

Case Study 11: Analysis of cattle production system data from the Swaziland Farm Animal Genetic Resources Survey

Moses Mefika Sithole¹ and Dorah Vilakati ²

¹Department of Agricultural Economics and Management, University of Swaziland, P.O. Luyengo, Swaziland

²Ministry of Agriculture and Cooperatives, Swaziland Government, P.O. Box 162 Mbabane, Swaziland

Contents

- Summary
- Background
- Research Strategy
- Objectives
- Study design
- Questions to be addressed
- Source material
- Data management
- Exploration & description
- Statistical modelling
- Findings, implications and lessons learned
- Reporting
- Study questions
- Related reading
- Acknowledgements

Summary

This case study describes the design of a questionnaire survey to assess the importance of cattle to subsistence farms (homesteads) of the Hhohho region in Swaziland and the methods of cattle production under which they are raised. These data represent a small subset of data taken from a large livestock breed survey carried out in Swaziland on cattle, sheep, goats, donkeys, pigs and chickens. One of the main purposes of the case study is to demonstrate methods for handling data from surveys and in particular the application of methods of analysing multiple response data. The statistical package INSTAT is used as well as GenStat.

Data exploration using tables and graphs, keeping in mind the multiple response nature of the data, are first undertaken to determine the general data patterns, trends and distributions. To confirm the exploratory results, further data analysis is carried out using an index method (based on ranking) to assess the importance of cattle to the homestead and the purposes of raising cattle.

In cases where conclusions are not obvious, Pearson's chi-square test is applied to test the

differences in proportions of homesteads giving different responses. REML is also used to compare average cattle numbers raised in homesteads in which livestock was either the primary source of income or not.

Finally, aspects of methods of reporting of results are briefly described in terms of the different types of people (from researchers to farmers) who will read the information.

Background

A lack of knowledge of the physical and performance characteristics of indigenous farm animal genetic resources in countries in sub-Saharan Africa and the extent of existing genetic diversity has, through pressures to increase production, led to the underutilisation, dilution and replacement of these resources through crossbreeding with 'specialised' exotic breeds. In order to obtain better knowledge of existing livestock breeds and the management thereof, FAO, together with ILRI, planned a series of surveys to gather this information across the Southern African Development Co-operation (SADC) region.

The overall objectives of the surveys were to obtain estimates of livestock population size and distribution of farm animal breed resources, as well as to determine management/production and socio-cultural practices employed by farmers in raising these animals. The plan was to use such surveys as a platform for subsequent simple, regular updating of breed information. With the information gained the idea was that countries would be able to:

- develop comprehensive plans for the management of farm animal genetic resources,
- develop and harmonise support policies for their management,
- facilitate development of appropriate animal recording systems and sustainable breeding programmes,
- facilitate development and implementation of relevant conservation activities.

A questionnaire for a Farm Animal Genetic Resources Survey was designed for preliminary evaluation in Zimbabwe. Other methodological aspects of sample design and survey implementation were also tested (Rowlands et al., 2003). A computer database system Breedsurv using Microsoft Access for data capture and data management was also developed by ILRI.

With the experiences gained from Zimbabwe recommendations were then made to the other SADC countries, namely, Angola, Botswana, Democratic Republic of Congo, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania and Zambia for surveys to be implemented in each of them.

The surveys generally focused on six livestock species, cattle, sheep, goats, chickens, pigs and donkeys. Hence, six sets of questionnaires were designed for the survey, one set for each species. The questionnaires were designed to capture household information, information on production systems, breeding, mating, disposal/purchasing and breeding strategies, health, age structure and breed characteristics.

One species was chosen as the primary species for which all questions were asked of a homestead. Data just on age structure and breed characteristics were also requested on either one or two other species.



Source: Thulile Sgwane

As well as the surveys conducted in other SADC countries a breed livestock survey has since also been carried out in Ethiopia using a similar survey design. The results have been reported by Ayalew et al. (2004). The photographs below show some of the cattle surveyed and enumerators discussing questionnaires between household visits.



Source: Anette van Dorland Source: Anette van Dorland

The questionnaires used for the SADC surveys have also been adapted for a survey of sheep in parts of Kenya (Kosgey, 2004)



Source: Isaac Kosgey

Research Strategy

This case study describes certain aspects of the Farm Animal Genetic Resources Survey carried out in Swaziland, using just results collected in one region (Hobbo) for cattle. The design of the survey followed largely that developed in Zimbabwe. Before embarking on the breed survey in Swaziland, however, the various considerations, as outlined by Rowlands et al. (2003), that need to be taken into account in planning a survey were considered. Realising the amount of work to be involved in such a survey certain fundamental questions were addressed:

- Is the survey approach the appropriate instrument to be used?
- If so, does the whole country need to be surveyed or only certain regions?
- For which livestock species are data required?
- Which of the questions provided in the Zimbabwe questionnaire are required?

and so on.



Source: Thulile Sgwane

Furthermore, if a survey were to be carried out:

- What type of sampling frame should be adopted?
- What forms of stratification might be appropriate?
- What proportion of homesteads needs to be sampled from the population?
- Is there census information on the number of homesteads within the population from which the sample would be drawn that can help with the design of the survey?

After due consideration it was decided to follow broadly the approach done in Zimbabwe and undertake a survey throughout Swaziland for all six species: cattle, sheep, goats, pigs, chickens and donkeys



Map showing the location of Swaziland together with the national flag

Map of Swaziland showing the Hhohho region.



Objectives

The overall objectives of the livestock breed survey were to:

- Document existing farm animal genetic resources for cattle, sheep, goats, donkeys, pigs and chickens within Swaziland.
- Characterise available indigenous livestock information to quantify the extent of breed diversity within Swaziland.
- Determine breed status and trends within the country
- Estimate population numbers for different breeds.
- Summarise performance characteristics.
- Use the results to contribute to improved use and conservation of indigenous breeds.



Source: Thulile Sgwane

Questionnaires were designed to take account of these overall objectives and to more specifically collect data to:

- Determine why farmers keep livestock.
- Describe methods of management of livestock including housing, types of feeding and responsibilities of different members of the household for different livestock activities.
- Assess levels of animal health and prevalent diseases affecting livestock as perceived by farmers.
- Ascertain annual levels of mortality and reasons for culling.

- Collect information on breeding methods.
- Assess the important qualities of different breeds as perceived by farmers.
- Collect phenotypic data for different breeds.

The present case study illustrates methods of handling some of the management data associated with the second sub-objective for cattle. Case Study 12, on the other hand, explores methods of estimating population size.



Source: Thulile Sgwane

Study design

The population for the survey was defined as those homesteads that kept livestock (180,000) in communal Swazi Nation Land (SNL) and commercial livestock farms (235). However, for the purpose of this case study, only the sample of homesteads from the communal land is used.

Stratification of the population of homesteads was designed to make sampling at the upper layers representative of the varying conditions within the country. This was done, on the basis of the country's four administrative regions - Hhohho, Manzini, Shiselweni and Lubombo (Primary strata). These regions represent distinct socio-economic patterns and modes of livelihood.

Within each of these regions further classification was done on the basis of all the available 28 livestock sub-regions (Secondary strata). The numbers of sub-regions in the Hhohho, Manzini, Shiselweni and Lubombo regions were 6, 7, 7 and 8, respectively, and each sub-region was sampled



Source: John Rowlands

Sampling was then done in two stages of cluster sampling. In the first stage, purposive sampling of dip-tank areas (Primary sampling units - PSUs) was done within each sub-region. These were selected so that the different sections of the sub-region were sufficiently represented in terms of ecological zone, proximity to commercial farms, proximity to urban areas, remoteness and livestock density. Remoteness was an important criterion for capturing breeds that might be present only in a particular area.

The numbers of dip-tank areas selected within a sub-region depended on the total number of dip-tank areas in that sub-region; in general, the more dip-tank areas (reflecting higher livestock numbers) the larger was the sample.

Taking into consideration cost, human resource and time, it was decided to sample a total of 60 dip-tank areas in each of the four regions, i.e. a total of 240 from the 430 dip-tanks in the SNL. The plan was then to divide them approximately proportionally between the various sub-regions whilst also taking into account livestock density, ecological zone, proximity to commercial farms, etc.

Sub-region	Dip-tank areas for Hhohho Region	
	Population	Sample
Sub-region1	13	6
Sub-region2	12	7
Sub-region3	20	15
Sub-region4	28	20
Sub-region5	13	7
Sub-region6	14	5

Dip-tank areas were used as a cluster in the sampling frame because they form an important basis of the livestock administrative structure and this is where the majority of the government's livestock departmental field staff is based.

In the second stage of the sampling procedure, homesteads (Ultimate sampling units - USUs) were randomly sampled within selected dip-tank areas.

In order to adequately cover each dip-tank area, it was decided to sample as far as possible 10 homesteads per dip-tank area (although some problems did arise) to give a total sample of 2,400 homesteads countrywide.

