

III Quantitative Analysis Guidelines

The Quantitative Analysis Guidelines are designed to help users interested in analysing the data derived from the quantitative methods available in *The BAG* and *The Grab Bag*. In this section, we provide detailed guidance on the use of computer spreadsheets, simple descriptive data analysis, and an introduction to more complicated statistical tests. Given the different levels of computer literacy that we have encountered within our own tests of assessment methods for social C&I, we concluded that 'starting simple' was the safest route. If you are already familiar with computer spreadsheets, you may want to skip over much of the following material. On the other hand, if you are a relatively new computer user, we hope these instructions will be helpful.

A – ANALYSIS OF RESULTS FROM A STAKEHOLDER IDENTIFICATION METHOD (CATPAC)

CatPac is a software package that uses neural network analysis to identify clusters in human speech. Most communications experts consider human speech to reflect human thought. CatPac has been used, in the context of assessment of human well-being in forests, to identify relevant stakeholders in forests in Cameroon, Indonesia and Brazil. Instructions on using the method are available in *The Grab Bag*. This manual deals only with analysis of the results.

CATPAC GUIDELINES[20]

This section describes, for the new CatPac user, how to enter the text/data, some of the computer-related features requiring decisions, and how to make the dendograms and cognitive maps that comprise CatPac results.

a. How to enter the text/data

- After you record the interview, the next step is entering the

20 There is a CatPac User Manual by Woelfel (1998). The manual contains more comprehensive information on this software. The purpose of our guidelines is only to give you the most crucial procedures in running CatPac.

text. You can enter the text using any Word Processing software, but remember to press **Enter** at the end of each line. This is to keep the text in a readable format[21] when you open the text using the CatPac software.

- After you finish entering the text from one participant, type a '–1' as a separator between participants. Be sure to enter the data from one category (men/women, different ethnic groups, occupation, etc.) in one file.
- Save the file in text format (*.txt). CatPac can only open text files.

b. How to choose the appropriate clustering method

All the clustering techniques provided by CatPac are based on an agglomerative method. There are no exact rules to choose. To know more about the advantages and disadvantages of each method, you can go to the **O**ption menu, and then choose Clustering method. A dialogue menu will prompt you. Choose **H**elp if you want to know more about each clustering method.

c. How to make an exclude file

Since you are interested in the clustering of significant words/concepts, you will likely want to exclude from your analysis certain grammatical and other words (such as 'is', 'very', 'it'). CatPac has a default exclude file[22] in English. In many cases this will not help you much, since most interviews should be in the local language. In that case, you have to set up an exclude file for that language. For initial purposes, you can translate the English file and use the translation as the new exclude file. You can do it by retrieving the English version file, translating the words line by line, and then saving

21 Many of our collaborators found the text that they already entered became a single long line of text when they opened the text file in the CatPac software. It occurred because they did not press Enter at the end of each line.

22 A file that contains a list of words appearing in the text of your interviews that have a primarily grammatical meaning, and do not contribute significantly to understanding of concepts related to the study.

them in another file with *.exc extension. It will perhaps be easier to imagine comparable grammatical markers in the local language and list them in the exclude file. If you find there are more irrelevant words that need to be excluded, you can add them to the list, by appending them at the bottom of your new exclude file.

d. How to make a dendogram

A dendogram is a figure representing the cognitive clusters of concepts, based on the speech (or text) you have entered and analysed. It normally looks something like a glove (see example below).

- Open the exclude file first, to activate the file. The exclude file will not be active unless you open it. Then open the text file that you want to analyse. You can use the default setting for options, including window size, slide size, clamping, clustering technique, etc.[23]
- Go to the **F**ile menu and choose **R**un, and then choose **M**ake dendogram. CatPac will give you the results in a few seconds. You can print the dendogram directly by choosing Print under the **F**ile menu. You also can save the dendogram as *.den file by choosing Save **A**s under the **F**ile menu (see Dendogram sample below).

23 The use of each option and how to choose the right option for your text can be found in the CatPac User Manual.

DENDOGRAM SAMPLE

TOTAL WORDS	96	THRESHOLD	0.000
TOTAL UNIQUE WORDS	25	RESTORING FORCE	0.100
TOTAL EPISODES	101	CYCLES	1
TOTAL LINES	27	FUNCTION	Hyperbolic
		CLAMPING	Yes

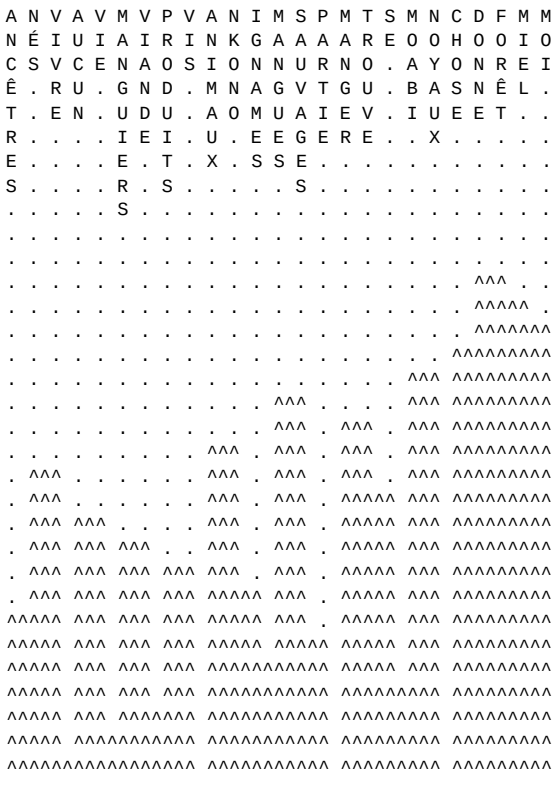
DESCENDING FREQUENCY LIST

WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
FORÊT (forest)	25	26.0	94	93.1
SAUVAGES (wild)	8	8.3	48	47.5
DONNE (give)	5	5.2	34	33.7
MANGUES (mango)	4	4.2	27	26.7
MIEL (honey)	4	4.2	31	30.7
NKONO	4	4.2	28	27.7
NOYAUX (pits)	4	4.2	28	27.7
TROUVE (find)	4	4.2	36	35.6
ANIMAUX (animals)	3	3.1	21	20.8
IGNAMES (yams)	3	3.1	28	27.7
MANGER (to eat)	3	3.1	21	20.8
MOABI [a tree]	3	3.1	28	27.7
MOI (me)	3	3.1	21	20.8
PARTIE (part)	3	3.1	9	8.9
CHOSE (thing)	2	2.1	10	9.9
MANGUIERS	2	2.1	10	9.9
NÉS (born)	2	2.1	13	12.9
PRODUITS (products)	2	2.1	14	13.9
SE [reflexive, self]	2	2.1	14	13.9
VIANDE (meat)	2	2.1	14	13.9
VIE (life)	2	2.1	14	13.9
VIS (lives)	2	2.1	14	13.9
VIVRE (to live)	2	2.1	14	13.9
ANCÊTRES (ancestors)	1	1.0	7	6.9
AUCUN (none)	1	1.0	7	6.9

ALPHABETICALLY SORTED LIST

WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
ANCÊTRES	1	1.0	7	6.9
ANIMAUX	3	3.1	21	20.8
AUCUN	1	1.0	7	6.9
CHOSE	2	2.1	10	9.9
DONNE	5	5.2	34	33.7
FORÊT	25	26.0	94	93.1
IGNAMES	3	3.1	28	27.7
MANGER	3	3.1	21	20.8
MANGUES	4	4.2	27	26.7
MANGUIERS	2	2.1	10	9.9
MIEL	4	4.2	31	30.7
MOABI	3	3.1	28	27.7
MOI	3	3.1	21	20.8
NKONO	4	4.2	28	27.7
NOYAUX	4	4.2	28	27.7
NÉS	2	2.1	13	12.9
PARTIE	3	3.1	9	8.9
PRODUITS	2	2.1	14	13.9
SAUVAGES	8	8.3	48	47.5
SE	2	2.1	14	13.9
TROUVE	4	4.2	36	35.6
VIANDE	2	2.1	14	13.9
VIE	2	2.1	14	13.9
VIS	2	2.1	14	13.9
VIVRE	2	2.1	14	13.9

WARDS METHOD



Interpretation: These data provide you with a visual expression of the clustering of words in the recorded text. The importance of these clusters derives from the perspective that people's speech reflects their thoughts, to a certain degree. So if, for instance, the word 'forest' and 'good' appear together frequently (forming a cluster), we might conclude that these people consider forests to be good, which one might interpret to favour sustainability. In the above example, the French words for thing (*chose*), give (*donne*), forest (*forêt*), honey (*miel*), and me (*moi*) represent the most important cluster.

