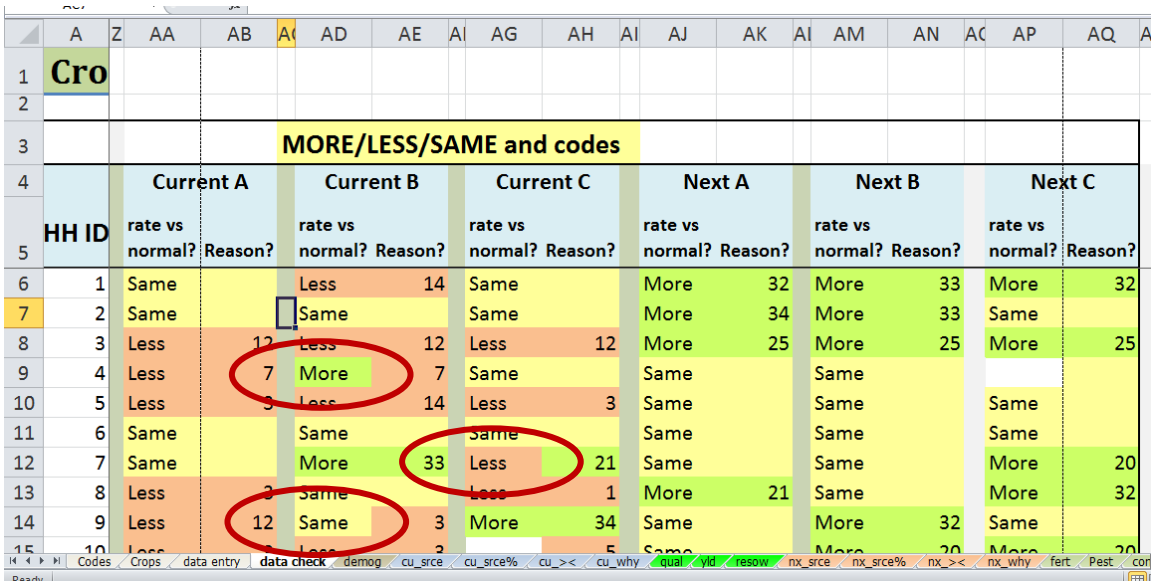
Here, each crop has two columns: the left column gives the answer to MORE / SAME/ LESS, with MORE shaded green, SAME light yellow, and LESS shaded red. The right column gives the code given for 'Why More or Less'.

Data-checking: look for cases where the two columns have **different colors.** Fig 4.3 gives some examples. In Fig 4.3, we can see that there are mistakes in records 4 and 9 for Current Crop B, and in record 7 for Current Crop C (circled).

Again, the mistake may be in data-entry (e.g. coding 'more' when it should have been something else), but it is also possible that the interview form recorded the wrong reason why for 'Why more or less'. If that is the case, speak to the person who did the interview to see if a different code may be appropriate. This sort of issue is best addressed in the field, soon after the interview form is handed in (see section 4 a, above.)

Again, once errors are corrected, the file should be SAVED, before proceeding.

Fig 4.3. Image from 'data check' worksheet, showing how errors in 'More/Less/Same' and Codes are identified.



MORE/LESS/SAME and codes

HH ID	Current A		Current B		Current C		Next A		Next B		Next C	
	rate vs normal?	Reason?	rate vs normal?	Reason?	rate vs normal?	Reason?	rate vs normal?	Reason?	rate vs normal?	Reason?	rate vs normal?	Reason?
1	Same		Less	14	Same		More	32	More	33	More	32
2	Same		Same		Same		More	34	More	33	Same	
3	Less	12	Less	12	Less	12	More	25	More	25	More	25
4	Less	7	More	7	Same		Same		Same			
5	Less	3	Less	14	Less	3	Same		Same		Same	
6	Same		Same		Same		Same		Same		Same	
7	Same		More	33	Less	21	Same		Same		More	20
8	Less	3	Same		Less	1	More	21	Same		More	32
9	Less	12	Same	3	More	34	Same		More	32	Same	
10	Less	3	Less	3		5	Same		More	20	More	20

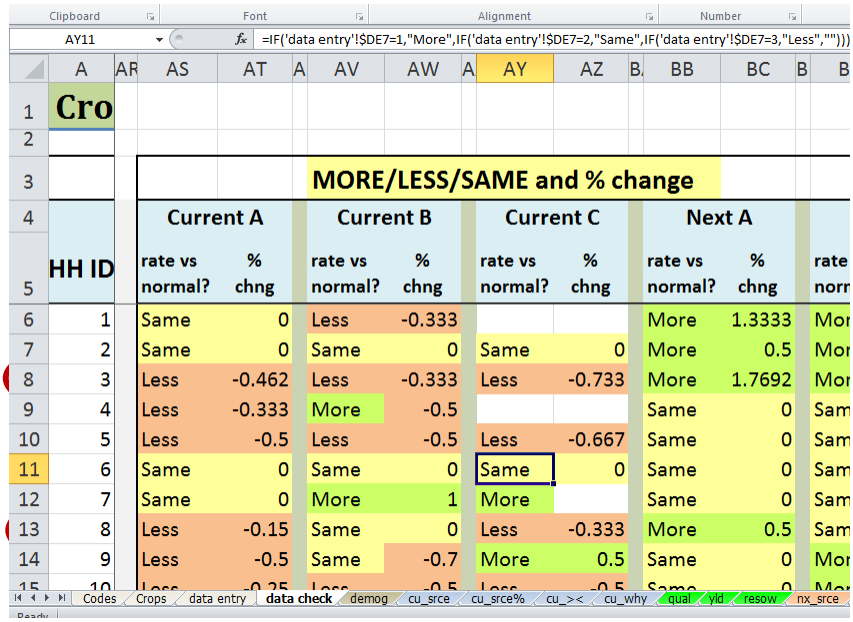
iii) **More / Less / Same and % change (Columns AS to BI)**

This section again looks at the **More / Same / Less** answer for each crop, but compares it to the difference between the figures for **Actual Sow** and **Normal Sow**. This difference is expressed as a % change, in terms of normal sow. So if the farmer is sowing 100% more than normal (i.e. 2 X normal), the number in the second column will be +1. If the farmer is sowing 50% less than normal, then the number will be -0.5.

Again, each crop has two columns: on the left, the answer to MORE / SAME / LESS, is given and shaded green, light yellow, or red depending on the answer. The right is **green for positive numbers** (the figure for Actual Sow is bigger than Normal Sow), light yellow for 0 (Actual Sow and Normal Sow are the same), and Red for negative figures (Actual Sow is smaller than Normal Sow).

Fig. 4.4 shows an example, with errors in Records 4 and 9 in Current Crop B. Again, these records on the data entry sheet need to be checked against the interview paper form. It may be that the wrong reason as entered, but it may also be possible that the mistake is with the number entered for Actual Sow. Again, once corrections are made, the file should be saved. Note that the record 7 for Current Crop C is actually not a mistake: in that case, the normal amount was 0 kg, so no % change figure is given as it is not possible to divide by 0. Checking this record shows that the 'More' code is indeed correct.

Fig. 4.4 Image from data check worksheet, showing detail for checking More / Less / Same and % change.



iv) **Checking for errors on Input use (Columns BK to CO)**

This section looks at PART III of the interview, where farmers answer YES or NO to whether they use a particular input (Fertilizer, Compost, Storage Chemicals, Pesticide). They either should give a reason why not (if they had answered NO) or name at least one crop if they had used that input (i.e. had answered Yes).