This sample size (approximately a 1.3% sampling fraction) is comparable to that normally used by

Homesteads with livestock in Sub-region 3		
Dip-tank area	Population	Sample
Dip-tank31	197	9
Dip-tank32	222	8
Dip-tank33	141	8
Dip-tank34	140	8
Dip-tank35	198	14
Dip-tank36	222	11
Dip-tank37	144	9
Dip-tank38	171	8
Dip-tank39	279	12
Dip-tank310	180	12
Dip-tank311	238	3
Dip-tank312	214	10
Dip-tank313	116	5
Dip-tank314	249	11
Dip-tank315	98	6

the Central Statistical Office in agricultural surveys for the country.

Total	2809	134
Total for Hohho Region	9047	456

Some questionnaires had to be discarded due to incomplete or missing information.

Sometimes prior information on the species kept by a particular homestead was found to be incorrect. Thus, on reaching a homestead, supposedly keeping pigs and the species that had been pre-selected as the primary one for the survey at this particular homestead, the enumerator may have found that the homestead only kept goats.

More than 10 homesteads were sometimes interviewed in some dip-tank areas. Use of the modified random walk method occasionally led to additional homesteads being interviewed.

Homesteads with livestock in Sub-region 3		
Dip-tank area	Population	Sample
Dip-tank31	197	9
Dip-tank32	222	8
Dip-tank33	141	8
Dip-tank34	140	8
Dip-tank35	198	14
Dip-tank36	222	11
Dip-tank37	144	9
Dip-tank38	171	8
Dip-tank39	279	12
Dip-tank310	180	12
Dip-tank311	238	3
Dip-tank312	214	10
Dip-tank313	116	5
Dip-tank314	249	11
Dip-tank315	98	6
Total	2809	134
Total for Hohho Region	9047	456

Questions to be addressed

The questions to be addressed by the case study are as follows:

- What is the distribution of household heads by age and gender?
- What importance do livestock play in the homesteads and how important are cattle?
- What are the primary purposes for keeping cattle?
- Which members of homesteads own cattle, and who are responsible for different cattle activities?
- Does cattle ownership among members of the homestead vary according to the gender of the head of the homestead?

Initially, a summary of the data is presented showing the number of homesteads sampled, the number of homesteads that kept cattle and the number of these recorded with cattle as the primary species.

The other questions are then addressed through exploratory and descriptive methods with frequencies using pie charts, bar charts and tables followed by the application of statistical tests as necessary.

Multiple response data are also summarised through the calculation of indices (based on

relative rankings) that provide further information on the relative importance of livestock, crops, home industries and salary/wages to the income of the homesteads, and also the purposes for which cattle are kept.

Source material

The data set for the Hhohho Region, from which data have been selected for analysis for this case study, is held in Breedsurv - a computer data capture and storage system designed and written in Microsoft Access 2000 by ILRI specifically for SADC Livestock Breed Surveys (Rowlands et al. 2003). The system was designed so that data entry screens match as closely as possible the formats of the questionnaires. It also contains a number of data validation checks to limit entry of erroneous data.

Data were collected on the homestead, on the production system, on the different breeds of animals kept, their ages and sex, on aspects of animal health, animal movements, and so on.

For the purpose of this case study, however, only a proportion of the data recorded in relation to cattle production systems is used to demonstrate methods of analysis. These are those recorded on Page 2c of the questionnaire CS11Quest2; the remaining data are recorded on Page 3c CS11Quest3. Cattle production system data are stored in a Breedsurv table called PRODCAT. This table is linked to the HOUSEHOLD table, which contains data from the title page of the questionnaire and from Page 1c CS11Quest1 which gives general information recorded on the household (homestead).

Sub-region, dip-tank area and homestead number appear in the data base as Ward, Village and Hhold, respectively. This was to provide 'generic' names that could be used across the majority of SADC countries. These are defined as index or key variables to provide the link between the Microsoft Access HOUSEHOLD and the PRODCAT tables.

Data management

Data files

For the storage, manipulation and presentation and analysis of the data we need a combination of database, spreadsheet and statistical computer software packages. Microsoft Excel has been used for some of the manipulation of the data, and GenStat and Instat for data description and analysis.

Access to the Breedsurv HOUSEHOLD table can be made by opening Breedsurv, clicking the 'Database Window' icon on the Access Tool Bar, then clicking the 'Tables' tab and finally double clicking HOUSEHOLD. The same process can be repeated to retrieve the PRODCAT table.

Each table can be saved as an Excel file by clicking 'File', then 'Export', changing the file type to Excel and clicking 'Export'. To avoid confusion later, we have saved the worksheets under the same names as the tables, so that the HOUSEHOLD table becomes the CS11HOUSEHOLD worksheet, and the PRODCAT table becomes the CS11PRODCAT

worksheet

Descriptions of the variables stored in the HOUSEHOLD and PRODCAT tables in Breedsurv are contained within the Access table structure files and these have been copied into CS11HOUSEDoc and CS11PRODDoc, respectively. The names of the variables in these Excel files are the same as those found in the Breedsurv data system and are abbreviations for the original questions or types of responses (in the case of multiple response questions).

In order to facilitate use of the data just those variables explicitly used in this case study have been extracted and put into CS11Data1, CS11Data2a and CS11Data2b, respectively. Large- and small-scale commercial farms have also been omitted by using the Excel Filter command on the variable farm_type.

The data used in the PRODCAT file have been split in two, thus, CS11Data2a contains the data for Questions 3 to 4 on page 2c of the questionnaire and CS11Data2b contains the data for Question 5 (CS11Quest2). This is to make the files easier to handle. All three data files have already been opened in GenStat so that the variables to be used as factors in the analysis could be predefined, and, where appropriate (age and gender) labels attached to their integer codes (Spread → Factor → Edit Labels...).

Multiple response questions

Data management techniques, especially in survey analysis do not end with data file preparation. As will be seen under Exploration & description considerable skills in data management and data handling techniques are required in setting up data in appropriate formats for analysis. It is appropriate, however, to introduce the handling of multiple response questions here.

Multiple response questions can result in series of variables coded as 1 or 0, when just ticks or yes/no responses are required for a series of items in a list (see, for example Question 4 in CS11Quest2), or as 1, 2, 3 etc when a series of items is ranked (e.g. Question 8 in CS11Quest1). These questions are shown below.

CATTLE Production system page 2c

4. Members of household who own cattle (*Tick one or more*)

Head ☐

Spouse ☐

Head/spouse together ☐

HOUSEHOLD General Information page 1c

8. Sources of income
(*Tick first column as appropriate, rank level of source of income in second column – 1 highest.*)

1. Crops ☐ ☐

2. Livestock and products * ☐ ☐

3. Home industries ☐ ☐

Sons ☐ Daughters ☐ Others * ☐ * Describe _____	4. Salary / wages ☐ ☐ Other (specify) 5. _____ ☐ ☐ * Include the value of non-cash outputs or products e.g. manure, traction etc.
---	--

In contrast, single response questions can result in a simple yes/no or 0/1 answer (e.g. Question 7 in CS11Quest1). Selection of one item from a list of alternatives (e.g. Question 6 in CS11Quest2) also results in a single code to define the selected item. This can be an integer 1, 2, 3 etc corresponding to the position of the item in the list.

Where responses other than the ones explicitly itemised are allowed (as in Question 6) an extra text variable needs to be defined to store the specific response.

Different ways for coding multiple response data are discussed in the Instat software included within this CD.

It is important to highlight the proper way of handling data from multiple response questions; such data can often be mishandled. An error that sometimes happens is to code the answers to questions in a way that does not recognise the multiple levels in the data that result from the different responses allowed.

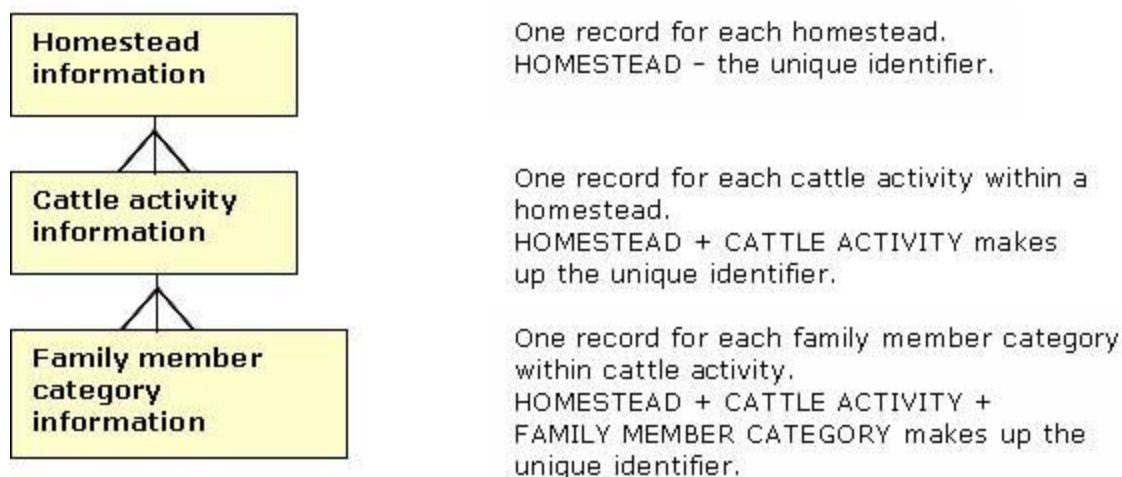
An example of multiple-level data in this case study is given by the responses required for cattle activities by different members of the homestead (Question 5 in CS11Quest2). Homestead can be thought of as being at the top of the hierarchy of responses, within each homestead one can think of there being different cattle activities, and within each cattle activity there can be different categories of members of the homestead responsible for that activity.