This suggests an important role of honey in local forest management. Pits (*noyaux*) and *moabi* (a Cameroonian tree) are in another important cluster close to the previous one, suggesting that these may be important forest products that people use. The clustering of words in people's speech about a topic such as forests (or any other researchable issue) gives important hints about relations between forest and other elements in the local environment, human and otherwise.

e. How to make a cognitive map[24] from CatPac

- CatPac also has the ability to produce a cognitive map. You can get the cognitive map by saving the dendogram file as a coordinate file (*.crd). You will be prompted by a dialogue box. Just click **OK** and type the name of your coordinate file.
- You can open the Thought View software[25], then open the coordinate file. You can do some operations on the picture (i.e., rotate the picture; shrink or enlarge; change the cube, grid, or font colour, etc.).
- After you are satisfied with the appearance of the picture, you can copy and paste it to another software by choosing **C**opy (Ctrl+Ins) under the **E**dit menu. You can see a sample of a cognitive map below.

24 The cognitive map can be used to help interpret the clusters. Most of the time, the cognitive map gives the same configuration as the dendogram, but sometimes it may provide new insights. If the results given by the two methods are conflicting, it is strongly recommended to use the results from the dendogram as they are more accurate.

25 Thought View is a complementary program for CatPac. You can install Thought View by clicking **Setup.exe** file in CatPac CD attached and following the instruction.

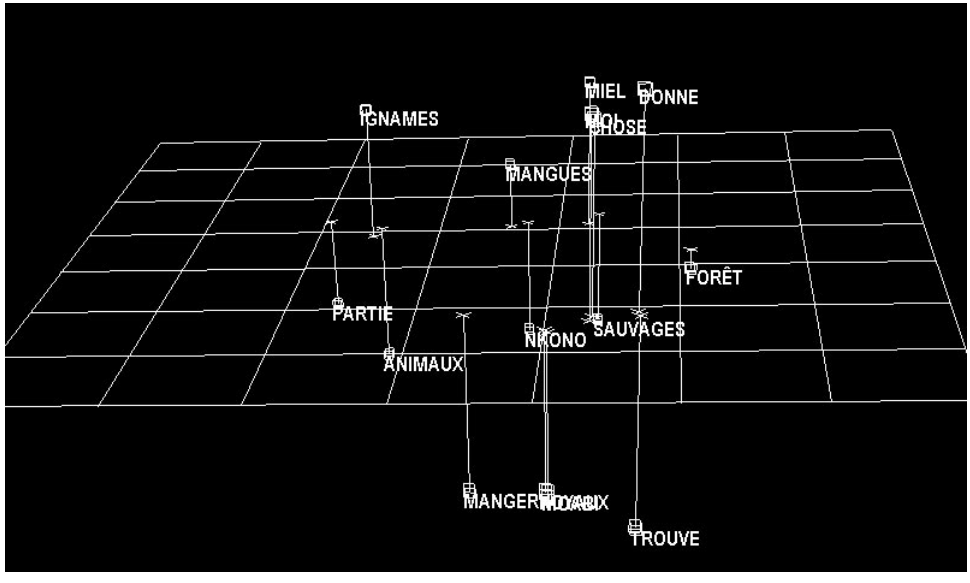


FIGURE 2. Cognitive map for Baka females (Dja Reserve, Cameroon)

The cognitive map is sometimes useful to get a rough visual representation of people's thinking. Again, you can see the close connections among the words honey (*miel*), me (*moi*), thing (*chose*), and give (*donne*). Forest (*forêt*) appears further away than one might guess from the cluster analysis. Toward the bottom of the plot the two unreadable terms are pits (*noyaux*) and *moabi* (a valuable Cameroonian tree).

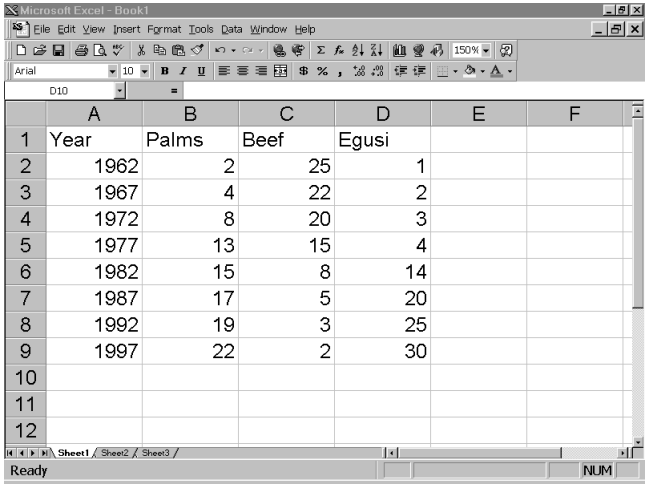
B – ANALYSIS OF METHODS ON SECURITY OF INTERGENERATIONAL ACCESS TO RESOURCES

This section contains analytical guidance for three social science assessment methods: the Histo-Ecological Matrix, Access to Resources by Generation, and Benefit Sharing among Stakeholders. These methods have been found useful in assessing security of intergenerational access to resources in certain forests among certain groups of people.

1. HISTO-ECOLOGICAL MATRIX GUIDELINES

a. Entering the data

You can follow the format as shown in Figure 3. The resources identified in the field are listed across the top (palms, beef, *egusi*), while the years are listed vertically. You do not have to enter the data in consecutive years. We would, however, suggest you do so for ease of reading.



	A	B	C	D	E	F
1	Year	Palms	Beef	Egusi		
2	1962	2	25	1		
3	1967	4	22	2		
4	1972	8	20	3		
5	1977	13	15	4		
6	1982	15	8	14		
7	1987	17	5	20		
8	1992	19	3	25		
9	1997	22	2	30		
10						
11						
12						

FIGURE 3. Spreadsheet format for histo-ecological data

b. Checking linear trend over time

You can determine whether the data show a trend over time by using the regression facility in Excel. Add the regression command into Excel by installing its Add-Ins program.

Follow these steps to install the program:

- Go to the Tools menu and choose Add-Ins (Figure 4),

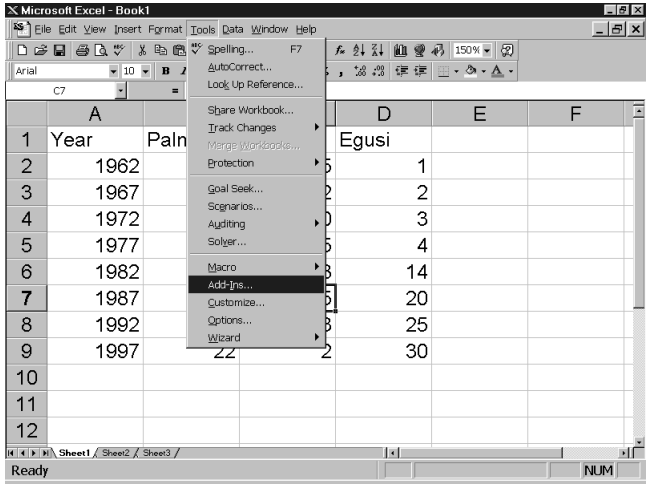


FIGURE 4. Choosing Add-Ins submenu (histo-ecological data)

Click on Analysis ToolPak and Analysis ToolPak-VBA (Figure 5). Do not click the options if they are already marked, which means the Add-Ins program has already been installed on your computer.

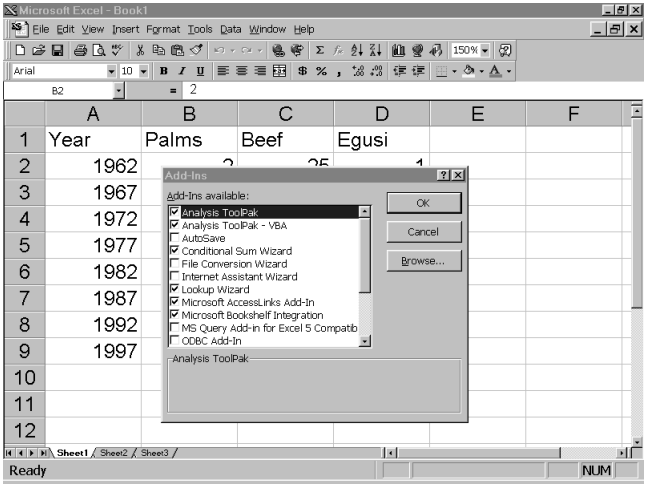


FIGURE 5. Choosing specified Add-Ins module (histo-ecological data)

- After you install this program, you will see there is a new sub menu under the Tools menu, that is: **D**ata Analysis, at the bottom of the **T**ools menu. You can do a regression analysis now by choosing the **D**ata Analysis sub menu.

- A dialogue menu as shown in Figure 6 will prompt you. Choose the regression option from the list of analysis tools and then click **OK** (Figure 6).