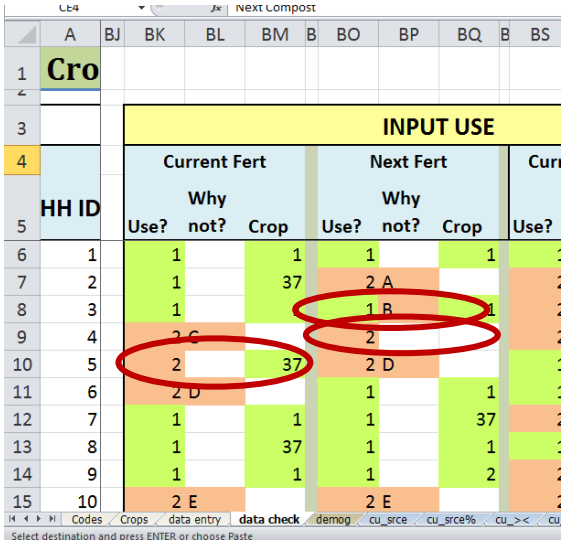
Each input is shown for Current Season, followed by the Next Season. Each input and season has three columns: the first column answers whether the input was used or not – **Yes is shaded green**, and **No is shaded red.** The middle column turns **red** if a reason was

entered for '**why not use this input?'.** It is blank if nothing is entered. The right column is **shaded green** if any crop has been entered here in relation to that input. Only 'Crop1' is checked for each input. For example, for Fertilizer use in the Current Season, the right column will check if something is entered in '**CuFerCr1'.**

CORRECT rows are either a) Green – blank cell – Green, or b) Red – Red – Blank cell. As before, if there are **two different colors** within one group of three cells in a row, there is an error. Fig. 4.5 shows an example of this, showing errors in record 5 for Fertilizer in the Current Season (answer coded as NO, but they are using it on a crop), and for Fertilizer Use Next Season there are errors in record 3 (answer coded as YES, but with a reason why not) and record 4 (answer coded as NO, but without a reason Why Not).

These questions should mainly be checked to ensure that the correct YES / NO code has been entered. It is possible that there may occasionally be missing data for 'crops' or for 'reason why not' – this is not ideal, but less of a problem.

Fig 4.5 Image from data check worksheet, showing error-checking columns for input use.



Once all errors have been checked and corrected, the dataset is ready to be combined with other datasets form the same site (if there are multiple files).

f) Merging or splitting files

Each spreadsheet will use ALL the rows in the data entry worksheet to calculate results. Therefore, you may wish to merge files, because:

- More than one person entered data for a site; and you want to combine them all (**once you have cleaned them of errors**) to get a complete dataset for the site.
- You want to combine all sites, to get national figures.

Equally, you may want to split a dataset - for example, to have a file that only has IDPs (Internally displaced people).

Merging and splitting these files is not difficult, but needs to be done carefully to avoid making mistakes.

i) Merging two files

- Open both Excel files in separate windows.
- **FOR the SOURCE DATASET:** Set your cursor in Cell A2, hold SHIFT, and SCROLL RIGHT to COLUMN JL. This should be the last column of data ('**aid3 yr'**).
- Still holding down SHIFT, SCROLL DOWN until you get to the last row with data. The last row should be apparent from the first four columns on the left (interv, org, date, and HHID).
- Once you are sure you have selected the FULL set of data (Columns A to LB, rows 2 to bottom of dataset), **Copy** them to the clipboard (using CTRL-C, or the menu).

- **IN the DESTINATION DATASET**: Go to Column A.
- SCROLL DOWN until you reach the first empty row (i.e. the first row below existing data).
- With your cursor in Column A, and the first blank row, **Paste data from the clipboard** (using CTRL-V, or the menu to do this).
- The new data should appear in the rows below where you started. You should quickly check that the data were pasted into the correct columns, e.g.
    - Does the date column have valid dates in all rows – from top to bottom?
    - Is column G (gender) reading M or F in all rows?
    - Does column O contain crop names in all rows?
    - Is there no data pasted after column LB?
- **If everything seems correct, then SAVE the file under a new name. Saving will re-run calculations, so the data analysis will now reflect all the records in the new, combined file.**

ii) Sorting datasets

It is often helpful to organize the dataset to display the records in order of their HHID number. This makes it easier to check back against records. Follow these steps for sorting the datasets:

- GO to the worksheet '**data entry'**
- Go to the '**Data'** menu, and click the '**Sort'** button
- All the data will now be highlighted in blue, and a dialog box will open.
- Tick the box 'my data has headers'
- Using the drop-down box on the left, select '**HHID'** as the column to sort the data by
- Select "smallest to largest"
- Click OK

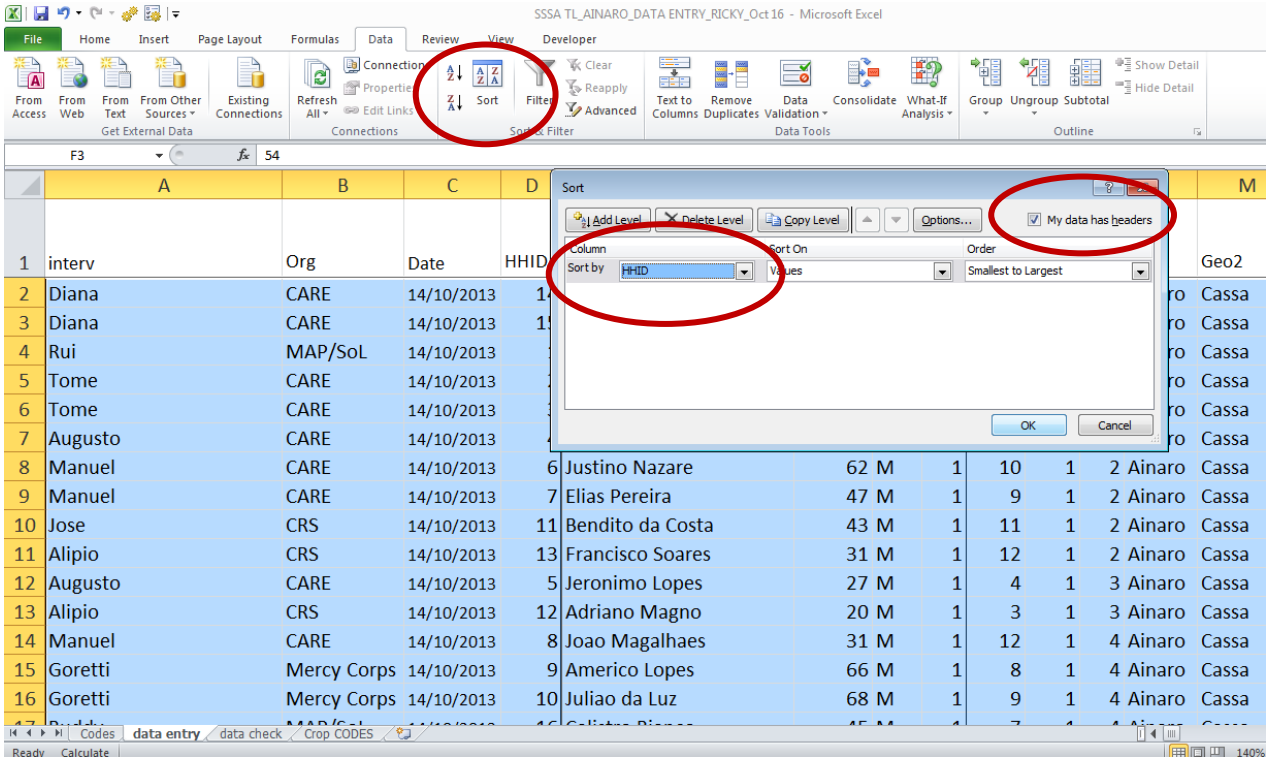
Fig. 4.6 shows the dialog box, and relevant fields to select.

Of course, it is possible to sort the data by other variables, such as (for example):

- whether the respondents are **residents** or **Internally-displaced people** (variable **resid**)
- a particular location (using **Geo1, Geo2,** or **Geo3** to sort), or
- farm size.