In a database that recognises this hierarchical data structure there would be a flat file for storing data about each homestead, such as contained in the HOUSEHOLD file CS11Data1. A second flat file could then be used to store the information on each cattle activity.

Rows of the cattle activity file would include a code to identify the particular cattle activity, while columns would correspond to information collected about the particular cattle activity.

Yet another flat file would be needed to store the information collected about the different family members involved in each cattle activity, and whether or not they were involved.

When these three flat files are considered together, they form a hierarchical structure, as illustrated.



The information at different levels of the hierarchy can be linked via index or key variables. Thus, in Breedsurv, three variables (WARD, VILLAGE and HHOLD) are used together to uniquely define the homestead and link the HOUSEHOLD and PRODCAT files. Use of index variables ensures efficient use of storage space (because data contained in the HOUSEHOLD file does not need to be repeated for each record in the PRODCAT file).

Recognition of the different layers in a hierarchy at which data are collected and the use of different files to represent those layers can also ensure that errors in the choice of denominator for the calculations of percentages and other statistics do not occur.

A hierarchy has not been established for homestead member within cattle activity in this case; both variables are stored together as 'flat' data in the PRODCAT file (CS11Data2b). This was because no additional information was collected about the different family members other than whether or not they were engaged in the activity. Some extra care may be needed, therefore, in analysing these data.



Source: Thulile Sgwane

As already mentioned, multiple response questions can be incorrectly analysed and interpreted when the wrong denominator is used for calculating percentage responses.

For example, if it is desired to compute the percentage of the homesteads in which boys are used to make dairy products, the researcher has two choices for the denominator.

- It could be the total number of homesteads.
- It could be only those homesteads which are involved in the activity.

The two results have different interpretations.

Alternatively, the researcher may wish to compute the number of times that boys are used to make dairy products as a percentage of the sum of times that all family member categories are involved in the activity. This results in a different denominator and a different interpretation.



Source: Thulile Sgwane

Exploration & description

Livestock species owned

First in the table below we provide the details of the number of homesteads sampled in the Hhohho region, the number of these homesteads that possessed cattle and the number selected for interview for cattle as the primary species.

	Number of homesteads	Percent
Number of homesteads sampled	456	100
Number keeping cattle	386	85
Number recorded with cattle as primary species	164	36

The number of homesteads sampled and the number selected for interview for cattle as the primary species correspond to the numbers of records in the HOUSEHOLD and PRODCAT files, respectively. Only those homesteads that were interviewed for cattle as the primary species provided data for the PRODCAT file. Thus, of the homesteads that kept cattle, 42% (164/386) were interviewed for cattle as the primary species.

Livestock species owned

Determination of the numbers of homesteads that kept each livestock species (Question 9 in CS11Quest1) can be obtained by analysing the answers in the form of a multiple response question. We have found that Intstat handles this relatively easily.

However, the data file contains the numbers of animals for each species that were recorded

in each homestead (see CS11Data1) rather than whether or not the homestead kept the species. Therefore, to produce multiple response tables, Instat requires new variables (factors) to indicate the presence (coded as 1, say) or absence (coded as 0, say) of each species in a given homestead.

In EXCEL, the filter option can be used to identify the homesteads that did not keep a given species, and hence also those that kept the species. New variables (Cattle, Sheep etc CS11Data1) can then be created indicating the presence or absence of the species.

Prior to doing this it was first important to verify that the numbers recorded for each species seemed reasonable. When the data were scanned it was noted that Homesteads 125 and 436 (see CS11HOUSEHOLD) had unusually high numbers of sheep. Hence for these homesteads the numbers of sheep have been marked as missing in CS11Data1. Similarly, homestead 370 had an unusually high number of chickens; this has also been marked as a missing value.

Note that in such situations where there are suspicions about the authenticity of certain data values and having first checked the raw data to see whether there may have been a mistake in data entry, any changes to the data file should always be documented and kept in what is sometimes referred to as an 'Audit trail' file. For the purpose of this case study a note has been made at the foot of CS11Doc1 - see below.

GenStat has a facility that defines values to be temporarily missing (**Spread → Column → Temporary missing values...**) but retains the value in the file. This can be useful during exploratory analysis to see what impact omitting a value may have on the analysis and saves preliminary editing of a data file and delays any necessity of documenting any alteration to the data.

Edited data

Values in the following records in CS11Data1 have been replaced by missing values:

Record	Homest	Variable	Deleted value
8	125	shp_no	555
254	436	shp_no	496
215	370	chk_no	2000

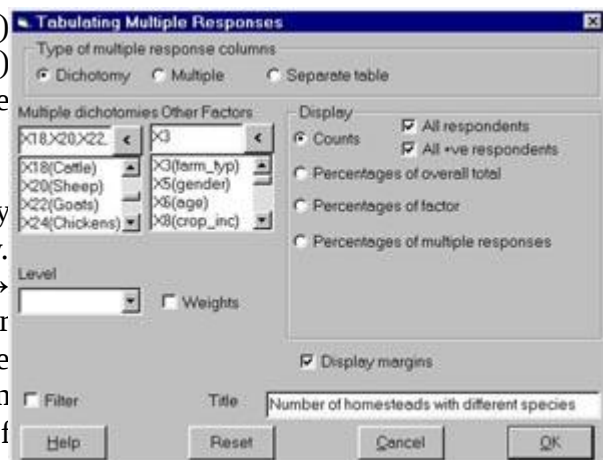
By using **Statistics → Tables → Multiple response...** to produce the dialog box below Instat can generate the multiple response table for the numbers of homesteads that kept the different species as shown below. By a slight modification of the dialog box the table of percentages of homesteads that kept the different species can also be obtained.

	farm_type			farm_type	
Summary	SNL	All	Summary	SNL	All
Cattle	85	85	Cattle	386	386
Sheep	15	15	Sheep	68	68
Goats	80	80	Goats	364	364
Chickens	91	91	Chickens	417	417
Pigs	29	29	Pigs	134	134
Donkeys	10	10	Donkeys	47	47
Othspec	0	0	Othspec	2	2
			Responding	456	456
			Respondents	456	456
All	311	311	All	1418	1418

The tables show that homesteads generally owned more than one species. Thus, the total percent of responses added to 312 (not 100 that would be the case if each homestead owned only one species) showing that homesteads owned on average three types of species.

Chickens were owned in the majority (92%) of the homesteads, followed by cattle (85%) and goats (80%). The other species were owned by fewer homesteads.

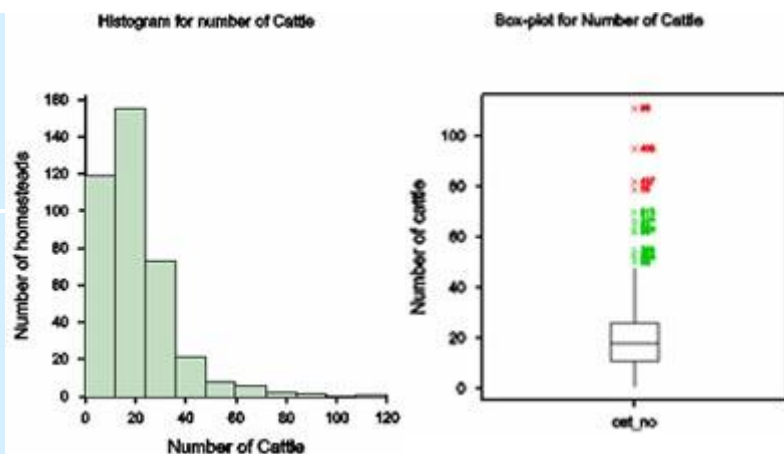
The same results can also be obtained simply by treating each response variable separately. Thus using **Stats** → **Summary Statistics** → **Frequency Tables** in GenStat for the factor 'Cattle' in CS11Data1 the number of the sampled homesteads that possessed cattle can be derived. Thus, again we see that 85% of the homesteads sampled kept cattle; the remaining 15% kept sheep, goats, etc., but not cattle.



Next, using GenStat, restricting cat_no to >0, then applying **Stats** → **Summary Statistics** → **Summarized Contents of Variates** and ticking in turn the 'Histogram' and 'Boxplot' boxes in the dialog box under 'Graphics', different presentations of summary statistics for the number of cattle kept by homesteads can be obtained. The aim is to look at the distribution of the number of cattle and see what implications it may have on the data analysis.