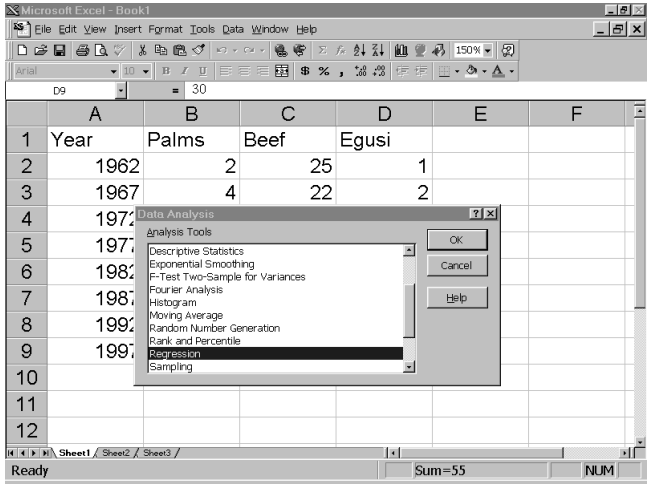


FIGURE 6. Starting regression analysis for histo-ecological data

- The next step is defining the range (the 'cells' in the spreadsheet that contain the data) for dependent variable (Y) and independent variable (X). Input the cell's range in a dialogue menu as shown in Figure 7. For instance, if you want to check the trends on palms, in the example above you can treat the palms variable as the dependent variable and year (time) as the independent variable. Do not forget to click on Residual plot, Line Fit Plot, and Normal Probability Plot, as you will need them later to check the model's adequacy. Then, click **OK**.

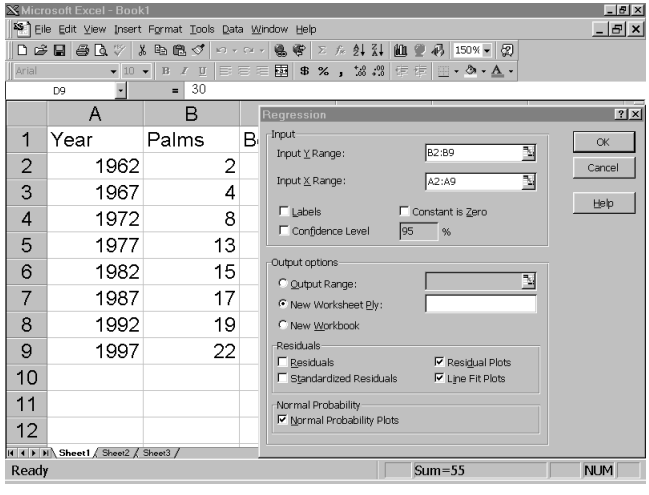


FIGURE 7. Defining variable's cell range (histo-ecological data)

- The output of the analysis that you will get is shown in Figure 8. You can check the model's significance by looking at the p-value (shown by the arrow). If the p-value falls below 5% (0.05), it can be concluded that there is a linear[26] trend over time. The line's slope will tell you important information about people's perceptions of changes in the resources over time. If the slope has a negative sign, it means the amount of that resource is perceived to be decreasing. If the slope of the sign is positive, the reverse is true. The resource is perceived to be increasing. You also can check how well the equation fits the data by looking at the R-square. The higher the R-square, the better the fit[27].

26 If you find the p-value is not significant, it doesn't mean the data do not have a pattern or trend over time. They may have a non-linear trend (i.e., quadratic, cubic, etc.). Excel does not provide analyses tools for non-linear trends. You can use more sophisticated statistical software, such as SPSS, SAS or Minitab, in such cases.

27 The R-sq is very sensitive to outliers and influential observation. There are several statistic instead of R-sq which has been introduced for assessing goodness of fit, including PRESS statistic (Draper & Smith 1982). However for our simple model, R-sq is adequate.

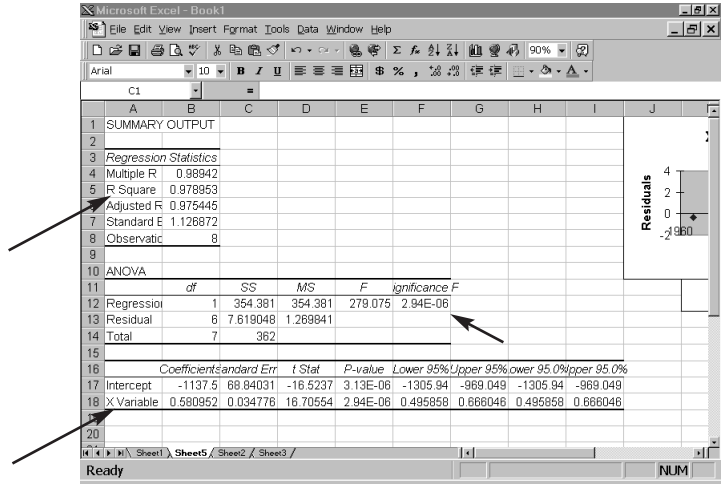


FIGURE 8. Output of regression analysis for histo-ecological data

c. Checking model validity

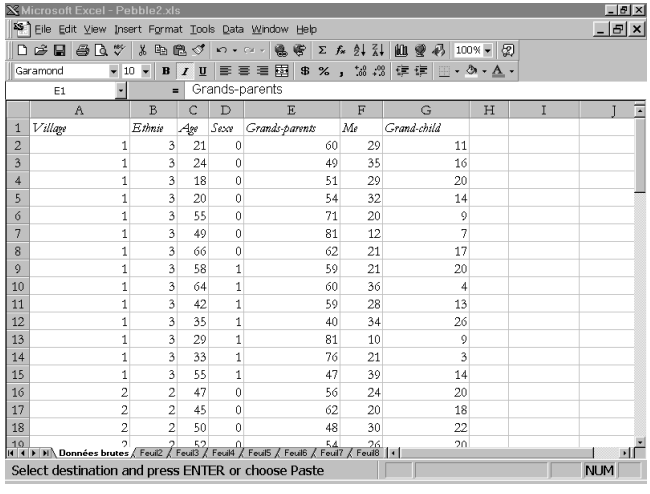
Regression analysis, which has been used widely to model the relationship between two variables, has several assumptions that need to be satisfied to get a valid (reliable) model. Those assumptions are normality of residuals, homogeneity of residual variances, uncorrelated residuals, etc. More detailed discussion on those assumptions can be found in (Draper & Smith 1981), although appendix A in this manual might give you insight how to investigate those assumptions using explanatory and graphic approach. However, we strongly suggest you use statistical software, such as Minitab, SAS, and SPSS, to perform more a detailed inspection. Because our data are similar to time series data, it is more likely there will be correlated residuals that violate the assumptions. In this case, autoregressive regression (Makridakis *et al.* 1983).

2. ACCESS TO RESOURCES BY GENERATION GUIDELINES

This section first provides information on how to set up a spreadsheet and enter the data for the pebble distribution method on access to resources by generation (in *The BAG*). Next, we explain how to perform simple statistical analyses on the data, including how to check data distribution and generate simple descriptive tables. For those of you who are satisfied with simple analyses, you can stop there. The final part is intended for those who want to conduct more advanced analyses, such as testing the significance of differences of opinion among different groups (based on ethnicity, age, gender, etc.).

a. Entering the data

- Enter the data using spreadsheet software, such as Excel or Lotus, if possible.
- Enter the demographic data, such as user group, gender, age, educational level, etc., in the left-hand columns (as shown in Figure 9, columns A–D). Insofar as possible, enter the demographic data using numeric codes. If you have categorical data, such as gender, educational level, etc., you need to convert them into numeric codes (e.g., female = 0, male = 1, etc.). It is very important to keep a coding sheet, so you will know the meaning of each code. You will need it when you undertake the analysis. We suggest you record the coding information within the spreadsheet.



	A	B	C	D	E	F	G
1	*Village*	*Ethnie*	*Age*	*Sexe*	*Grands-parents*	*Me*	*Grand-child*
2	1	3	21	0	60	29	11
3	1	3	24	0	49	35	16
4	1	3	18	0	51	29	20
5	1	3	20	0	54	32	14
6	1	3	55	0	71	20	9
7	1	3	49	0	81	12	7
8	1	3	66	0	62	21	17
9	1	3	58	1	59	21	20
10	1	3	64	1	60	36	4
11	1	3	42	1	59	28	13
12	1	3	35	1	40	34	26
13	1	3	29	1	81	10	9
14	1	3	33	1	76	21	3
15	1	3	55	1	47	39	14
16	2	2	47	0	56	24	20
17	2	2	45	0	62	20	18
18	2	2	50	0	48	30	22

FIGURE 9. Spreadsheet format for *inter-generation* data

- You can use the next column to enter the amount of pebbles for each generation (i.e., grandparent, me, grandchild[28]) allocated by the respondent. You only need one row for each respondent.
- After you finish entering the data for all respondents, do not forget to save the file in Lotus 123 format (*.wk1).

b. Conducting simple data analyses using SPSS[29]

We use SPSS (Statistical Package for Social Sciences) to analyse the data.