Sorting in this way can allow you to SPLIT a file, and only look at results for a specific group. This is described below.

Fig 4.6. The sort dialog box, showing key steps – ticking the 'my data has headers' box, and choosing 'HHID' as the column for sorting data.

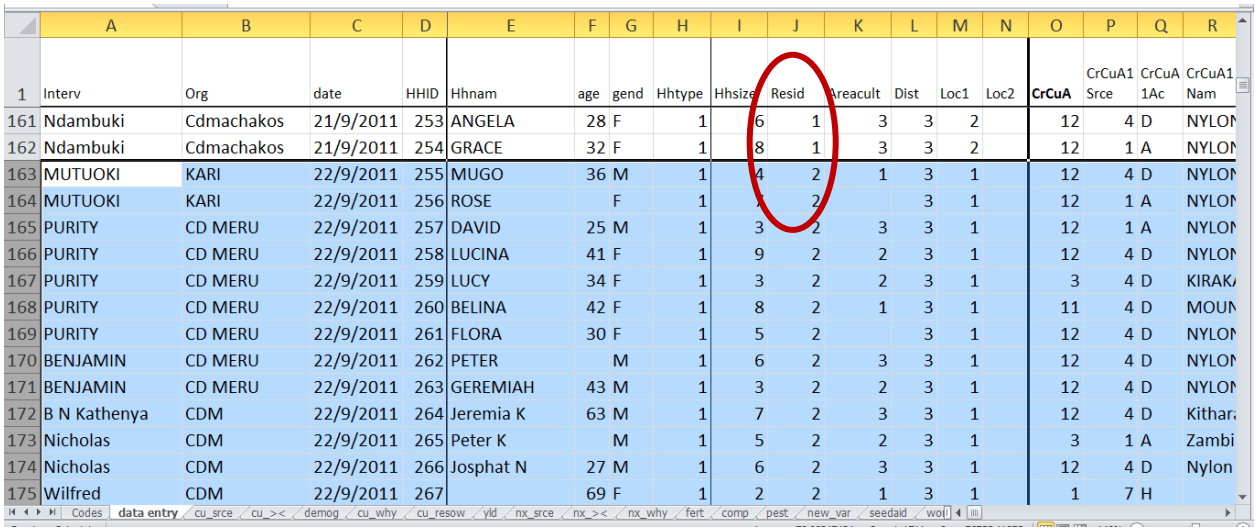


iii) Splitting datasets

It is quite simple to make a smaller dataset from a larger file. Once you have sorted the data in the way you want (see 4(f)ii above), and have saved the file, you can remove the unwanted records by deleting the unwanted rows. For example, imagine your dataset contains both farmers who are permanent residents and those who are IDPs (Internally-Displaced people). You want to look at the analysis results for residents and IDPs separately. TO do so, follow these steps (see also Fig 4.7):

- Sort the full dataset by '**resid', and SAVE this file**
- Once this is sorted, go to the Resid column and scroll down until you see the point where Resid changes from **1 (resident)** to **2 (IDP)**
- **To create a file containing only records for residents ('resid = 1'),**
  - Select ALL the rows where **Resid = 2,** by clicking on the numbers at the left.
  - DELETE THESE ROWS
  - Immediately **SAVE** this as a new file name (e.g. with 'residents only' in the file name)
- **To create a file containing only records for IDPs**
  - **OPEN** the original file which was sorted, and has both Resid = 1 and Resid =2
  - Select all rows where **Resid = 1**, and delete these
  - Immediately **SAVE** this as a new file (e.g. including 'IDPs only' in the file name)

Fig. 4.7 Example of using the 'sort' feature to organize the data, and to delete only certain records. In this example, the data have been sorted by 'Resid', and all the records with Resid=2 (row 163 and below) are about to be removed



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Interv	Org	date	HHID	Hhnam	age	gend	Hhtype	Hhsize	Resid	Areacult	Dist	Loc1	Loc2	CrCuA	CrCuA1 Srce	CrCuA 1Ac	CrCuA1 Nam
161	Ndambuki	Cdmachakos	21/9/2011	253	ANGELA	28	F	1	6	1	3	3	2		12	4	D	NYLON
162	Ndambuki	Cdmachakos	21/9/2011	254	GRACE	32	F	1	8	1	3	3	2		12	1	A	NYLON
163	MUTUOKI	KARI	22/9/2011	255	MUGO	36	M	1	4	2	1	3	1		12	4	D	NYLON
164	MUTUOKI	KARI	22/9/2011	256	ROSE		F	1	2	2		3	1		12	1	A	NYLON
165	PURITY	CD MERU	22/9/2011	257	DAVID	25	M	1	3	2	3	3	1		12	1	A	NYLON
166	PURITY	CD MERU	22/9/2011	258	LUCINA	41	F	1	9	2	2	3	1		12	4	D	NYLON
167	PURITY	CD MERU	22/9/2011	259	LUCY	34	F	1	3	2	2	3	1		3	4	D	KIRAKA
168	PURITY	CD MERU	22/9/2011	260	BELINA	42	F	1	8	2	1	3	1		11	4	D	MOUN
169	PURITY	CD MERU	22/9/2011	261	FLORA	30	F	1	5	2		3	1		12	4	D	NYLON
170	BENJAMIN	CD MERU	22/9/2011	262	PETER		M	1	6	2	3	3	1		12	4	D	NYLON
171	BENJAMIN	CD MERU	22/9/2011	263	GEREMIAH	43	M	1	3	2	2	3	1		12	4	D	NYLON
172	B N Kathenya	CDM	22/9/2011	264	Jeremia K	63	M	1	7	2	3	3	1		12	4	D	Kithara
173	Nicholas	CDM	22/9/2011	265	Peter K		M	1	5	2	2	3	1		3	1	A	Zambi
174	Nicholas	CDM	22/9/2011	266	Josphat N	27	M	1	6	2	3	3	1		12	4	D	Nylon
175	Wilfred	CDM	22/9/2011	267		69	F	1	2	2	1	3	1		1	7	H	

Codes | data entry | cu_srce | cu_>< | demog | cu_why | cu_resow | yld | nx_srce | nx_>< | nx_why | fert | comp | pest | new_var | seedaid | wo(

KEY LESSONS FROM THIS SECTION:

1. **It's very important to review survey forms in the field, as soon as possible after data collection. It's also very important to enter data into the Excel workbook as quickly as possible and conduct data checks, in order to clean any errors.**
2. **The Excel workbook automatically converts vegetatively-propagated crops (cassava, sweet potato, etc.) into kilogram of maize equivalents. Please contact info@seedsystem.org for any questions about this process.**
3. **When adding a new crop to the database, it's important to communicate the new addition among the entire SSSA team as soon as possible.**
4. **Carefully merge files from different data entry staff members.**

5. INTERPRETING OUTPUTS

a) Preliminary steps – running analysis and displaying only crops with data

If you have done any editing changes to the data, you need to do two steps before viewing the results.

i) **Running calculations.** First, when data are ready, press **F9, and wait.** In the lower right of the screen, a "**calculating**" will appear, with a % counting – slowly! – up to 100%.

ii) **Displaying only crops with data.** The crops tables in the spreadsheets are set up to analyze up to 200 possible crops. Of course, in any one location, only a few crops will have data, and the rest of the rows will be blank. There is a **FILTER MACRO** embedded in the spreadsheet file, which will only display crops with data in them.

> To run the macro, press **CTRL-q**. This will check ALL the tables listing crops, and display only those with data in them. When done, you will get a message "**Filter macro complete".**

This macro needs to be run **whenever new data are added**: otherwise, any new crop added since the last time a FILTER MACRO had run will not be displayed.