Summary statistics for cat_no
 Number of observations = 386
 Number of missing values = 0
 Mean = 20.63
 Median = 18.00
 Minimum = 1.00
 Maximum = 111.00
 Lower quartile = 11.00

Upper quartile = 26.00



The results show that the distribution of the numbers of cattle per homestead is skewed. From the box plot it can be seen that most homesteads possessed fewer than 25 cattle. However, a few homesteads had large numbers of cattle, some above 50 and up to 111. Hence, any statistical comparison with animal numbers may require a transformation to normalise the data.

Age and gender of head

Let us now look at some of the homestead information that may have a bearing on the management of cattle, e.g. gender and age of the head of homestead (Question 2 in

CS11Quest1). We shall present pie charts at the regional level and a bar chart at the sub-regional level; there are insufficient data to consider distributions further down the hierarchy at the dip tank level.

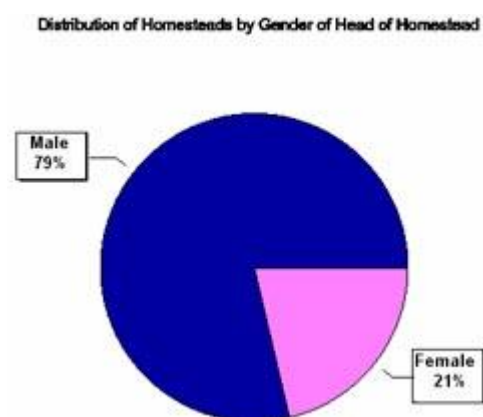
A pie chart can be created in Excel but we shall demonstrate here how this can be done in GenStat. To create the pie chart for the distribution of gender of heads of homesteads, we first need to run the Frequency Tables option on the variable 'gender' and store the frequencies in a spreadsheet. This is done via **Stats** → **Summary Statistics** → **Frequency Tables....** and clicking 'Display frequencies in spreadsheet'. This gives the frequency table:

One Factor Tables [Lo...]		
Row	gender	_counts
1	Female	97
2	Male	358

Next we must change the table to a vector so that the data contained in the table can be used as data to form a pie chart. This is done by right clicking 'gender' and selecting Convert in the menu that appears. We convert the table to a vector so that each column becomes a variable and also convert gender_1 to a text variable so that it can be used as a label in the pie chart.

Click Spread → Set as Active Work Sheet to ensure that GenStat recognises this as the only active sheet and then click Graphics → Pie Chart.... Type '_counts' under 'Data values:' and enter 'gender_1' under 'Labels' to form the pie chart.

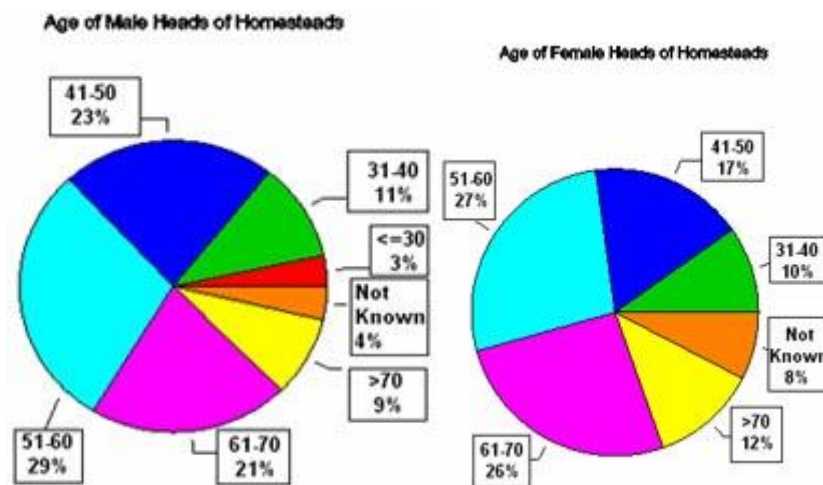
The results show that heads were male in the majority (79%) of the homesteads and female in the remaining 21%.



By using the Spread → Restrict/Filter command to select male and female homestead heads in turn we can similarly produced separate pie charts by age for each category.

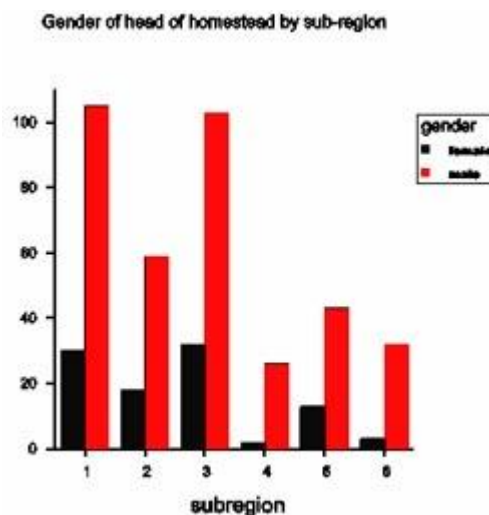
The male heads were distributed throughout all ages. In contrast, there were no female

heads 30 years old or younger. Young females are not likely to head a homestead, as it is customary for males to provide the headship, and it is not until the husband dies or gets a job away (perhaps in South Africa) that his wife assumes headship.



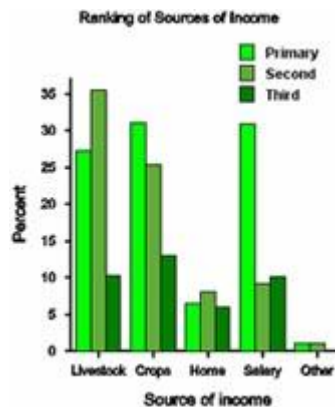
To create a bar chart to compare ratios of genders of homestead heads across sub-regions we start, with the original spreadsheet selected, by pressing Graphics → Create Graphs... → 2-way Barchart. Click 'Create Table' in dialog box that is shown, and then 'Groups' in the next small dialog box that appears, followed by 'OK'. Enter factors 'Subregion' and 'gender' under 'Selected Groups' and click 'OK'. The original dialog box then reappears with 'table' in the 'Data:' box. Click 'Next' and enter a title to produce the output as shown.

Across all the sub-regions, there were more homesteads with male heads of homesteads than there were with female heads, and this was perhaps more so for Sub-region4 and Sub-region6 than it was for the other sub-regions.

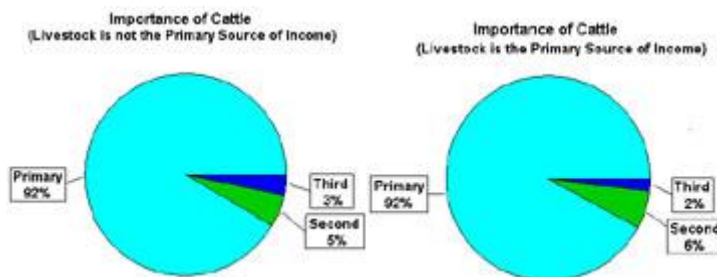


Importance of cattle

Using similar procedures in GenStat as described above, and with careful handling of the data, we can produce pie charts and bar charts to examine firstly the importance of livestock generally to the homestead and secondly the specific importance of cattle. The bar chart for the ranking of sources of income (Question 8 in CS11Quest1) shows that similar proportions of homesteads ranked livestock, crops and salary/wages as the primary sources of income.

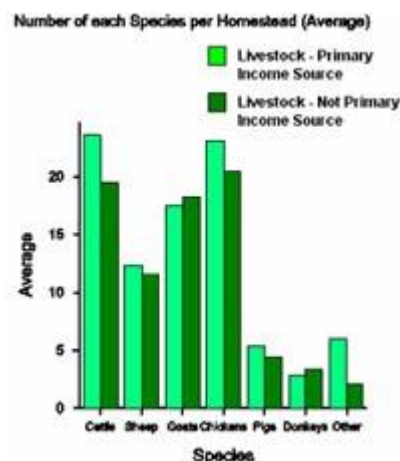


Regardless of whether or not livestock was the primary source of income, the pie charts show that over 90% of homesteads considered cattle as the most important species. In subsequent analysis, therefore, it looks as though we can generally ignore primary source of income in analysing aspects of cattle production.



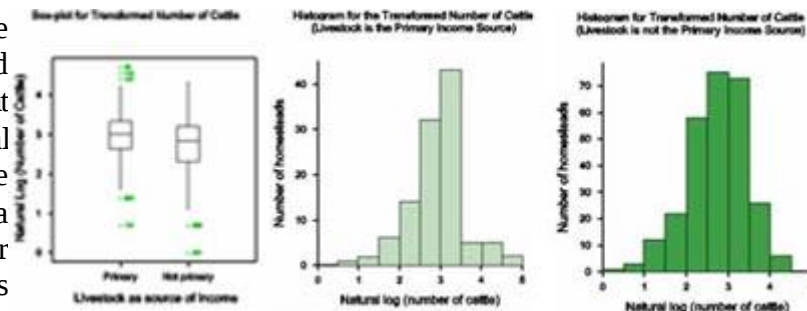
The bar chart for average numbers of different species kept by homesteads (for those homesteads that kept that species) showed that the numbers of cattle kept were as many as those of chickens. This seems a little surprising but, at the same time, indicates the overall importance of cattle to the homestead.