- **Opening the data** — You can open the Lotus 123 file in an SPSS spreadsheet. Use the read variable name[30] option.

28 Usually we use three categories, grandparents, me, and grandchildren, but the East Kalimantan team (1997) added two more categories: parents and children.

29 We have also made guidelines for those of you who want to undertake simple analyses using Excel (see Part c of this section).

30 From the SPSS menu bar, you can choose **F**ile → **O**pen, then change the file type to Lotus (*.wk1) and choose the file you want to open. The SPSS processor automatically will ask you whether you want to read the variable name or not. Just click on the 'read the variable names' option.

- **Labelling the data** — Use the coding sheet (the one that you made when you coded the demographic data into a numeric code to label the data). For labelling the data, place the cursor at the column where the demographic data are. Below is an example of labelling gender data.

From menus, choose: Data
Define variable (Figure 10)

Click **Label** button (shown by the arrow).
You will be asked to enter your label (Figure 11).

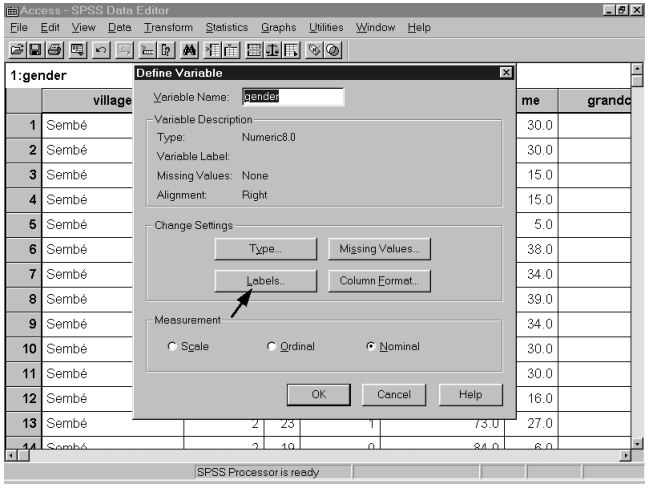


FIGURE 10. Label button (inter-generation data)

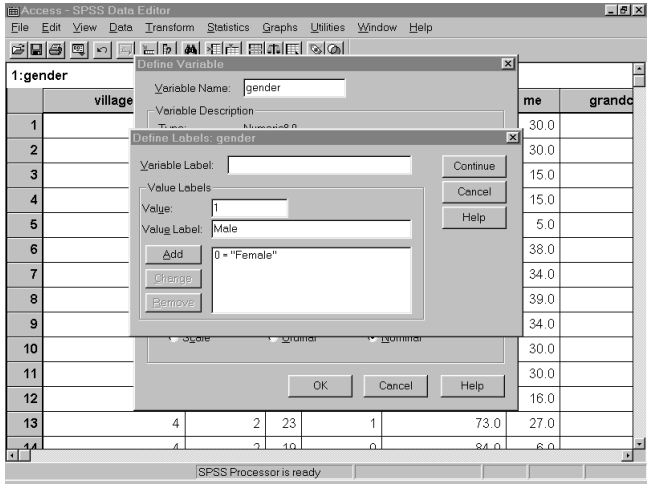


FIGURE 11. Defining labels for gender on inter-generation data

Enter the demographic label (e.g., Female, in the example above) and when you finish, click **Continue**.

Follow the same procedure for the other demographic data.

- **Checking data distribution** — Commonly, we calculate a *mean/average* to measure the distribution of pebbles according to the respondents. However, using the mean has some disadvantages when the data are not spread symmetrically or when there are one or more outliers. If this occurs, we can use *median value* instead. To check whether the data are spread symmetrically or not we can use a *boxplot*. We check the data for each generation separately, since each generation was considered as a separate variable. If the majority of the generations were distributed asymmetrically, we would use the median instead. On the other hand, if the majority of generations was distributed symmetrically, we could use the *mean/average*.

You can obtain a boxplot by following these instructions:

From menu, choose **G**raphs
Bo**x**plot

Choose *simple* and *summaries for separate v**a**riables* (Figure 12).

Click on **Define** button.

Enter all the generations as variables (Figure 13) and then click **OK**.

You will get the Boxplot as shown in Figure 14. The black line inside the box shows the median position. The upper and lower edges of the box show the position of the 3rd and 1st quartiles, respectively. The data are distributed symmetrically if and only if the median is located in the middle of the box, at approximately the same distance from both the 3rd and 1st quartiles.

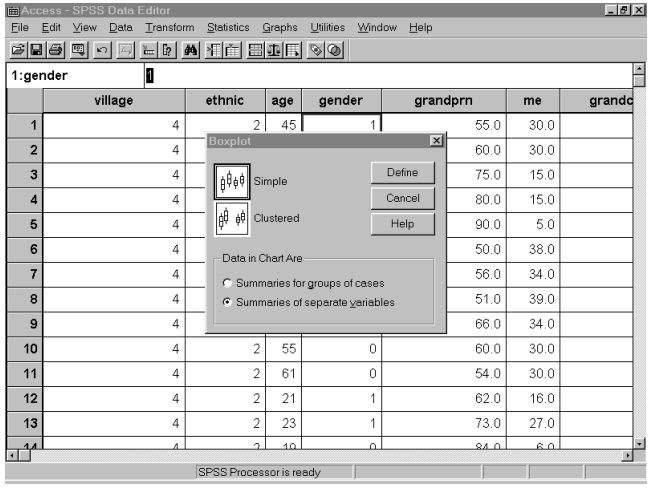


FIGURE 12. Choosing type of boxplot for inter-generation data

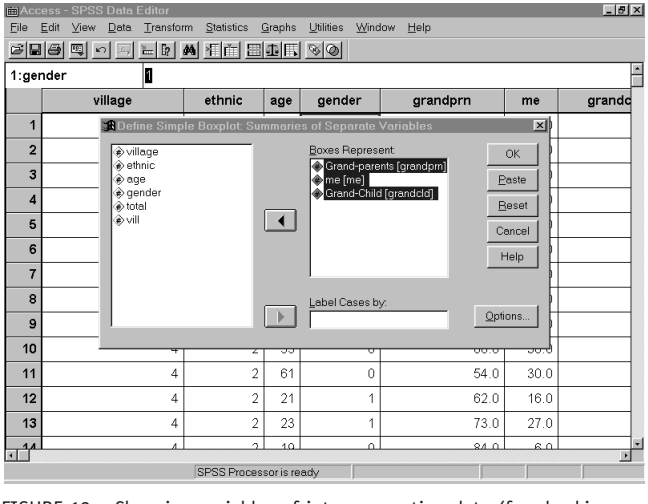


FIGURE 13. Choosing variables of inter-generation data (for checking data distribution)

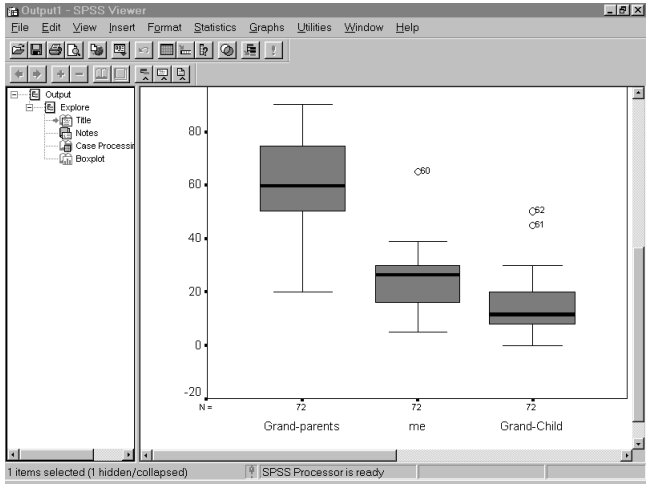


FIGURE 14. Boxplot for inter-generation data

If the data distribution is highly skewed or there are some outliers (usually indicated by asterisks or circles), we suggest you use the median instead. In these boxplots above, we considered me and grandchild as having an asymmetric distribution since the median position is skewed and there are also some outliers. Hence, we use a median for analysing inter-generation data.

- **Generating the table that describes the distribution of access to resources across generation, according to all respondents**

From the menu, choose:

Statistics
Custom **T**ables
Basic Tables

Put all generations in *summaries* item. You can do this by clicking the variable name and then clicking on the arrow in front of the *summaries* box (Figure 15).