> There is also an **UNFILTER MACRO**, which removes any sorting and displays all crops again. The shortcut for this is **CTRL-Shift-K.**

b) Interpreting data analysis

The findings are grouped into a range of different sets of worksheets, grouped by color. These are all listed in Appendix 1, but summarized here:

- 5 sheets around data entry (+ codes and checking), in light gray.
- 1 sheet on demographics in darker gray: '**demog'** gives overall descriptive figures on the sample
- 9 sheets on seed sources and performance for CURRENT (or most recent) season, in bright green.
- 6 sheets on seed sources and performance for NEXT season, in light orange.
- 5 sheets in input use, in light purple
- 2 sheets on new varieties and seed aid, in light green:
- 1 sheet, on **money spent on seed**, in bright yellow
- 2 sheets on **statistical analysis**, in light blue

These are discussed below.

i) Data entry sheets

These sheets have been discussed in sections 3 and 4: **'Codes'** lists every variable; **'Crops'** is used to add new crops; **'MSL codes'** lists all reasons why farmers may plant more or less in the current year versus normal; **'data entry'** and '**data check'** for entering and verifying the data, respectively.

ii) Demographics

This sheet has 7 tables which summarize the characteristics of those in the sample (remember that the 'sample' is what is in 'data entry': as you merge files, the sample will get bigger). There are separate tables for gender of household head, nature of household head (adult, child, "grandparent") resident/IDP, and area under cultivation. Both total count (N) and % of the sample are given. Tables 5 and 6 give means and standard deviation for household size and age of household head, as well as the maximum and minimum figures in the sample.

Finally, Table 7 lists ranks the crops given most often as key crops for Current Season, and for Next season. Occasionally there is a small glitch, especially with small samples: if two or more crops are grown by the same number of farmers, the name of the first crop will be repeated. In this case, refer to the sheets **'Cu_<>'** and '**Nx_<>**' to get the correct crop names – Fig. 5.1 gives an example here.

Fig. 5.1 Demographic table 7; two crops are grown by 14 farmers, so name of 1st crop (sweet potato) is repeated. Cu_<> sheet confirms that this second crop is Taro, as it is also grown by 14 farmers.

7) Key crops - crops named most frequently as 'most important' by households

	Recent (current) season			Next Season		
	Crop	N of HHs	% of HHs	Crop	N of HHs	% of HHs
	Maize	54	90.0%	Maize	55	91.7%
	Common beans	50	83.3%	Common beans	38	63.3%
	Cassava	33	55.0%	Cassava	29	48.3%
	Sweet potato	14	23.3%	Sweet potato	19	31.7%
	Sweet potato	14	23.3%	Taro	10	16.7%

Amounts for current/most recent season: more, less, or same?

Crop	Number of HHs	% of HHs			Change sowing quantites for all growing the crop
		MORE	SAME	LESS	average % change
Maize	54	13.0	59.3	27.8	-5.07
Common beans	50	16.0	52.0	32.0	8.87
Cassava	33	6.1	63.6	30.3	-4.05
Sweet potato	14	0.0	78.6	21.4	-9.52
Taro	14	7.1	85.7	7.1	4.46
Irish Potato	5	20.0	40.0	40.0	-28.33

iii) Seed use in Current (or most recent) Season

The important sheets include: **'cu_srce'** and **'cu_srce%'**, which calculate amounts of seed for each source by crop; : **'cu_acq'** and **'cu_acq%'**, which calculate amounts of seed based on how they were acquired by crop; **'cu_><'** which looks at whether the amount of seed is more or less than normal; **'cu_why'**, which focuses on the reasons farmers give for sowing more or less; and **'qual'**, **'yield'**, and **'resow'** which summarize farmers' views of seed quality and yield, and whether they plan to resow.

Cu_Srce - this sheet calculates, for every crop in the Current season, the total amount of seed planted, and then breaks this down by the amount supplied from each source. These amounts add up all the farmers who mentioned that crop and source together. Amounts are in **kg** for seed crops, and '**kg of maize equivalents'** for vegetatively-propagated crops, as explained above, in section 4 (b) ii. So, for cassava, 400 cuttings are converted into 1 kg, or 10,000 cuttings to 25 kg. At the bottom of the table, the "TOTAL – all crops" row adds the entire column above it. These totals are why vegetatively-propagated crops are converted to maize equivalents – without converting, even one ha of cassava (10 000 cuttings) would dominate the total results. **Cu_Srce** gives the raw figures, but is not easy to interpret quickly – the following sheet is much better for this.

Cu_Srce% - This sheet gives the % of seed supplied by each seed source - for every crop, as well as for the sample as a whole. It calculates the total kg sowed for each crop, and uses figures from **Cu_Srce** to calculate the % of seed supplied from each source. Each row should add up to 100% - if they do not, that means that the data still need some cleaning. The bottom row gives the percentages for all crops combined. To signal important trends, any percentage above 30% is highlighted in orange (Fig 5.2). These figures are for the whole sample – e.g. Fig 6.2 shows that 15.2% of **all maize in the sample** came from local markets; one individual farmer, however, might get all of her maize from a local market, another farmer none at all.

Fig 5.2 Cu_Srce%. Farmers in this sample planted 2,395.6 kg of maize, with 69.5% coming from their own stocks, and 15.2% from local markets.

2) Proportion (%) of seed supplied by source for each crop.

Crop	Total kg sowed	% of total										
		Home saved /own stock	Carryover - maize hybrids	friends, neighbours, relatives	local market	agro-input dealer	community-based seed groups	government	NGO / FAO	contract seed growers	Other	TOTAL %
Maize	2395.6	**69.5**	0.0	5.2	15.2	0.0	0.1	6.1	3.9	0.0	0.0	**100.0**
Common beans	1800.5	**71.9**	0.0	10.0	15.4	0.0	0.0	0.0	0.0	0.0	2.8	**100.0**
Cassava	165.8	**92.0**	0.0	6.0	0.0	0.3	0.0	1.3	0.4	0.0	0.0	**100.0**
Sweet potato	37.2	**86.1**	0.0	10.1	0.4	0.0	0.0	3.2	0.2	0.0	0.0	**100.0**
Taro	30.5	**99.8**	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	**100.0**
Irish Potato	266.0	26.7	0.0	7.5	28.2	0.0	0.0	**37.6**	0.0	0.0	0.0	**100.0**
Peanut	57.6	**45.3**	0.0	4.3	**50.3**	0.0	0.0	0.0	0.0	0.0	0.0	**100.0**
Rice	2506.0	**76.3**	0.0	7.5	6.1	0.0	0.6	8.1	1.4	0.0	0.0	**100.0**
Pigeonpea	10.7	**84.5**	0.0	7.0	8.5	0.0	0.0	0.0	0.0	0.0	0.0	**100.0**
Arrowroot	81.0	**100.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**100.0**
TOTAL-all crops	7387.2	**71.8**	0.0	7.2	12.3	0.0	0.2	6.1	1.8	0.0	0.7	100.0

NOTE: the '**total kg**' by crop comes from the single 'total kg' figure (ex CrCuAkgTot, CrCuBkgTot, etc). Quantities fore **individual sources**

Codes | data entry | data check | demog | cu_srce | cu_>< | cu_why | qual | yld | resow | nx_srce | nx_>< | nx_why | fert | Pest | comp | chm | newvar | seed

After demographics, this sheet should be the first place you look when considering the findings. It immediately shows the key sources for each crop. Often, local markets are more important than expected, while agro-input dealers, NGOs, and friends/neighbors/relatives are often less important than expected. There may be important differences by crop. Reviewing this table with the team can help everyone appreciate which sources are important to farmers. Caution: Be sure to take a look at the column 'Total kg sowed.' For small amounts, the findings are hard to interpret (i.e. to draw conclusions). (Note that this rule applies more generally: for small samples, interpret findings cautiously.)

Cu_Acq - this sheet calculates, for every crop in the Current season, the total amount of seed planted, and then breaks this down by how the seed was acquired. These amounts add up all the farmers who

mentioned that crop and route of acquisition together. Amounts are in **kg** for seed crops, and '**kg of maize equivalents'** for vegetatively-propagated crops. **Cu_Acq** gives the raw figures, but is not easy to interpret quickly – the following sheet is much better for this.