The bar chart also indicated that the homesteads, for which livestock provided the primary source of income, tended to have slightly higher average numbers of cattle, sheep, chickens and pigs than those where livestock did not provide the primary source of income.

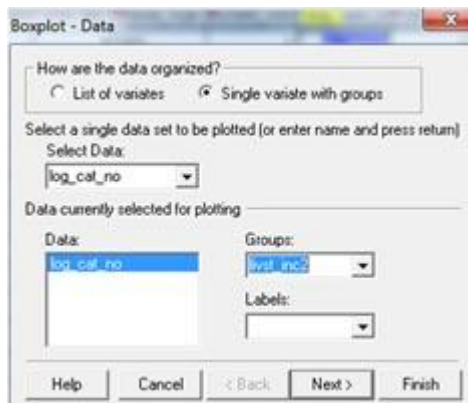


However, although to be expected, it is possible that the differences may not be statistically significant, but influenced (as we have already seen) by a few homesteads with large numbers of the species concerned. We therefore need to compare distributions after transformation of the data by a natural logarithm transformation stored in variable `log_cat_no`.

The histograms show the data still to be skewed slightly to the right but nevertheless more normal than those for the untransformed data (histogram shown earlier under Livestock species owned).



To investigate further the distribution of cattle numbers on the logarithmic scale we can generate a box-plot via Graphics → Boxplot...and complete the dialog box as shown below. The three outliers at the top all correspond to homesteads for which livestock provided the primary source of income, namely rows 85, 405 and 437 in CS11Data1.

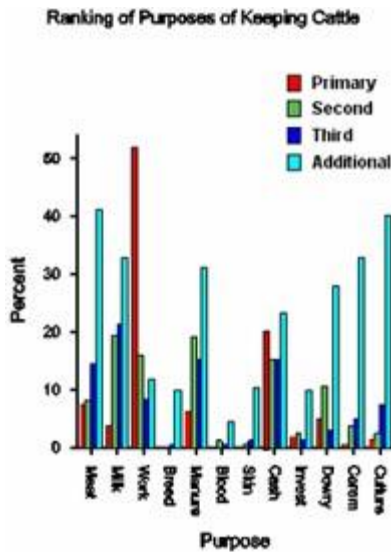


These data points may need further scrutiny, such as checking the questionnaires to verify that the numbers of cattle have been correctly entered. There are also a few homesteads with many fewer cattle (one or two) than the average.

Next, we look at the purposes for which the homesteads kept cattle (Question 3 in CS11Quest2). After opening CS11Data2a a bar chart can be produced in GenStat in a similar way to that for the ranking of sources of income. The bar chart shows the relative frequencies (as a percentage of the responding homesteads) with which each of the purposes of keeping cattle were ranked as: primary, secondary, tertiary, or alternatively ticked as an additional purpose.

Combining by eye the ranks for primary and secondary purposes we see that the major reasons for keeping cattle are for work/draft and cash from sales. Manure and milk were also rated important, followed by meat, investment and dowry.

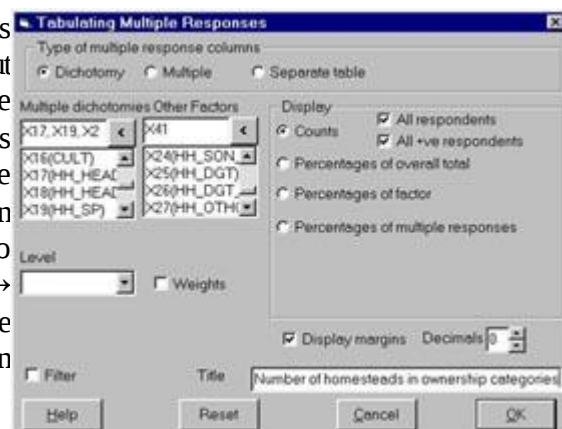
A striking feature of the bar chart is the number of times that traits were ticked but not ranked among the top three. This indicates the multiple reasons why cattle are important to the homestead. In the final data analysis it will be useful to be able to capture and summarise this information further.



Ownership

We describe the ownership of cattle by using the frequencies of homesteads (for each category of the members of the homestead) where members of the category owned cattle (Question 4, CS11Quest2).

Here again we demonstrate the use of Instat as we found it relatively easy to use. The input needed to generate the multiple response table for the numbers of different family categories owning cattle in relation to the gender of the head of the homestead is shown alongside. In order to access the dialog box one needs to open CS11Data2a and click Statistics → Tables → Multiple response... The corresponding percentage table is generated in a similar way.



Number of homesteads in ownership categories			
	GENDER		
Summary	Male	Female	All
HH_HEAD	121	18	139
HH_SP	11	11	22
HH_HDSP	7	2	9
HH_SON	26	7	33
HH_DGT	3	2	5
HH_OTHCD	9	3	12
Responding	133	30	163
Respondents	133	30	163
All	177	43	220

Percent of homesteads in ownership categories			
	GENDER		
Summary	Male	Female	All
HH_HEAD	91	60	85
HH_SP	8	37	13
HH_HDSP	5	7	6
HH_SON	20	23	20
HH_DGT	2	7	3
HH_OTHCD	7	10	7
All	133	143	135

The total number of categories is shown to be 220, which is more than the number of homesteads, 163. As a result, the percentages of homesteads in which a particular category of ownership occurred do not add up to 100 percent. Again this can be seen to be a feature of multiple response data.

The tables show that ownership of cattle mainly rested with the head of the homestead. This was more so when the head was a male. The tables do not show, however, those homesteads where ownership is restricted to only one category, i.e. all cattle owned by one member. Therefore, we used the Advanced Filter command in Microsoft Excel to identify the relevant homesteads where this was the case and then constructed a new variable (factor) indicating whether the head only or spouse only, etc. owned all the cattle in the homestead.

	GENDER		
Summary	Male	Female	All
HH_HEAD	121	18	139
HH_SP	11	11	22
HH_HDSP	7	2	9
HH_SON	26	7	33
HH_DGT	3	2	5
HH_OTHCD	9	3	12
Responding	133	30	163
Respondents	133	30	163
All	177	43	220

	GENDER		
Summary	Male	Female	All
HH_HEAD	91	60	85
HH_SP	8	37	13
HH_HDSP	5	7	6
HH_SON	20	23	20
HH_DGT	2	7	3
HH_OTHCD	7	10	7
Responding	133	143	135
Respondents	133	143	135
All	133	143	135

The values against 'Responding' in these new tables show the numbers of homesteads in which ownership was confined to one member.

When the head was male, he himself owned all the cattle in 64% (85/133) of all homesteads. This compares with 37% (11/30) when the head was female.

In 13% (4/30) of female-headed homesteads the spouse owned all the cattle. Ownership of cattle was shared between the head and his/her spouse in 6% (9/220) of the homesteads.

	GENDER		
Summary	Male	Female	All
HH_HEAD_ONLY	85	11	96
HH_SP_ONLY	1	4	5
HH_HDSP_ONLY	3	2	5
HH_SON_ONLY	4	1	5
HH_OTH_ONLY	1	0	1
Responding	94	18	112
Respondents	133	30	163
All	94	18	112

	GENDER		
Summary	Male	Female	All
HH_HEAD_ONLY	64	37	59
HH_SP_ONLY	1	13	3
HH_HDSP_ONLY	2	7	3
HH_SON_ONLY	3	3	3
HH_OTH_ONLY	1	0	1
Responding	71	60	69
Respondents	71	60	69
All	71	60	69

In order to examine further the influence of gender of head of homestead on cattle ownership, we summarised the results in a contingency table categorised by ownership

solely by the head versus other combinations of ownership.

The percentages in the table show that when the head was male he was almost twice as likely to be the sole owner of all the cattle in the homestead compared with all other categories of ownership. On the other hand, when the head was female, the pattern was reversed.

However, these results really need to be tested statistically before a final conclusion can be made. This is done later under Statistical modelling.

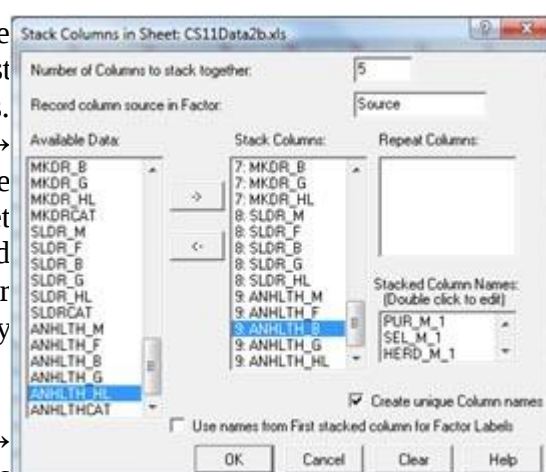
Ownership Category	Gender of head of homestead		
	Male	Female	All
Head only	85(64 ^a)	11(37)	96(59)
Other combinations	48(36)	19(63)	67(41)
Total	133(100)	30(100)	163(100)

^a Percentages in italics.