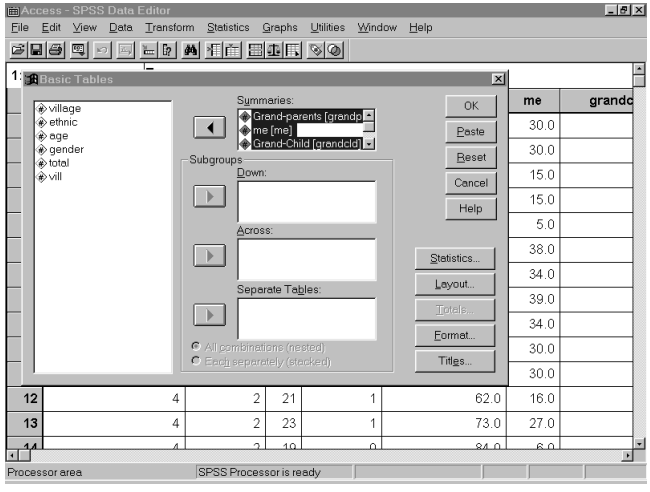


FIGURE 15. Selecting inter-generation variables

Click on the **S**tatistics button and add **M**ean or **M**edian (depend on which measure you decide to use) as the cell's statistics.

By clicking the **OK** button, you will get the results (Figure 16).

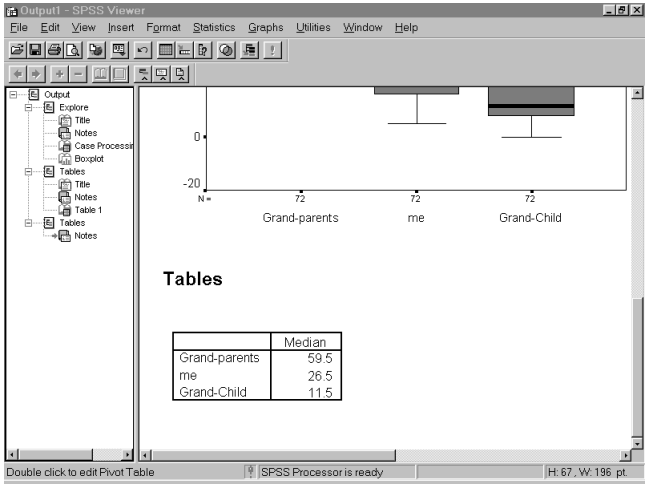


FIGURE 16. SPSS table describing distribution of access to resources across generation, according to all respondents

c. Conducting simple analyses using Excel

- **Generating the table that describes the pebble distribution, according to all respondents**

From the **D**ata Menu, choose **P**ivotTable Report. You will see the dialogue box as shown in Figure 17.

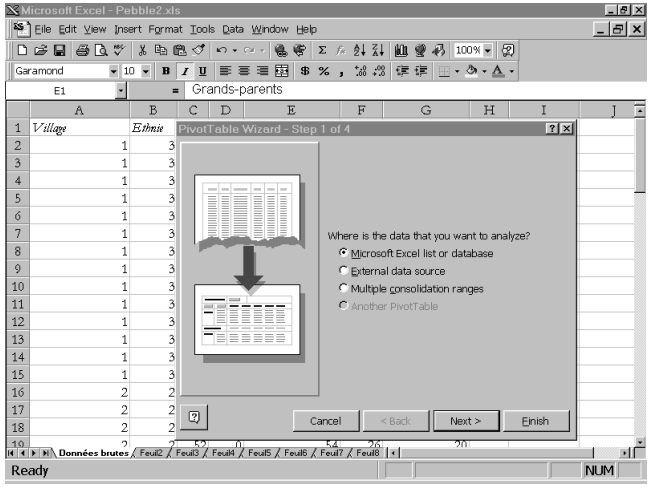


FIGURE 17. Pivot table wizard (inter-generation data)

Click on **Next>** button twice, and you will see a dialogue box as shown in Figure 18.

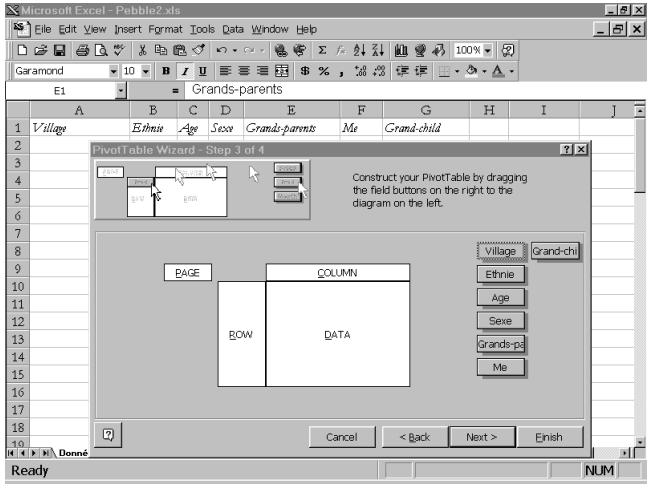


FIGURE 18. Constructing a pivot table for simple analysis of inter-generation data (Step 1)

Drag all the generations into the data space (Figure 19).

The sum is the default statistic. You can change the statistic by clicking on the button and choosing the statistic that you want to calculate[31] (Figure 20).

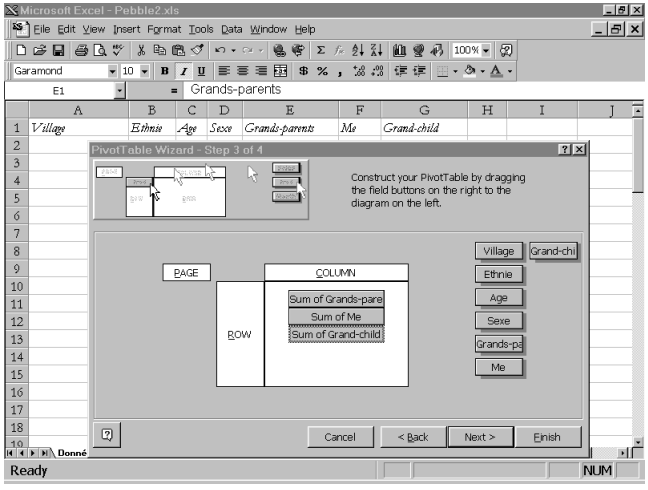


FIGURE 19. Constructing a pivot table for simple analysis on inter-generation data (Step 2)

31 The statistics provided by Excel do not include Median. If you want to use median, you must use another software, such as SPSS.

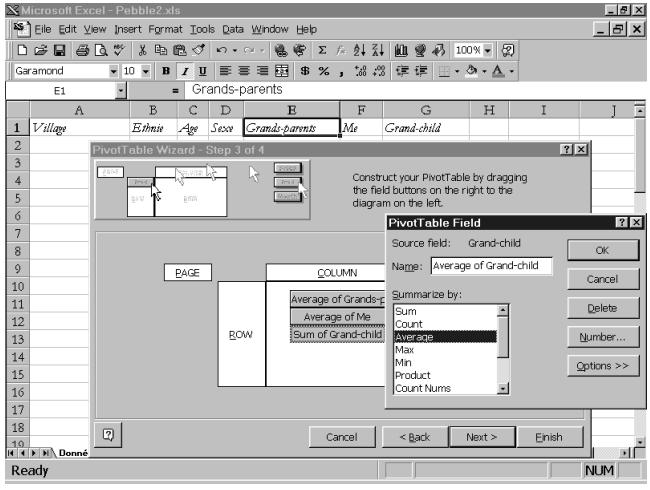
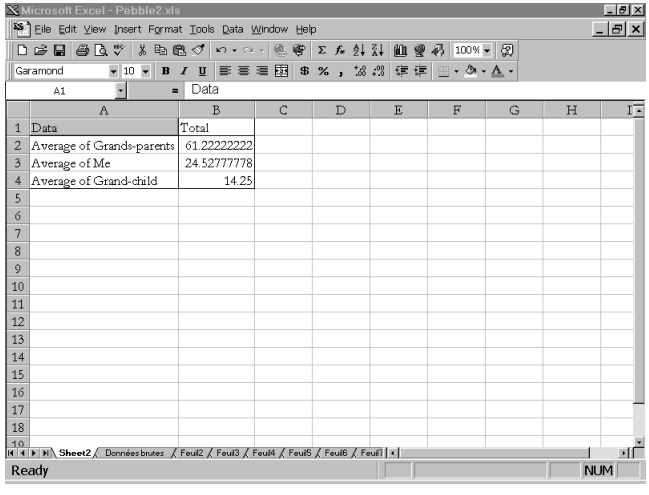


FIGURE 20. Changing the default statistic on inter-generation data

After you have decided on the appropriate statistic, you can click the **F**inish button and you will get the table as shown in Figure 21. If you prefer a smaller number of digits, you can set the decimal place[32].