Cu_Acq% - This sheet gives the % of seed by how it was acquired - for every crop, as well as for the sample as a whole. It calculates the total kg sowed for each crop, and uses figures from **Cu_Acq** to calculate the % of seed supplied from each source. Each row should add up to 100% - if they do not, that means that the data still need some cleaning.

Cu_>< This sheet gives the number of households growing each crop in the Current season, and the proportion of these households who are sowing More, Same, or Less than normal for that crop. The final column gives the average change in sowing rates for all crops with five or more cases. This is calculated as the mean of individual changes in sowing amounts. So if, for one crop, two individuals sowed 100% more than normal, two more sowed normal amounts (0% change), and one individual sowed half her normal amount (-50%), the average change would be +30% (100 + 100 + 0 + 0 – 50 = 150; 150 / 5 = 30%).

In the % change column, positive means are highlighted orange, negative means are in italics (Fig 6.3). This table is also important for review, to understand how the amount of seed (and by extension, the crop area) sowed is different from normal amounts. Often, aid agencies may assume that a crisis means that farmers have reduced the amount they sow – and that seed must be distributed to farmers as a result. However, farmers often make very modest reductions, and others may be increasing the amount they sow – for instance, Fig 5.3 shows almost no difference from normal amounts. Note that farmers do change strategies and dynamism in quantities sown, -10% or +10% may fall within the range of normal. (Be sure to tie this quantitative table to the one on reasons 'why'. Cu_why.)

Fig. 5.3 Example for Cu_<> sheet.

Amounts for current/most recent season: more, less, or same?

Crop	Number of HHs	% of HHs			Change sowing quantites for all growing the crop
		MORE	SAME	LESS	average % change
Maize	178	15.7	67.4	16.3	**1.25**
Common beans	54	16.7	51.9	31.5	**9.04**
Cassava	128	7.8	74.2	18.0	***-0.93***
Sweet potato	38	2.6	78.9	18.4	***-4.83***
Taro	19	5.3	89.5	5.3	**3.29**
Peanut	12	0.0	100.0	0.0	0.00
TOTAL-all crops	**554**	**10.6**	**72.4**	**17.9**	***-0.26***

Cu_why – This summarizes the reasons given by farmers for sowing LESS or MORE than normal. There are two tables: the top one lists all the reasons why farmers gave for sowing less than normal, the bottom table the reasons for sowing more. These reasons are grouped as **seed-related, non-seed factors of production,** and **other priorities.** The number of responses and % of the total are given for each answer (and also for sub-totals for the first two groups: Seed-related factors, and non-seed factors). The bottom of each table gives a total, which is the total of all farmers who said less (or more, in the lower table). The % in the bottom row may be less than 100% - this may occur if some farmers answered less (or more) but did not give a valid reason why. But it may also indicate some errors need correcting. Any percentage above 20% will be highlighted in orange (Fig 5.4).

These tables often yield surprising results. Aid agencies may assume that farmers reduce sowing amounts because seed is unavailable. However, seed availability is often not a major factor; farmers tend to change

the amounts they sow for other reasons. These reasons (money, labor, chronic stresses) may suggest issues to confront when developing an Action Plan.

Fig. 5.4 Section of first Cu_>< table. Here, only 30% of farmers who sowed less than normal did so for seed-related factors (20% was for seed access, i.e. often money), while labor was the reason that 35% of farmers sowed less than normal.

Why are you sowing LESS or MORE than normal?
Current/most recent season

Reasons farmers gave for planting LESS than normal in most recent (current) season

Reasons	N	% of responses
SEED- RELATED (or indirectly linked to seeds)		
Seed availability		
No seed available in market	0	0.0%
No seed/cuttings available from neighbors	3	6.1%
Seed access		
No money to buy seed/poor finances or seed too high	10	**20.4%**
Seed quality		
Seed available is not good quality or the variety is not liked	2	4.1%
Sub-total: seed-related	**15**	**30.6%**
NON-SEED FACTORS OF PRODUCTION (limits)		
No/insufficient labor	17	**34.7%**
Illness/health problems	0	0.0%
No/insufficient land or land not appropriate/sufficiently fertile	5	10.2%

demog | cu_srce | cu_>< | **cu_why** | qual | yld | resow | nx_srce | nx_>< | nx_why

'qual', 'yld', and **'resow'** – These three sheets have similar structures, each with two tables.

For **'qual'**, the top table looks at farmers' assessments of seed quality, **crop by crop**. Alongside each crop is the number of farmers growing that crop in the Current season, followed by how many rated the seed GOOD, AVERAGE, or POOR. The final three columns are the percentages giving that answer. The second table looks at farmers' assessments of seed quality according to the **source of seed** used (this table combines all crops that use a particular source). The sheet **'yld'** follows the same pattern, but gives farmers' assessments of the production in the Current (or most recent) season. Finally, **'resow'** summarizes farmers' responses to whether they plan to sow a particular seed lot again, or not.

In the % column of these tables, any figure above 20% will be highlighted for **Poor quality, Poor yield,** or **Will NOT resow.** These tables are useful for identifying if farmers note a problem of seed quality for particular crops or seed sources. Often, farmers tend to be quite positive about most seed quality – at least for the sources they choose themselves (they are sometimes more negative about seed given to them from NGOs, for instance). If they are more negative about yield than about seed quality, that suggests that the problems are with factors other than seed (e.g. climate, soil fertility, inputs).

iv) Sheets for Next season

These sheets are **'nx_srce'** and **'nx_srce%'**, which calculate the amounts and % of seed supplied by each source for the next season; **'nx_acq'** and **'nx_acq%'**, which calculate the amounts and % of seed by how it will be acquired for the next season **'nx_><'** shows whether the intended sowing amounts for next season are more or less than normal, and **'nx_why'** gives the reasons why there may be a difference.

These sheets work the same way as with Current season. Findings for the next season are important to show results for different crops (if Next Season features different crops than the Current Season), and highlight farmers' plans and longer-term responses to stress. For instance, farmers might sow less than normal for a key crop in the Current Season, but plan to sow much more than normal for Next Season. It is important to note where plans are dynamic, and where farmers are 'bouncing back' from stress.

v) Input use

This section contains sheets named: **'fert'**, **'pest'**, **'comp**', **'stor'**, and **'chem'**, which respectively explore farmers' use of fertilizer, pesticides, and compost, seed storage issues, and seed storage chemicals. For example, the sheet **'fert'** starts with a table giving the proportion of farmers using mineral fertilizer for both Current and for Next Season. This is followed by a list of the crops to which fertilizer is applied, with the % column being the percentage that all fertilizer applications that are given to that crop. Fig 5.5 shows these first two tables, and how % above 20% are highlighted in orange. In this example, slightly over half the farmers in the current season used mineral fertilizer, with most of them applying it to maize or cotton.

Fig. 5.5 Example, from top part of 'fert' worksheet.

1) Proportion (%) of HHs using minreal fertilizer

HHs using mineral fertilizer			
Current/most recent season		Next season	
Yes	50.8%	Yes	61.3%
No	49.2%	No	38.7%
N total	124	**N total**	124

2) Use of fertilizer by crop

Fertilizer use by crop				
Crop	Current/most recent season		Next season	
	N	%	N	%
Maize	23	**46.0%**	15	**27.8%**
Sorghum	5	10.0%	5	9.3%
Cotton	20	**40.0%**	33	**61.1%**
Tobacco	2	4.0%	1	1.9%
TOTAL-all crops	**50**	**100.0%**	**54**	**100.0%**

Table 3 is normally skipped over. It lists the types of fertilizer, and frequency of use, but needs coding after data-collection. Table 4 gives the reasons for NOT using fertilizer, for both Current and Next season. As before, this table tallies the frequency of each answer, and the % of the total who gave that reason.