Cattle responsibilities

In order to establish the proportions of homestead members that were involved in each of the cattle activities (Question 5 in CS11Quest2) we need to first count the numbers of homestead members involved for each activity.

As the different activities for each member are stored in different fields in CS11Data2b we first need to stack them into single activity columns. We can do this in GenStat by using **Spread** → **Manipulate** → **Stack...** and then completing the dialog box alongside to produce a spreadsheet with columns which have been renamed Purchase, Selling etc. preceded by the factor 'Source' with levels 1 to 5 relating to family member category.



Following **Stats** → **Summary Statistics** → **Frequency Tables...** and producing 2-way tables for Source by each activity in turn we can derive a series of frequencies, as shown below for breeding.

Breeding Source	Breed		
	0.0000	1.0000	Count
1	55	109	164
2	153	11	164
3	163	1	164
4	164	0	164
5	163	1	164
Count	698	122	820

As can be seen, in the table, the total numbers of members involved in each activity varies and is different from the number of homesteads.

Activity	No. of homesteads involved in activity	No. of member categories involved	Frequency (percentage) of homesteads involved			
			Adult males	Adult females	Boys	Girls
Purchasing	152	172	135 (89)	33 (22)	2 (1)	1 (1)
Selling	144	159	126 (88)	31 (22)	2 (1)	0 (0)
Herding	154	187	64 (42)	22 (14)	57 (37)	2 (1)
Breeding	117	122	109 (93)	11 (10)	1 (1)	0 (0)
Feeding	64	69	40 (62)	7 (11)	9 (14)	1 (2)
Milking	139	159	59 (42)	19 (14)	53 (38)	2 (2)
Making dairy products	37	40	12 (32)	17 (46)	5 (14)	2 (5)
Selling dairy products	16	17	5 (31)	9 (56)	2 (12)	0 (0)
Animal health	154	174	137 (89)	22 (14)	11 (7)	1 (1)

The involvement or not of a homestead in an activity has already been stored in CS11Data2b as PURCAT, SELCAT etc. By stacking these variables in the same way and forming a frequency table, counts of 152, 144, 154, 117, 64, 139, 37 16 and 154 for the different activities, respectively, are obtained. We can add these to the table and use them as denominators to form the percentages.

Adult males were involved in purchasing, selling, breeding and animal health in approximately 90% of homesteads. Fewer homesteads reported that they were involved in breeding compared with purchasing or selling cattle, presumably because of reliance on random mating of their cows with bulls in the community. Similarly, only a small proportion responded to the question on feeding indicating that only two fifths or so of homesteads fed supplementary feed.

Activity	No. of homesteads involved in activity	No. of member categories involved	Frequency (percentage) of homesteads involved			
			Adult males	Adult females	Boys	Girls
Purchasing	152	172	135 (89)	33 (22)	2 (1)	1 (1)

Selling	144	159	126 (88)	31 (22)	2 (1)	0 (0)	0 (0)
Herding	154	187	64 (42)	22 (14)	57 (37)	2 (1)	42 (27)
Breeding	117	122	109 (93)	11 (10)	1 (1)	0 (0)	1 (1)
Feeding	64	69	40 (62)	7 (11)	9 (14)	1 (2)	12 (7)
Milking	139	159	59 (42)	19 (14)	53 (38)	2 (2)	26 (19)
Making dairy products	37	40	12 (32)	17 (46)	5 (14)	2 (5)	4 (11)
Selling dairy products	16	17	5 (31)	9 (56)	2 (12)	0 (0)	1 (6)
Animal health	154	174	137 (89)	22 (14)	11 (7)	1 (1)	3 (2)

Manufacturing and selling of dairy products was more often done by adult females than males in the few homesteads involved in these activities. Boys were mainly involved in herding and milking. These activities were also often hired out, presumably because boys will often be attending school. Girls were not much involved in the homestead cattle activities.

Statistical modelling

Results of data exploration

Surveys often need less formal statistical analysis than experiments. Tables can often be compiled from the frequency tables calculated during the exploratory phase and presented without further analysis. It is only when one is unsure whether an observed differences is real or not that some statistical verification may be important. This may result in the use of simple t- or Chi-square tests or sometimes, where the data warrant it, in more complicated analysis of variance or logistic regression techniques. So far, under data exploration, we have discovered the following:

1. Males headed four fifths of the homesteads. No females 30 years of age or younger were heads of homesteads.
2. Livestock was important, as it was ranked as one of the three major sources of income. The others were crops and salary/wages. Homesteads ranking livestock as the primary source of income tended to have slightly more cattle on average than other homesteads although the average difference might not be statistically significant.
3. Cattle were important to the homesteads and were reported as the species of primary importance by over 90% of the homesteads. Cattle were mainly kept for the purposes of work/draft and cash from sales, though meat, milk and manure were also ranked as important. A method for summarising the relative importance of the different purposes by pooling the different rankings might be helpful.
4. Ownership of cattle mainly rested with the head of the homestead especially when the head was male. Female heads appeared more willing to share cattle ownership with other members of the homestead than their male counterparts. This, however, needs to be statistically tested.
5. The numbers of homesteads involved in the different activities varied with few involved in the manufacture and selling of dairy products.
6. Adult males were mainly involved in the purchasing, selling, breeding and in the care of the health of cattle. Adult females were mainly involved in manufacturing and selling of dairy products. Boys and hired labour were mainly involved in herding

and milking cattle, while girls participated little in the cattle activities.

So we now need to:

- Test formally whether there is a significant difference in average number of cattle between homesteads where livestock provided the major source of income and those for which livestock did not provide the main source of income.
- See whether we can form an index that pools the various rankings to a multiple response question.
- Test if the proportion of heads of homesteads that own all cattle themselves is significantly greater for male than female heads.



Numbers of cattle

In order to test whether the homesteads, for which livestock provides the major source of income, possess more cattle on average than those for which livestock does not, a linear mixed model (REML), with sub-region and dip-tank area within sub-region assumed to be random effects and primary source of income as a fixed effect, was fitted to the transformed cattle numbers in CS11Data1. To perform the analysis, the relevant GenStat input is entered in the dialog box obtained by clicking **Stats** → **Mixed Models (REML)** → **Linear Mixed Models** and putting the variable `livst_inc2` in the **Fixed model:** window and `subregion/diptank` in the **Random model:** window.

There is a significant difference ($P < 0.05$) between the two groups of homesteads.

```
**** REML Variance Components Analysis ****
```

```
Response Variate : log_cat_no
```

```
*** Estimated Variance Components ***
```

Random term	Component	S.e.
-------------	-----------	------

subregion	0.0154	0.0173			
subregion.diptank	0.0341	0.0198			
*** Residual variance model ***					
Term	Factor	Model	Parameter	Estimate	S.e.
Residual		Identity	Sigma2	0.442	0.0341
*** Wald tests for fixed effects ***					
Fixed term	Wald statistic	d.f.	Wald/d.f.	Chi-sq prob	
livst_inc2	6.03	1	6.03	0.014	

The variance components show that there was no significant variation among sub-regions, but there was some indication of differences between dip tank areas within sub-regions.

It should be noted, however, that we are making an assumption that the dip tank areas were selected at random, when in fact their selection was not strictly random; purposive sampling was involved in some cases as indicated under Study design.

**** REML Variance Components Analysis ****

Response Variate : log_cat_no

*** Estimated Variance Components ***

Random term	Component	S.e.
subregion	0.0154	0.0173
subregion.diptank	0.0341	0.0198

*** Residual variance model ***

Term	Factor	Model	Parameter	Estimate	S.e.
Residual		Identity	Sigma2	0.442	0.0341

*** Wald tests for fixed effects ***

Fixed term	Wald statistic	d.f.	Wald/d.f.	Chi-sq prob
livst_inc2	6.03	1	6.03	0.014

Comparison of sources of income

A number of questions provide ranked answers (e.g. Question 8 CS11Quest1, Question 3 CS11Quest2). So far we have produced bar charts to illustrate proportions of different characteristics receiving different rankings, but can we quantitatively combine these results? Rowlands et al. (2003) propose a method for developing an index based on the individual rankings. Their formula is calculated as follows:

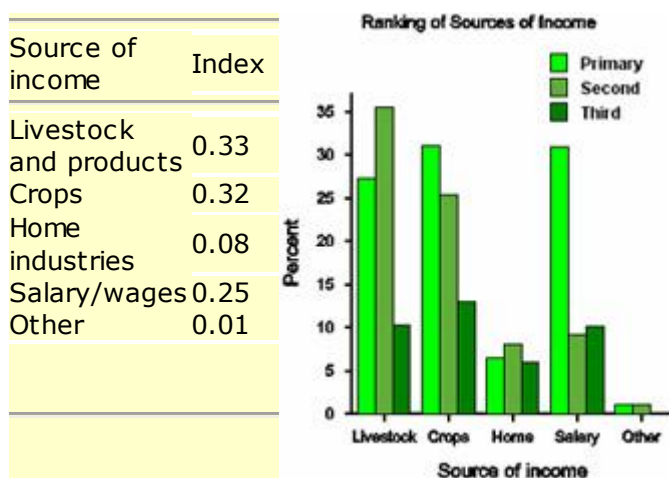
$$\text{Index} = \frac{\text{sum of [3 if rank 1 + 2 if rank 2 + 1 if rank 3] for a given source of income}}{\text{sum [3 if rank 1 + 2 if rank 2 + 1 if rank 3] over all sources of income.}}$$

Thus, if in a particular homestead the ranking is: crops 1, salary/wages 2 and livestock and products 3, contributions to the index for crops would be 3 for the numerator and $(3 + 2 + 1 = 6)$ for the denominator. For the index for salary/wages the contribution to the numerator would be 2 and the denominator 6 as before, and for livestock 1 and 6, respectively.