32 To set the decimal place, go to the **F**ormat menu, choose cells and under category choose numbers. Then you can set the decimal place by increasing or decreasing the default decimal place (default=2).



	A	B
1	Data	Total
2	Average of Grands-parents	61.22222222
3	Average of Me	24.52777778
4	Average of Grand-child	14.25

FIGURE 21. Excel table describing distribution of access to resources across generation, according to all respondents

d. Conducting advanced data analysis using SPSS

- **Generating the table that describes the distribution of access to resources across generation, according to a particular subgroup**

  Sometimes, it is interesting to compare different subgroups within each demographic element (ethnicity, gender, educational level, etc.) according to their opinions on the distribution of pebbles among stakeholders. To compare those subgroups' opinions, we need to know the distribution of pebbles among generations, according to each subgroup. You can follow our instructions below to get those figures. Imagine that we have a set of respondents from three ethnic groups, *Baka*, *Kako*, and *Nzime*[33].

33 These ethnic groups inhabit the Dja Reserve area, Cameroon (Tchikangwa *et al.*1998).

From the menu, choose: **S**tatistics
Custom **T**ables
Basic Tables

Put all generations in *summaries* item. You can do this by clicking the variable name and then clicking on the arrow in front of the *summaries* box (Figure 22).

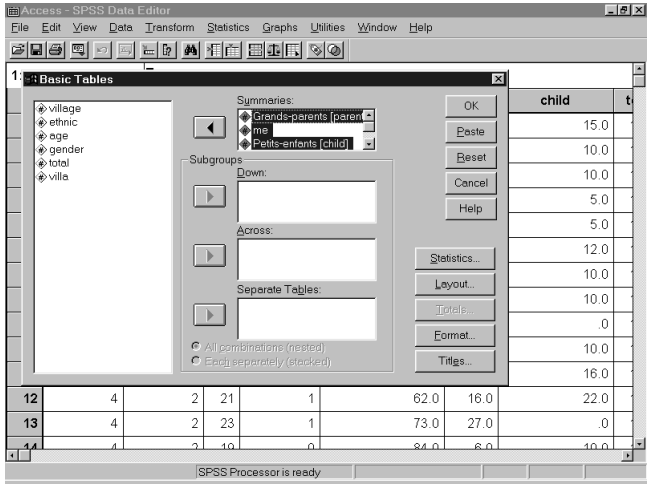


FIGURE 22. Selecting inter-generation variables (for summarising)

Put the column containing data of subgroups we want to compare into *subgroups-down* item (Figure 23). In our case, information about ethnicity is in the *Ethnic* column. Then, click **OK**.

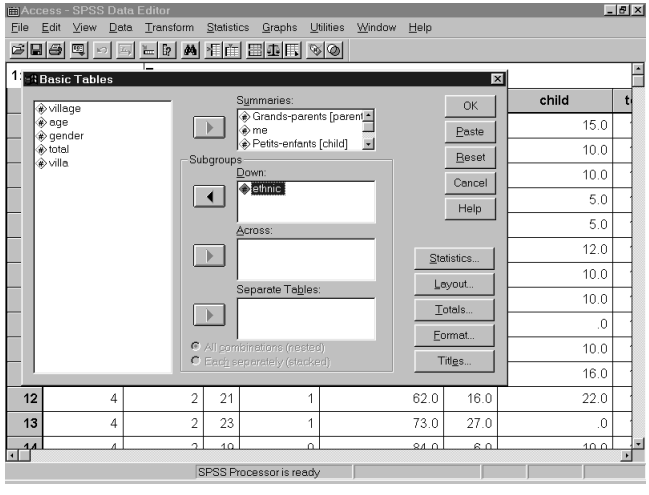


FIGURE 23. Selecting demographic variable on inter-generation data

Now, we can compare the differences of pebble distribution for each generation according to the three ethnic groups (Figure 24).

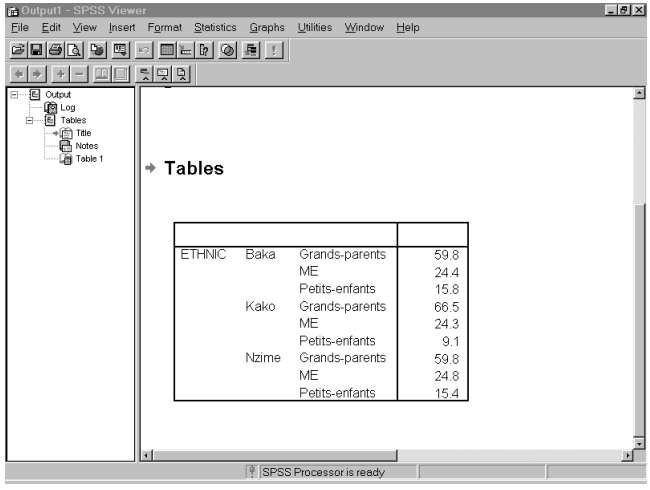


FIGURE 24. SPSS table describing comparison of distribution of access to resources across generation based on ethnicity

If you have more than three categories in your demographic data (e.g., four or more ethnic groups), you can continue this same step to get the *mean/median* value of pebble distribution among stakeholders for the other categories.

- **Testing agreement/disagreement among different groups on the distribution of access to resources across generation—** You can check whether there is agreement or disagreement among the different groups[34] about the amount of pebbles attributed to each generation.

From the menu, choose: Statistics
Nonparametric test
2 independent samples or
K independent samples[35]
(See Figure 25)

34 Groups here could be gender, educational level, ethnicity, etc.

35 If you would like to test data with only two categories or two groups, you should choose two independent samples and then use the Mann-Whitney test. If you are dealing with data in more than two categories, you should choose K-independent samples, and then use Kruskall-Wallis test.

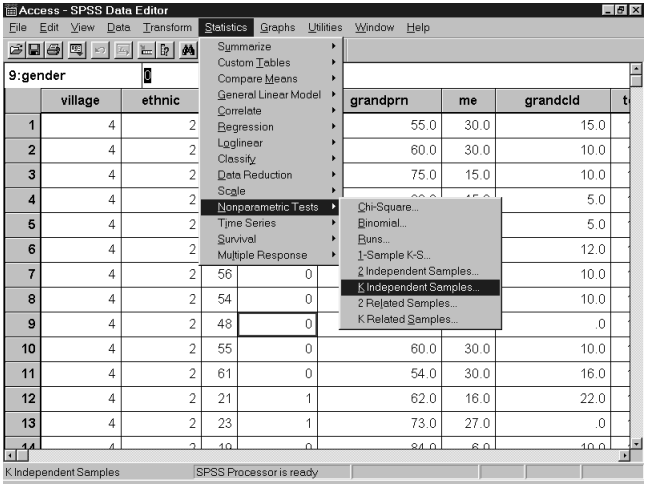


FIGURE 25. Selecting statistical test for testing agreement on inter-generation data

Put the test variables[36] into *test variable list box* and grouping variable into *grouping variable box* (Figure 26). You then should define a minimum and maximum value of the grouping variable (you can check the meaning of each code on your coding sheet). See Figure 27 for details.

36 Test variables should be the column containing the pebbles data for each generation. The grouping variable is the column containing information of different groups that we want to compare.

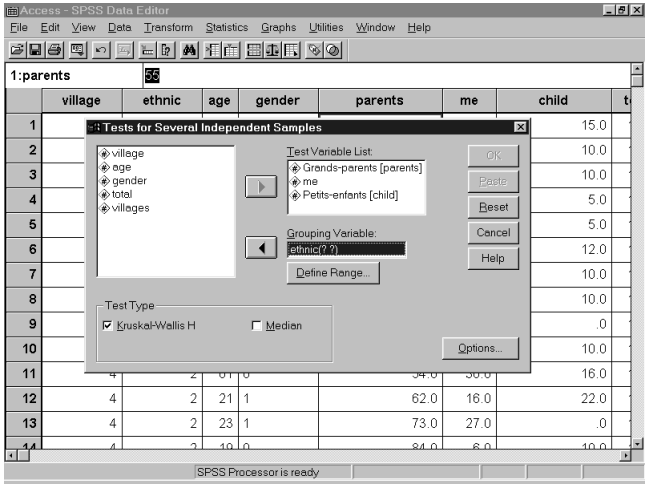


FIGURE 26. Selecting test variables and grouping variable (inter-generation data)

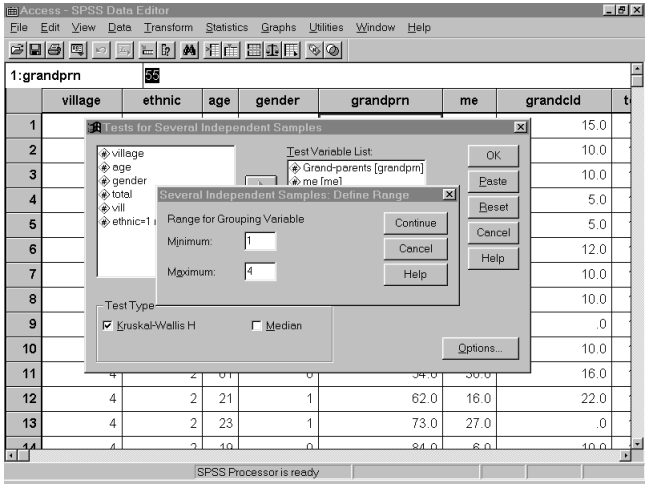


FIGURE 27. Defining the range for grouping (demographic) variable (inter-generation data)

Click **Continue** and then **OK**.