The sheets **'pest'** and **'comp'** have identical structures, though the types of compost are already pre-coded so Table 3 is already filled in. The **'stor'** sheet starts out analyzing the proportion of farmers who used storage chemicals over the two seasons, and then the proportion of farmers who noted losses in storage after the last harvest. The third table gives the average storage losses, by crop. This counts the number of times a specific crop was mentioned as having suffered losses, and the average losses for each crop – this is the average of the losses given by each farmer for that crop. Fig 5.6 gives an example here (losses above 20% shaded orange). In that example, average storage losses of both maize and sorghum were over a third; millet was higher, though this reflected only two cases.

Fig. 5.6 Example of Table of Average storage losses by crop.

What losses did you have in storage?		
Crop	N	mean loss (%)
Maize	51	**36.2%**
Sorghum	37	**34.8%**
Millets	2	**85.0%**
Rice	7	28.6%

Finally, **'chm'** summarizes on which crops storage chemicals had been used, which storage chemicals were used (again, this needs coding afterwards, and is usually skipped) and the reasons for not using storage chemicals.

vi) New Varieties and Seed Aid

There are two sheets in this section: **'newvar'**, and **'seedaid'.** Each has five tables.

For New varieties, the first table gives an overview of the proportion of farmer who obtained a new variety in the past five years. For those who did obtain a new variety over the past five years, the right-hand columns of this table calculate the mean number of varieties received average (and standard error), as well as the minimum and maximum numbers of new varieties obtained by any single farmer in this sample. This table is important for understanding farmers' access to innovation. Fig 5.7 gives an example:

Fig 5.7 Example of new varieties table: 35 farmers obtained at least one new variety in the past five years (28.2% of the sample); those who did get a new variety got on average 1.7 varieties in this period, with a maximum of 6.

1) Obtained any new variety in last five years?

Number of HHs*	Obtained a new variety in past 5 years? (%)			HHs who received	N of varieties received in past 5 years			
	Yes	No	total		Mean	Std Dev	Min	Max
124	28.2%	71.8%	100.0%	35	**1.7**	1.02	1	6

* total who answered either 'yes' or 'no' to this question

The second table lists which sources provided new varieties, and the third table summarizes the means of access (purchase, gift, direct seed distribution, etc.). These tables are often interesting, as they may show if farmers rely on only a few sources for innovation – for instance, if most new varieties come from NGOs providing free seed via direct seed distribution, then this is generally unsustainable.

Table 4 on this sheet lists the crops for which farmers obtained new varieties, and what proportion of farmers are still sowing these crops. This table can show if new varieties are concentrated in only a few crops. It may also show if many farmers have stopped growing new varieties, which again may suggest issues around sustainability. Finally, Table 5 summarizes the years when new varieties have been received. As the interview question specifies 'received in the past FIVE years', we should expect the data to be within a five-year bloc; this table has a wider range of years, to facilitate future use.

The **'seedaid'** sheet has the same structure, but reports on whether farmers received seed aid in the past five years.

vii) Money spent on seed

The sheet '**money'** provides interesting and useful estimates of the amount of money the 'average farmer' may spend on seed in a season. This section requires you to enter in data, selecting key crops, and inputting seed prices. Once you have done this, you can run the analysis using F9, and obtain estimates of money spent on seeds, and how this breaks down by crop and source.

Filling in Tables 1 and 2

Table 1 is for the current season, Table 2 for next season. Each table provides space for you to enter three key crops, which you must identify yourself as important for understanding money spent on seed. Ideally, these should be crops that are grown by the majority of farmers (to check, see **demog** Table 7, or **Cu_><** and **Nx_><** to see numbers growing each crop). Once you have arrived at the most popular crops, if you still need to narrow down to three crops, then focus on those crops where a good proportion is purchased (see **'Cu_Srce%'** and **'Nx_srce%'** and crops with high % via local markets and agro-input dealers). Once you select these crops, enter their names in both tables, in Column B. **You must also add in the "crop code" alongside each crop, in column C.** These codes can be found on sheet 'Crops', or in Table 3.2 above. If crops repeat between seasons, put them in the same ROW (top, middle, or bottom) in each table, so that comparisons can be made between seasons.

Finally, columns D and E ask you to enter **typical prices** for these crops in local markets or agro-input stores, respectively. Prices should be in local currency. For 'local market', the prices should be for 'potential seed' (i.e. for grain that could potentially be used as seed), and for agro-input shops the price for 'seed'. These prices should reflect what farmers would normally pay in for these crops at those sources – so input shops would normally be much higher, as their prices are often for Certified Seed. Prices should be entered as the **price per kilogram (**or **'kg maize equivalent'** if the crop is vegetatively-propagated). Table 4.3 gives conversion rates – so, for example, if sweet potato vines are sold, then you need to give the price for enough vines, or a bundle, that will give 500 plants. These typical prices can be gathered from visits to input shops, and to local markets – ideally, the local market price should be the mean of prices actually given by market traders for potential seed of that crop. Prices in local markets can vary through the season, but prices should be for the period when most farmers are buying seed.

Fig. 5.8- Tables in the sheet 'money' where you must enter information, in the yellow cells. Note in this example that input shops are more expensive, and that sesame seed is cheaper next season than the current one.

1) Most recent (current) season

most important crops		typical price / kg*	
Name	Code	local market	input shops
maize	1	12	20
groundnut	8	9	15
sesame	14	4	8

2) Next season

most important crops		typical price / kg*	
Name	Code	local market	input shops
maize	1	12	20
groundnut	8	9	15
sesame	14	2.5	6

* 1) the price during the period when most farmers are buying seed
2) for vegetative crops, a 'kg' represents a 'kg of maize equivalent'
(E.g. the price for 400 cassava cuttings)

Once you have added in this information, **re-run the calculations (Press F9).**

An example of results is shown in Fig. 5.8. the results are based on the prices you entered and the actual amounts of seed farmers obtained from local markets, agro-input shops, and from neighbors (meaning from any other farmer, but ONLY IF that seed were **purchased**). The figures are in local currency, and represent the average amount spent by farmers on seed from these three sources. In the example here, farmers spent more in local markets than elsewhere, and input shops were only really used for maize, though still relatively minor amounts were spent there. The totals at the bottom only add up the three crops in the table. The total for all sources (184.8 of local currency, in this case Kwacha) represents a 'typical' expenditure for a farmer who is growing all three crops listed here. This total is broken down by crop in the last two columns, showing that groundnuts comprise more than half the amount spent on seed.

Fig 5.8 Example of table on money spent on seed in current season.

3) Analysis of spending on seed from markets and input shops for 3 most important crops in most recent (current) season

most important crops	N growing this crop	Average Spending*				
		Neighbors	local market	input shops	All sources	% of total
maize	123	14.5	32.1	18.9	**65.5**	**35.4%**
Ground nut	107	9.7	86.1	1.7	**97.5**	**52.8%**
sesame	85	5.4	16.4	0.0	**21.8**	**11.8%**
total (of 3)		**29.6**	**134.6**	**20.6**	**184.8**	**100.0%**

* local currency

This analysis is done for the current season (Table 3) as well as for the next season (Table 4). Comparing these two tables can yield interesting findings – e.g. if the total amount spent on seed is going up or down,

or if the proportion of money spent by crop or by source varies between the seasons. Such figures might also be used to determine if farmers (or select farmers) might need some financial support.