This index adds up to 1 when summed over the different sources of income; the higher the value of an individual index for a particular source of income the greater the overall importance of that source of income.

We used Excel to calculate the values of the index for each source of income. Crop and livestock systems provided similar levels of sources of income for the homesteads. Salary/wages also provided important source of income.

The bar chart produced earlier (and repeated below) showed that crops were most frequently ranked first. Livestock, however, followed closely and were ranked second more often than crops. A combination of the index information and ranking information provides a good overall picture.



Purposes of keeping cattle

The purpose of keeping cattle can also be analysed by the 'index' method.

The table combines both index and ranking information to provide a good overall picture. The table confirms the high importance of work/draft. Indices for meat, milk, manure and cash from sales tended to be more similar. Cash from sales ranked second to work/draft and more frequently than meat, milk and manure as the primary purpose for keeping cattle. However, each of these last three frequently as second and third ranked purposes. This resulted in the comparatively higher ratings based on the index than when only first position ranks were considered.

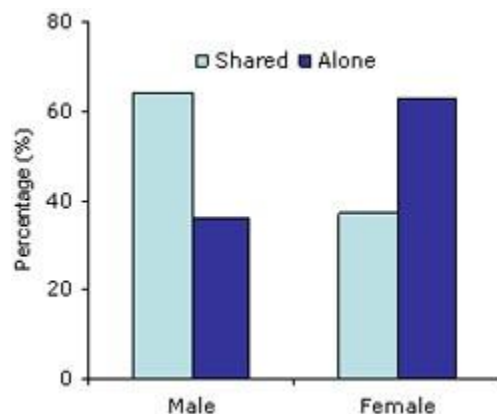
Purpose	Ranked as the primary purpose (% of homesteads)		Index	
	Livestock - primary income source	Livestock - not primary income source	Livestock - primary income source	Livestock - not primary income source
Meat	7.5	7.2	0.11	0.11
Milk	3.8	3.6	0.12	0.13
Work/draft	47.2	54.1	0.22	0.26
Stud breeding	0.0	0.0	0.02	0.01
Manure	1.9	8.1	0.10	0.14
Blood	0.0	0.0	0.01	0.01
Hide	1.8	0.0	0.02	0.01
Cash from sale	28.3	16.2	0.19	0.14
Investment	3.8	0.9	0.04	0.02
Dowry	5.7	4.5	0.07	0.08
Ceremonies	0.0	0.9	0.06	0.04
Cultural	0.0	1.8	0.05	0.06
Other	0.0	0.0	0.00	0.00

Comparison of corresponding indices for the two types of homesteads shows that cash from sales featured higher for homesteads where livestock was the major source of income. The opposite tended to be true for work/draft and manure. This might be expected as the homesteads for which livestock did not provide the main source of income would be likely to rely more on their cattle for work/draft and manure for the production of crops as the alternative source of income.

Purpose	Ranked as the primary purpose (% of homesteads)		Index	
	Livestock - primary income source	Livestock - not primary income source	Livestock - primary income source	Livestock - not primary income source
Meat	7.5	7.2	0.11	0.11
Milk	3.8	3.6	0.12	0.13
Work/draft	47.2	54.1	0.22	0.26
Stud breeding	0.0	0.0	0.02	0.01
Manure	1.9	8.1	0.10	0.14
Blood	0.0	0.0	0.01	0.01
Hide	1.8	0.0	0.02	0.01
Cash from sale	28.3	16.2	0.19	0.14
Investment	3.8	0.9	0.04	0.02
Dowry	5.7	4.5	0.07	0.08
Ceremonies	0.0	0.9	0.06	0.04
Cultural	0.0	1.8	0.05	0.06
Other	0.0	0.0	0.00	0.00

Sixty four percent (85/133) of male heads and 37% (11/30) of female heads were the only

owners of cattle in their homesteads. Are these differences significant?



The Chi-square test is an appropriate test to test this. To perform the test, click Stats → Statistical Tests → Contingency Tables.... Next, click the Create Table button in the dialog box to create a 2-way table with a blank spreadsheet. Rename column 1 as Male and column 2 as Female and fill in the spreadsheet with the values 85 11 48 and 19. Click 'OK' and complete the Chi-square test. This gives a chi-square value of 7.50 with 1 degree of freedom and a P-value of 0.006.

The result shows that there is sufficient evidence ($P < 0.01$) to believe this difference to be real. Thus, female heads were more willing to allow ownership to be distributed between members of their households.

Findings, implications and lessons learned

Some statistical implications of this case study are as follows.

1. It is seen from this case study that survey analysis does not always require formal statistical evaluation. Inferences can be made from tables and graphical presentations when the results are obvious. The graphical presentations give overall trends. Of course, when the results are not obvious, then they must be statistically quantified, and inferences made only if the results are statistically significant.
2. The case study has shown how sample size determines how far the data can be classified. Thus, it was appropriate to present the distribution of the gender and age of head of homestead only down to the sub-regional level and not dip-tank area level. It is important when planning a survey to decide what questions need to be asked and whether the sample size is large enough. It is generally necessary in terms of cost and manpower to limit the sampling to what can be managed. However, there may be a corresponding 'cost' in terms of how far the data can be classified.
3. One also has to be aware that proposed sample sizes are not always achieved. Thus, some questionnaires in this survey had to be discarded due to incomplete or missing information, particularly the key information needed before any subsequent entries could be made, and so the sample size in each dip tank area often ended up being less than the target of 10 homesteads.

4. The case study has demonstrated how to handle multiple response data and describes how they can often be mistreated, especially when the wrong choice of the denominator is made in calculating percentages. It is important to code the questions in a way that recognises the multiple levels in the data. For example, the data in this case study were recorded at the homestead level, as homesteads were the ultimate sampling units, but for certain questions such as the members of the homestead responsible for cattle activities there was the activity level underneath the homestead level, followed by the homestead member category level (adult males, adult females, boys, girls or hired labour) underneath the activity level.
5. The case study has also demonstrated the value of combining multiple ranked responses into the form of an index which adds to the interpretation of the relative importance of different responses.
6. The large amount of work that had been done in this case study to analyse a limited amount of data emphasises the difficulties that can arise in the handling of data from surveys. Many researchers underestimate the amount of analysis that will result. It is therefore very important to be careful when planning a survey to collect only the data that meet the objectives of the survey.

Reporting

There is an art in reporting survey results.

Surveys can provide lots of information in a vast number of tables. The data analyst needs to sift through the various tabulations that can be done and select those that provide the most valuable information.

Then he/she needs to decide whether to present results in tabular or graphical form. Graphical presentations are useful when the author wishes to demonstrate trends, but they do not give precise information on mean values nor, when included, also their standard errors. A good report will use both methods of presentation.

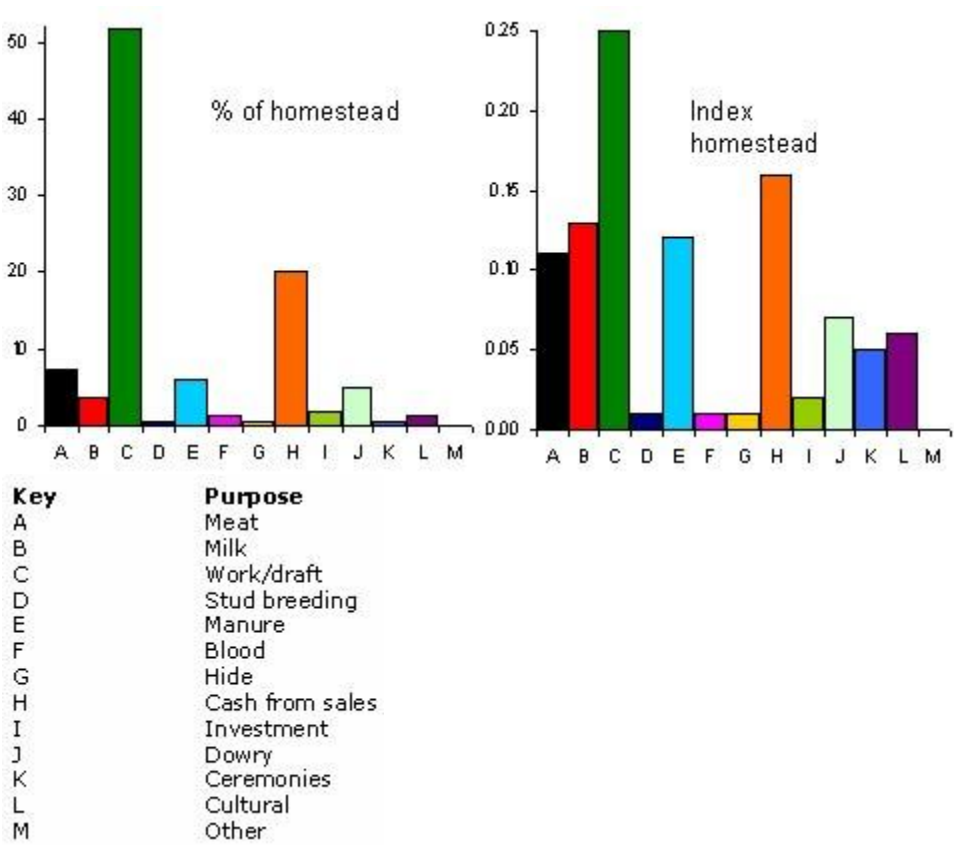
Tables should be presented as simple as possible as shown here - the table shown previously but now averaged over the two different types of homestead.