You will get the results as shown in Figure 28. The test will have to be carried out for each generation. Be sure to check the significance at each test.

See the P-value (indicated by the arrow). If the P-value falls below 0.05, we reject the null hypothesis[37] and conclude there is disagreement among the groups about the amount of pebbles belonging to each generation. If the P-value falls above 0.05 then we conclude there is agreement among the different groups. The test was carried out for each generation; be sure to check the test significance for each generation.

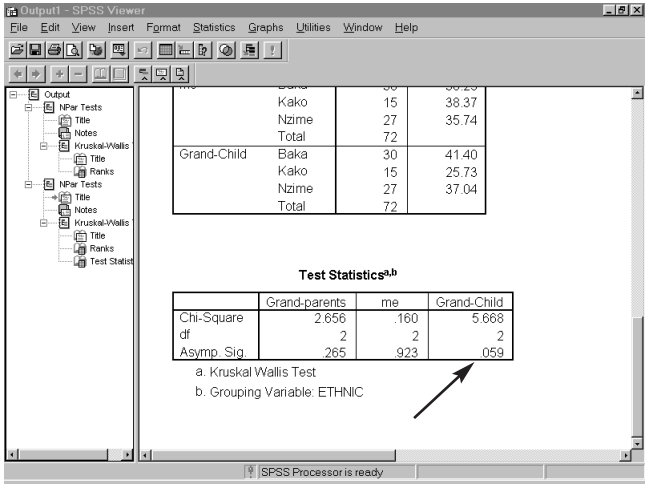


FIGURE 28. Output of the statistical test for each generation (inter-generation data)

37 The null hypothesis is that all the subgroups allocated the same amount of pebbles for each generation. In the other words, they agreed about the amount of pebbles attributed to each generation.

e. Conducting advanced data analysis using Excel

- **Generating the table that describes the distribution of access to resources across generation, according to particular subgroups.**

You can use a similar procedure to what you did to get a table of pebble distribution according to all respondents (Part c of this section). The differences will be when you construct the pivot table. Instead of just dragging all generations to the data space, you must also drag the column where the particular subgroups are into the column space[38] (Figure 29). Here we use ethnic groups as an example.

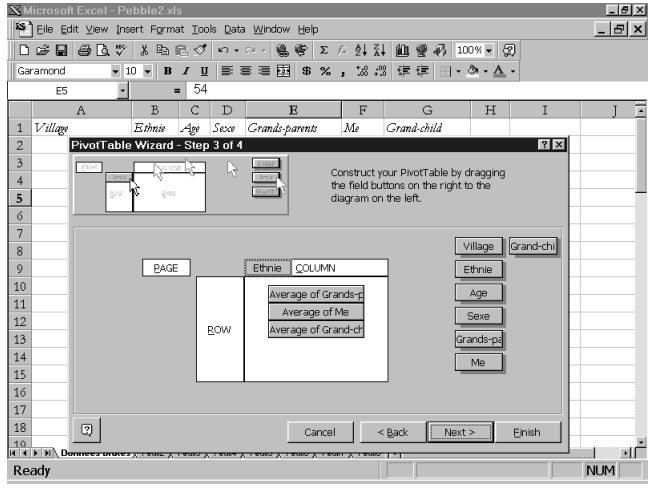


FIGURE 29. Constructing a pivot table for advanced analyses on inter generation data

38 Subgroups here can be gender, educational level, age, ethnicity, etc.

By clicking the **Finish** button you will get the result (Figure 30). Notice the two subtables for each ethnic group (in this example, as in the previous one, ethnic1 means *Baka*, ethnic2 means *Kako,* while ethnic3 means *Nzime*).

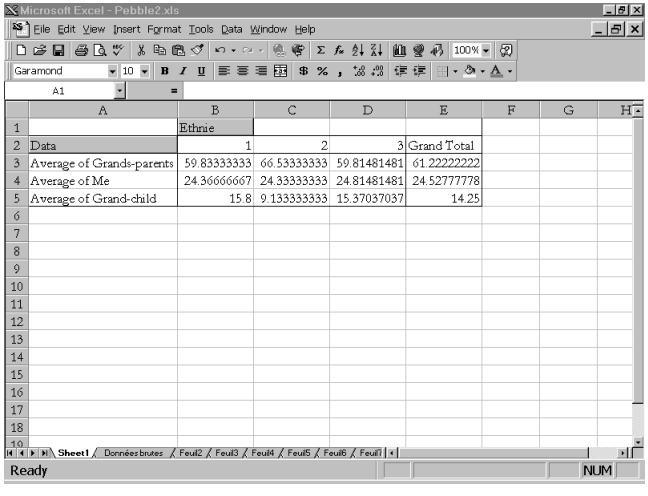


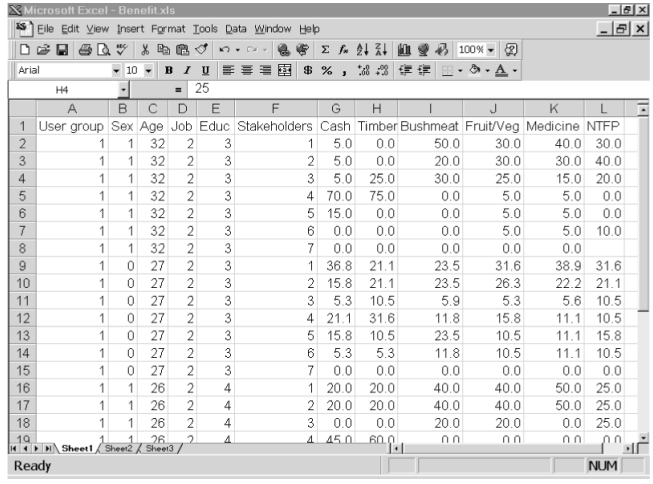
FIGURE 30. Excel table describing comparison of distribution of access to resources across generation based on ethnicity

3. BENEFIT SHARING AMONG STAKEHOLDERS: GUIDELINES

This section provides information on how to set up a spreadsheet for the pebble distribution method on benefit sharing among stakeholders (method described in *The Grab Bag*), and how to enter the data. Next we explain how to perform simple statistical analyses on the data, mainly how to see how the data are distributed and how to generate simple descriptive tables. For those of you who are satisfied with simple analyses, you can stop there. Finally, we provide help for those who want to have more advanced analysis such as testing the significance of opinion differences among different groups (based on ethnicity, age, gender, etc.).

a. Entering the data

- Enter the data using spreadsheet software, such as Excel or Lotus.
- Enter the demographic data, such as user group, gender, age, educational level, etc., in the left-hand columns (Figure 31, columns A–F). Insofar as possible, enter the demographic data using numeric codes. If you have categorical data, such as gender, educational level, etc., you need to convert them into numeric codes (e.g., female = 0, male = 1). It is very important to keep the coding sheet, so you will know the meaning of each code. We recommend you also store the coding information somewhere in the spreadsheet.
- Use one row for each stakeholder. For example, in Figure 31, we include a data set from Long Segar where we identified 7 stakeholders. You should use rows 2–8 (7 rows) for respon-



	A	B	C	D	E	F	G	H	I	J	K	L
1	User group	Sex	Age	Job	Educ	Stakeholders	Cash	Timber	Bushmeat	Fruit/Veg	Medicine	NTFP
2	1	1	32	2	3	1	5.0	0.0	50.0	30.0	40.0	30.0
3	1	1	32	2	3	2	5.0	0.0	20.0	30.0	30.0	40.0
4	1	1	32	2	3	3	5.0	25.0	30.0	25.0	15.0	20.0
5	1	1	32	2	3	4	70.0	75.0	0.0	5.0	5.0	0.0
6	1	1	32	2	3	5	15.0	0.0	0.0	5.0	5.0	0.0
7	1	1	32	2	3	6	0.0	0.0	0.0	5.0	5.0	10.0
8	1	1	32	2	3	7	0.0	0.0	0.0	0.0	0.0	
9	1	0	27	2	3	1	36.8	21.1	23.5	31.6	38.9	31.6
10	1	0	27	2	3	2	15.8	21.1	23.5	26.3	22.2	21.1
11	1	0	27	2	3	3	5.3	10.5	5.9	5.3	5.6	10.5
12	1	0	27	2	3	4	21.1	31.6	11.8	15.8	11.1	10.5
13	1	0	27	2	3	5	15.8	10.5	23.5	10.5	11.1	15.8
14	1	0	27	2	3	6	5.3	5.3	11.8	10.5	11.1	10.5
15	1	0	27	2	3	7	0.0	0.0	0.0	0.0	0.0	0.0
16	1	1	26	2	4	1	20.0	20.0	40.0	40.0	50.0	25.0
17	1	1	26	2	4	2	20.0	20.0	40.0	40.0	50.0	25.0
18	1	1	26	2	4	3	0.0	0.0	20.0	20.0	0.0	25.0
19	1	1	26	2	4	4	45.0	60.0	0.0	0.0	0.0	0.0

FIGURE 31. Spreadsheets format for forest benefits data

dent1, rows 9–15 (7 rows) for respondent2 and so on. You need one column in which to enter the stakeholder's code[39], indicating which rows belong to which stakeholders. Column F in our example above contains the stakeholder's code.