Finally, you can also use this sheet to **model different scenarios** and explore how they might affect the amounts farmers spend on seed. For instance, you can enter different prices into Tables 1 and 2, re-run the analysis (F9), and see how this affects the amount spent. For instance, doubling the seed price in an agro-input shop may have a relatively small impact on what a typical farmer spends, if these shops are not used to a great extent. You may also wish to see how entering different crops (with different codes and prices) affects the outcome. These tables can have a strong impact on policy-makers and key stakeholders, as they are often surprised by just how much money farmers spend on seed, which crops they spend it on (e.g. on legumes), and where (e.g. in local markets). These figures, while illustrative, can help you make valuable points about the seed systems that farmers actually use.

viii) Statistical analysis

The final pair of sheets, **'stat1'** and **'stat2',** run a set of simple statistical analyses, which start to examine if there are key trends that are statistically significant. Every table has a significance test, giving a probability (p) value, with the cell highlighted **orange for p<0.05 (alpha=5%), and yellow for p between 0.05 and 0.1 (alpha between 5% and 10%).**

All of these statistical tests work best if there are reasonably large sizes for each category. For example, if there are only a few female-headed households, or few 'other' households that are not headed by adults, then the statistical tests will not be very powerful for these comparisons - in other words, you will probably not see any clear results. Also, if entire categories are empty (e.g. there are no farmers <0.5 ha, or no farmers >2.0 ha), then the chi-squared tests will not function at all. In such cases, statistical tests may still be possible, but you would probably need to combine categories, and would have to run the tests in a separate statistical program – the categories for the automatic analyses here are pre-set.

In the context of an SSSA, it is probably best not to spend too much time examining these statistical tests. If there are significant trends emerging, they will be signaled by orange (or yellow) highlighting, which then prompts closer scrutiny. It can be useful to refer to such trends in presentations and reports, especially if you have other evidence around these trends. In the end, it is important not to get overly-fixated on statistical tests, as the evidence produced in the many other worksheets is far more informative and useful for understanding seed systems and developing an action plan.

Firstly, **'stat1'** looks if there are any significant differences between households headed by men or by women. Table 1 compares the number of times seed aid was received by male-headed versus female-headed households, using a t-test (2-sided) to compare means (Fig 5.9). The Null Hypothesis is that the two means are the same; the t test probability is the likelihood that this Hypothesis is true.

Fig 5.9 Example of statistical table comparing mean number of times received seed aid, by gender of household head – significance test is circled.

1) Number of times received seed aid in last 5 years

	Seed aid last 5 years			
HH Head	N	Mean number	SD	t test Prob.
Male	53	1.63	0.31	0.0473
Female	21	1.10	0.20	
total	74	1.31	0.21	

p < 0.05, so mean # times seed aid received is statistically different (at 5%) between men and women-headed HH

NOTE - values for probabilities of 5% and 10% are 0.0500 and 0.1000; a value of 0.0000 means a probability of <0.01%

Table 2 runs a Chi-square test, comparing the proportion of male- and female-headed households who had received seed aid, against the numbers that would be expected if there were no bias according to gender. Fig. 5.10 gives an example, where almost exactly the same proportion of male- and female-headed households received seed aid (just less than a half): thus it is not surprising that the probability figure (0.9795) is nowhere near significant (far above 0.05).

Fig. 5.10 Example of Chi-squared test, which assesses whether the proportion of answers (yes/no) in this case, is different than what would be expected, given the overall proportion of male and female-headed households.

2) Proportion receiving seed aid in last 5 years

		Received seed aid?		
HH Head	N	Yes	No	Chi-Sq prob.
Male	170	80	90	**0.9795**
Female	19	9	10	
Total	189	89	100	

The other tables in **'stat1'** follow a similar pattern. Table 3 runs a t-test on the mean number of times seed aid was received, according to gender, while Table 4 runs a chi-squared test on gender and the proportion receiving seed aid. Tables 5 – 10 run similar chi-squared tests on gender and use of inputs (fertilizer, compost, storage chemicals). Tables 11 and 12 look at frequency of sowing more / same / less than normal by gender. In Fig. 5.11, we can see that the one female-headed household which sowed more than normal was a smaller number of households than would be expected ((54 / 559) X 59 = **5.7**), while the 41 women sowing normal amounts is slightly more than expected ((401 / 559) X 54 = **38.7**). The significance of these trends is below 10%, but not below 5% (hence is shaded yellow)

Fig. 5.11. Example of chi-squared test of gender, in relation to quantity of seed sowed, in relation to normal amounts.

11) Quantity sowed most recent (current) season in comparison with the normal quantity (all crops)

		Quantity sowed most recent (current) season			
HH Head	N	More	Same	Less	Prob. Chi. Sq.
Male	505	58	360	87	**0.0772**
Female	54	1	41	12	
Total	559	59	401	99	

Table 13 uses a chi-squared test to see if the proportion of farms of different sizes differs according to gender. Finally, Tables 14 and 15 run t-tests to see if there is a difference in the mean family size, or mean age of household head, according to gender.

'stat2' makes comparisons based on other criteria, such as farm size. Table 16 shows the mean number of times seed aid was received for the four different ranges of farm size. There is no statistical test here, as that would require an ANOVA, and therefore a dedicated statistical package: but the means in Table 16 may indicate if there is a basis for further exploration. Table 17 runs a chi-squared test on the proportion of farms of different sizes that received seed aid. Tables 18 and 19 are similar, but look at the number of times, and proportions, farms of different sizes obtained new varieties. Tables 20 – 25 look at the proportions using inputs. Fig 5.12 gives another example, which may help in understanding how these tests operate. In this example, almost exactly the same numbers **overall** used fertilizer as did not. Such equal proportions are seen among the smallest farms (<0.5 ha), while well over half of the largest farms used fertilizer. Thus, the significant Chi-squared test is not surprising. Be aware that the low p-value signals a significant difference farm area groups, though it doesn't specify which ones (i.e. this is an overall chi-squared test, not pairwise).

Fig 5.12. Example of chi-squared test of mineral fertilizer use, according to farm area.

20) Use of mineral fertilizer in most recent (current) season?

Farm area (ha)	N	fertilizer most recent (current) season? Yes	No	Chi Sq. Prob.
< 0.5	8	4	4	**0.0007**
0.5 - 1.0	17	2	15	
1.0 - 2.0	30	12	18	
>2.0	68	44	24	
Total	123	62	61	

Tables 26 and 27 compare proportions sowing more / same / less than normal, against farm size, and Table 28 presents mean ages of household head against farm size, though without a statistical test.

Finally, Tables 29-34 explore trends according to whether or not households are headed by an adult (as normal), or are headed either by a child or a grandparent with dependent children (listed as 'other' in the Tables.) These operate tests similar to those done for gender of HH head, comparing two groups.

KEY LESSONS FROM THIS SECTION:

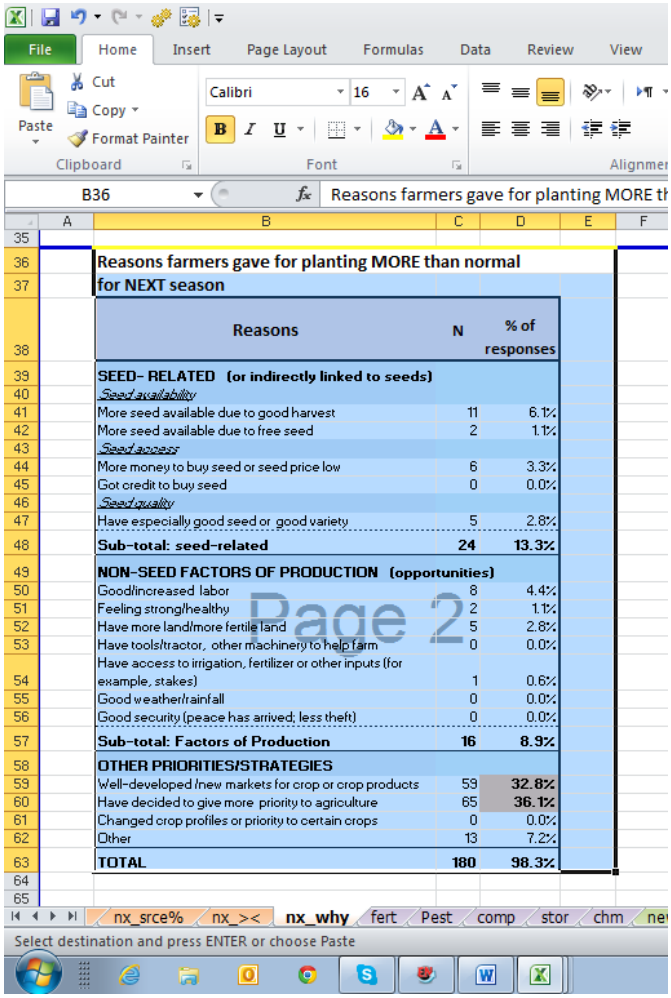
1. **Immediately after data entry, data analysis tables will populate.**
2. **The tables will display information about the characteristics of households, their current season seed sourcing, next season seed sourcing, input use, approximate money spent, and basic statistical analyses.**
3. **Often it is best to begin with the "demog" worksheet, followed by the "Cu_srce%" and "Cu_why" worksheets.**
4. **Statistical analyses deserve modest attention, but action plans should be primarily based on other worksheets.**