Note that the numbers of decimal places that have been chosen are sufficient to discriminate between different values in the table. There is nothing worse than having a table with superfluous decimal places.

Purpose	Ranked as primary purpose (% of homesteads)	Index
Meat	7.3	0.11
Milk	3.7	0.13
Work/draft	51.8	0.25
Stud breeding	0.6	0.01
Manure	6.1	0.12
Blood	1.2	0.01
Hide	0.6	0.01
Cash from sales	20.1	0.16
Investment	1.8	0.02
Dowry	4.9	0.07
Ceremonies	0.6	0.05
Cultural	1.2	0.06
Other	0.0	0.00

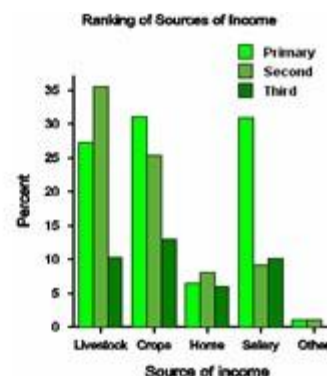
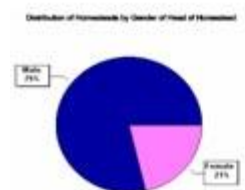
A graphical representation of the same data might, however, provide a more visual comparison of the two methods of ranking the different purposes for keeping cattle. Results in the previous table have been averaged over homesteads where livestock is or is not the primary source of income.

Note that two sets of bar charts have been scaled so that they occupy similar vertical widths. The graph demonstrates that, although work/draft is most frequently the primary purpose for keeping cattle, this method of calculation hides the importance of other purposes that may be ranked second or third.



The report needs to be geared towards its potential readership. The types of information required by government ministry officials, by other stakeholders, by researchers or farmers are different. Whilst series of tables may be important for agricultural development officers and researchers, they will also need to be provided with graphical presentations of the results, as just shown, to indicate overall trends. Some documentation will also be required to indicate the important, and, where appropriate, the statistically significant results. Care must be taken in ensuring that inferences are made only when the results are obvious or can be qualified statistically.

Farmers are at the other end of the spectrum. They need simple presentations of the results in a pictorial form (such as many of the graphical representations shown in this case study) that they can clearly understand. They and other organisations involved in the execution of the survey are often neglected. Feed back to them is important and is often not done.



Pie charts and bar charts are a particularly useful method of presentation. Case Study 7 discusses a different method of reporting back to farmers.

A report is best organised if the results can be presented to match the questions that the survey set out to address. In this way the results and the accompanying documentation will give direct answers to the questions of the study. Hence the report will guide the government ministry officials in policy making and the farmers in making informed decisions.

We have, in this case study, tried to follow this guideline in presenting the results according to the questions posed at the outset, namely:

- What is the distribution of household heads by age and gender?
- What importance do livestock play in the homesteads and how important are cattle?
- Which members of homesteads own cattle, and who are responsible for different cattle activities?

Other examples of methods of reporting are given amongst the material included under

Study questions

1. Describe what is meant by a random walk. A number of questionnaires in this survey had to be disposed when the enumerator found on reaching a homestead that the homestead did not possess the required 'primary' species. Plan an alternative method for ensuring that the required samples of homesteads are surveyed for 'primary' species. With this method, and assuming that all species are raised describe how you would use the random walk method to select 2 homesteads each for cattle, sheep, goats and chickens and one each for pigs and donkeys in a dip-tank area.
2. Write an essay on interviewing techniques for a survey such as the one described in this case study.
3. Use GenStat to reproduce the bar chart for the ranking of sources of income shown under Importance of cattle.
4. Repeat the analysis under Cattle responsibilities but this time produce two separate tables for homesteads headed by males and females, respectively. It will be necessary to copy the gender field into CS11Data2b. Include just the activities for purchasing,

- selling, herding and milking in the analysis. Comment on the results. If deemed appropriate, apply a statistical test to confirm your conclusions.
5. Apply the steps considered in this case study under Research strategy to the situation in your country. What changes, if any, would you recommend?
 6. Assume that you have been given the responsibility for supervising the data entry for this survey and that you have four data entry clerks. Describe how you would organise the data entry and plan a suitable checking procedure to ensure that the data entry errors are minimised.
 7. Demonstrate that you can use GenStat as shown in this case study to analyse of proportions of households involved in different activities.
 8. Demonstrate that you can use Instat to generate the numbers of households shown in this case study to own different species.
 9. Assume that members of your class are farmers. Prepare a talk using results from this case study that you think will be of interest to them.
 10. Discuss the advantages and disadvantages of database management systems such as Access compared with a system based on spread sheets such as Excel and the types of situations where you would prefer to use one or the other.

Related reading

Ayalew, Workneh and Rowlands, J. (eds). 2004. *Design, execution and analysis of the livestock breed survey in Oromiya Regional State, Ethiopia*. OADB (Oromiya Agricultural Development Bureau), Addis Ababa, Ethiopia, and ILRI (International Livestock Research Institute), Nairobi, Kenya. Full text

Kosgey, Isaac Sanga. 2004. *Breeding objectives and breeding strategies for small ruminants in the tropics*, Ph.D Thesis. Animal Breeding and Genetics Group, Wageningen University, The Netherlands. pp. 216 Full text

Rowlands, John, Nagda, Sonal, Rege, Edward, Mhlanga, Flora, Dzama, Kennedy, Gandiya, Faith, Hamudikwanda, Humprey, Makuza, Stanley, Moyo, Siboniso, Matika, Oswald, Nangomasha, Edward and Sikosana, Joseph. 2003. *A report to FAO on 'The design, execution and analysis of livestock breed surveys - a case study in Zimbabwe*. ILRI (International Livestock Research Institute), Nairobi, Kenya. Full text

Rowlands, John, Ayalew, Workneh and Hanotte, Olivier. 2005. *Cattle breed distributions across Borana Zone as determined from an analysis of phenotypic data collected in the Oromiya Regional State Livestock Breed Survey. A Report to Oromiya Agricultural Development Bureau*. International Livestock Research Institute, Addis Ababa, Ethiopia, 17pp. Full text

Acknowledgements

The survey in Swaziland was conducted by the Ministry of Agriculture and Co-operatives (Department of Veterinary and Livestock Services), in collaboration with the University of

Swaziland. We give sincere thanks to all those who were involved from the planning stage to the data collection and data capture and cleaning. Also, many thanks to the farmers in the Hhohho region for their kindness and time they took to answer the questions in the survey questionnaire.

We also thank the Department of Veterinary and Livestock Services for permission to include some of the raw data collected in the survey for teaching and learning purposes. Teachers and other users of this case study are welcome to use the data provided **but must not publish any results without permission** of the Ministry of Agriculture and Co-operatives and until they have themselves published all the findings from the survey.

The first author made his contribution to the work presented in this case study while he was a Visiting Research Fellow at the School of Applied Statistics, University of Reading. He thanks the School and the University for hosting his visit and for all the support and assistance afforded to him during the visit. In particular, he would like to give sincere thanks to the Head of School, Dr. Derek Pike, for supervising the work.

Furthermore, the first author visited the School through a Commonwealth Fellowship Award offered by the Commonwealth Scholarship Commission in the United Kingdom under the Commonwealth Scholarship and Fellowship Plan. He would like to sincerely thank the Commission for the Award. The Commission's Secretariat is the Association of Commonwealth Universities, which was responsible for all academic aspects of the Award. The first author gives due thanks to the Association and to the British Council, which was responsible for all welfare aspects of the Award.

During the tenure of the first author's Commonwealth Fellowship Award in Britain, he was on paid Sabbatical Leave by his employer, the University of Swaziland. He would like to sincerely thank the University of Swaziland for approving his Sabbatical Leave and for the financial support the University afforded him to have his family join him in the United Kingdom during his stay at the University of Reading.