- In the next column, you can enter the amount of pebbles for each stakeholder allocated by the respondents. You need one column for each benefit. In our example above, we use column G for cash, column H for timber, and so on (Figure 31).
- After you finish entering the data for all respondents, do not forget to save the file into Lotus 123 format (*.wk1).

b. Conducting simple data analyses using SPSS[40]

We use SPSS (Statistical Package for Social Sciences) to analyse the data.

- **Opening the data —** You can open the Lotus 123 file in an SPSS spreadsheet. Use the *read variable name*[41] option.
- **Labelling the data —** Use the coding sheet that you made when you coded the demographic data into a numeric code to label the data. For labelling the data, place the cursor at the column where the demographic data are. Here, we give an example to label educational level data.

39 Each stakeholder was coded using a number. If we have *n* stakeholders, we can assign a unique number to each stakeholder arbitrarily. But we suggest you code the stakeholders in a mnemonic fashion that will make your analysis easier. You might, for example, assign 1 to local people, 2 to contractors and traders, and 3 to government and timber concessions. Your coding scheme will depend on your planned use.

40 We also made guidelines for those of you who want to carry out simple analyses using Excel (see Part c of this section).

41 From the SPSS menu bar, you can choose **F**ile → **O**pen, then change the file type to Lotus (*.wk1) and choose the file you want to open. The SPSS processor automatically will ask you whether you want to read the variable name or not. Just click on the '*read the variable name*' option.

From the menu, choose: **D**ata
Define variable
(Figure 32).

Click **Label** button (shown by the arrow). You will be asked to enter your label (Figure 33).

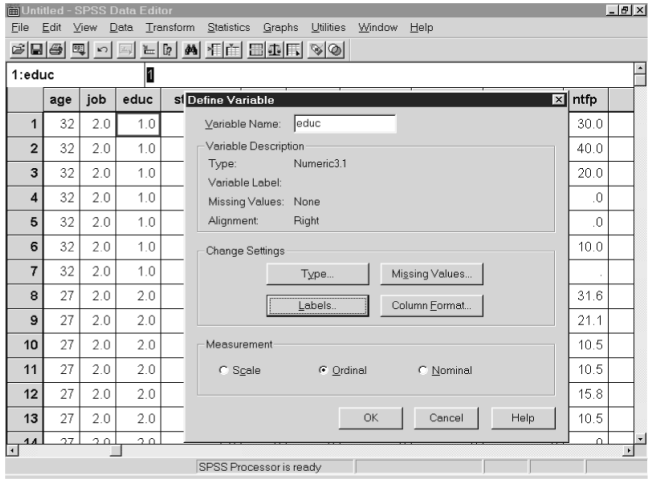


FIGURE 32. Label button (forest benefits data)

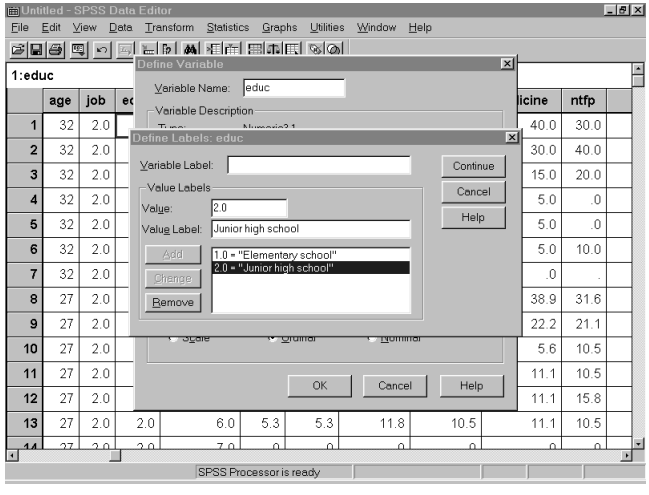


FIGURE 33. Defining labels for educational level on forest benefit data

Enter the demographic label (e.g., junior high school, in the example above) and when you finish click **Continue**.

Follow the same procedure for the other demographic data.

- **Checking data distribution**

  Commonly, we calculate a *mean/average* to measure the distribution of pebbles according to the respondents. However, using the mean has some disadvantages when the data are not spread symmetrically or when there are one or more outliers. If this occurs, we can use the median value instead. To check whether the data are spread symmetrically or not, we use a *boxplot*. With this method we check the data for each benefit separately, because those benefits were considered as separate variables. If the majority of the benefits were distributed asymmetrically, we would use a median instead. On the other

hand, if the majority of benefits were distributed symmetrically we should use the *mean/average*.

You can obtain a *boxplot* by following these instructions:

From the menu, choose: Graphs
Boxplot

Choose *clustered* and *summaries* for separate variables (Figure 34).

Click on **Define** button.

Enter all the benefits as variables and *stake*[42] as the category axis (Figure 35).

You will get the *Boxplot*, as shown in Figure 36. The black line inside the box shows the median position. The upper and lower edges of the box show the position of the 3rd and 1st quartiles, respectively. The data are distributed symmetrically if and only if the median is located in the middle of the box, at approximately the same distance from both the 3rd and 1st quartiles.

42 *Stake* is the column containing the stakeholder's code. In this example, we use *stake*, an abbreviation of **stakeholders**, to name the column. You can use a different name, of course.

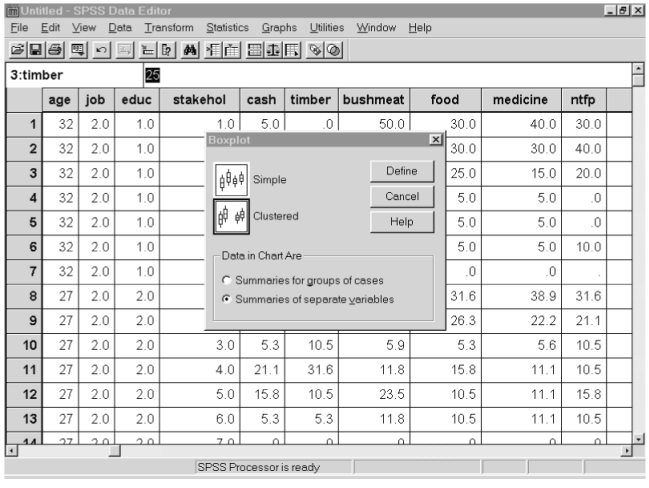


FIGURE 34. Choosing type of boxplot for forest benefits data

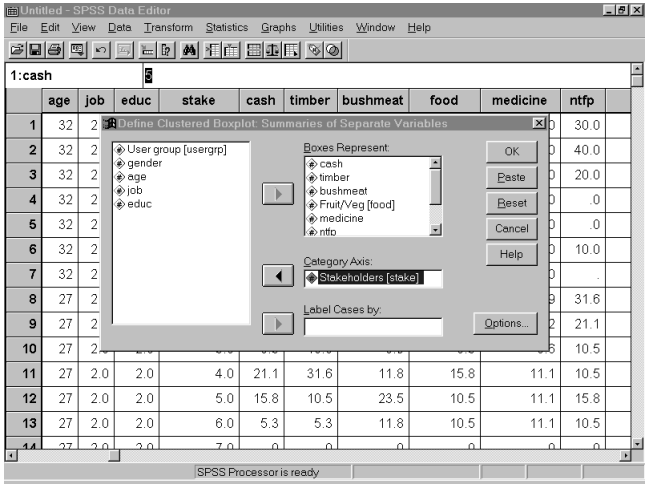


FIGURE 35. Choosing variables of forest benefits data (for checking data distribution)