6. HOW TO COPY AND PASTE TABLES FOR USE IN REPORTS

The Excel tables have been formatted so that it is possible to copy them, with formatting intact, into other documents such as PowerPoint presentations or Word reports.

- In Excel, **select** the Table you wish to paste into another document (Fig 6.1 gives an example)

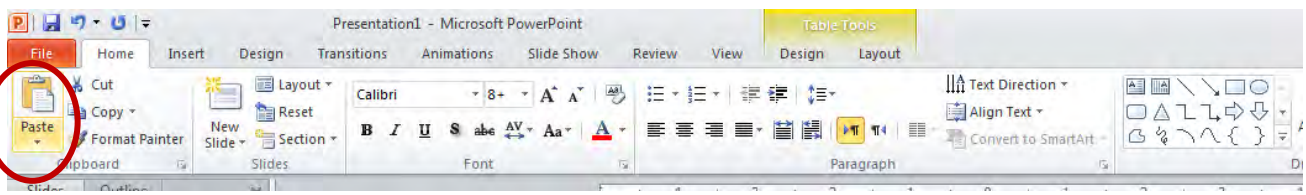
Fig 6.1 – selecting a table: reasons farmers plan to sow more next season



Reasons	N	% of responses
SEED- RELATED (or indirectly linked to seeds)		
Seed availability		
More seed available due to good harvest	11	6.1%
More seed available due to free seed	2	1.1%
Seed access		
More money to buy seed or seed price low	6	3.3%
Got credit to buy seed	0	0.0%
Seed quality		
Have especially good seed or good variety	5	2.8%
Sub-total: seed-related	**24**	**13.3%**
NON-SEED FACTORS OF PRODUCTION (opportunities)		
Good/increased labor	8	4.4%
Feeling strong/healthy	2	1.1%
Have more land/more fertile land	5	2.8%
Have tools/tractor, other machinery to help farm	0	0.0%
Have access to irrigation, fertilizer or other inputs (for example, stakes)	1	0.6%
Good weather/rainfall	0	0.0%
Good security (peace has arrived; less theft)	0	0.0%
Sub-total: Factors of Production	**16**	**8.9%**
OTHER PRIORITIES/STRATEGIES		
Well-developed /new markets for crop or crop products	59	**32.8%**
Have decided to give more priority to agriculture	65	**36.1%**
Changed crop profiles or priority to certain crops	0	0.0%
Other	13	7.2%
TOTAL	**180**	**98.3%**

- **Copy** your selection (**CTRL-C,** right-clicking to copy, or using the menu) to the clipboard
- Go to the destination document
- Select among options for **pasting.** This can be done either by clicking on the 'Paste' icon on the 'Home menu' (fig 6.2), or right-clicking with your mouse (You can use **CTRL-V** to paste, but this does give you any choice of <u>how</u> you paste – see below)

Fig. 6.2 Location of paste icon.



- You will be presented with several options for pasting, which may include
    - **'Keep source formatting'** – will preserve the look (font, colors, headings) in the original text. This is often the best choice.
    - '**Use destination styles'** – will match style used (e.g. font) where the material is pasted.
    - **'Link and keep source formatting'** –keeps original formatting, but will also maintain a link to the source file, and will update the pasted text with any changes made to the source file.
    - **'Link and use destination styles' –** will match style used where material is pasted, but maintain a link to source file to update test with any changes made to the source file.
    - '**Picture' –** inserts the table as an image. This means you cannot edit contents, only change size or shape of the image.

It is generally best to use 'KEEP SOURCE FORMATTING'. This allows you to edit different features, such as borders in a table, or width of cells. If you have copied text along with a table from Excel, then the text (and any blank spaces around the text) will be framed within hidden boundaries, in the same way that table cells are set. For the text to display the way you want, you may need to merge these cells, or delete blank rows.

APPENDIX 1: Worksheets

Data entry and management (light gray)

1. **Codes** – Lists all 272 variables for data entry, along with codes or notes on entry
2. **Crops** – Lists all crops, with space to add additional crops
3. **MSL codes** – Lists all possible coded reasons that a farmer planted more or less compared to the previous season.
4. **Data entry** – Used to enter in records from survey forms
5. **Data check** – Highlights inconsistencies in data; used to check for errors

Demographics (dark gray)

6. **Demog** – gives demographic data of sample (e.g. gender, distribution of farm sizes, main crops, mean and distribution of age of HH head and family sizes

Seed sources and performance for CURRENT (or most recent) season (bright green)

7. **cu_srce** – gives quantities of seed provided from each source, broken down by every crop.
8. **cu_srce%** - gives the % of seed supplied by each seed source, broken down by every crop.
9. **cu_acq** – gives quantities of seed by how they were acquired, broken down by every crop.
10. **cu_acq%** - gives the % of seed by how they were acquired, broken down by every crop.
11. **cu_>< ** - number of households growing each crop; proportions sowing More, Same, or Less than normal for that crop, and the average change in sowing rates, by crop
12. **cu_why** - reasons why farmers sow LESS or MORE than normal
13. **qual** – farmers' assessments of seed quality, by crop and source
14. **yld** - farmers' assessments of yield, by crop and source
15. **resow –** farmers' intentions to sow again, by crop and source

Seed sources and performance for NEXT season (light orange)

16. **nx_srce** - gives quantities of seed provided from each source, broken down by every crop.
17. **nx_srce% -** gives the % of seed supplied by each seed source, broken down by every crop.
18. **nx_acq** – gives quantities of seed by how they were acquired, broken down by every crop.
19. **nx_acq%** - gives the % of seed by how they were acquired, broken down by every crop.
20. **nx_>< ** - number of households growing each crop; proportions sowing More, Same, or Less than normal for that crop, and the average change in sowing rates, by crop
21. **nx_why** - reasons why farmers sow LESS or MORE than normal

Input use (light purple)

22. **fert –** fertilizer use: overall %, by crop, and reasons for not using
23. **pest –** pesticide use: overall %, by crop, and reasons for not using
24. **comp -** compost / manure use: overall %, by crop, type of compost, and reasons for not using
25. **stor -** % who use storage chemicals, and mean losses in storage by crop
26. **chm –** storage chemical use: overall %, by crop, and reasons for not using

New varieties / Seed aid

27. **newvar –** how many obtained a new variety in past five years, along with sources, means of access, and whether still sowed
28. **seedaid -** how many received seed aid, along with mode of aid, organization and crop

Money (yellow)

29. **money -** estimates of the amount of money the 'average farmer' may spend on seed in a season

Statistical analysis (blue)

30. **stat1** – general statistical comparisons, by gender of household head
31. **stat2** – general statistical comparisons, by farm size, and type of household head (adult vs. other)