

Data File

R031

| | |
|-------------------|---|
| Content | This dataset file contains information related to household information |
| Cases | 248219 |
| Variable(s) | 18 |
| Version | Version 1.0 |
| Producer | National Bureau of Statistics |
| Missing Data | * indicate the missing data |
| | Data processing consisted of the following processes: |
| | <ul style="list-style-type: none">· Data entry· Data structure formatting· Batch validation· Tabulation |
| | Data Entry |
| | Scanning and ICR data capture technology for the small holder questionnaire were used on the Mainland. This not only increased the speed of data entry, it also increased the accuracy due to the reduction of keystroke errors. Interactive validation routines were incorporated into the ICR software to track errors during the verification process. The scanning operation was so successful that it is highly recommended for adoption in future censuses/surveys. In Zanzibar all data was entered manually using CSPro. |
| | Prior to scanning, all questionnaires underwent a manual cleaning exercise. This involved checking that the questionnaire had a full set of pages, correct identification and good handwriting. A score was given to each questionnaire based on the legibility and the completeness of enumeration. This score will be used to assess the quality of enumeration and supervision in order to select the best field staff for future censuses/surveys. |
| | CSPro was used for data entry of all Large Scale Farm and community based questionnaires due to the relatively small number of questionnaires. It was also used to enter data from the 2,880 small holder questionnaires that were rejected by the ICR extraction application. |
| | Data Structure Formatting |
| | A program was developed in visual basic to automatically alter the structure of the output from the scanning/extraction process in order to harmonise it with the manually entered data. The program automatically checked and changed the number of digits for each variable, the record type code, the number of questionnaires in the village, the consistency of the Village ID Code and saved the data of one village in a file named after the village code. |
| Processing Checks | Batch Validation |
| | A batch validation program was developed in order to identify inconsistencies within a questionnaire. This is in addition to the interactive validation during the ICR extraction process. The procedures varied from simple range checking within each variable to the more complex checking between variables. It took six months to screen, edit and validate the data from the smallholder questionnaires. After the long process of data cleaning, tabulations were prepared based on a pre-designed tabulation plan. |
| | Tabulations |
| | Statistical Package for Social Sciences (SPSS) was used to produce the Census tabulations and Microsoft Excel was used to organize the tables and compute additional indicators. Excel was also used to produce charts while ArcView and Freehand were used for the maps. |
| | Analysis and Report Preparation |
| | The analysis in this report focuses on regional comparisons, time series and national production estimates. Microsoft Excel was used to produce charts; ArcView and Freehand were used for maps, whereas Microsoft Word was used to compile the report. |
| | Data Quality |
| | A great deal of emphasis was placed on data quality throughout the whole exercise from planning, questionnaire design, training, supervision, data entry, validation and cleaning/editing. As a result of this, it is believed that the census is highly accurate and representative of what was experienced at field level during the Census year. With very few exceptions, the variables in the questionnaire are within the norms for Tanzania and they follow expected time series trends when compared to historical data. Standard Errors and Coefficients of Variation for the main variables are presented in the Technical Report (Volume I). |

Variables

| ID | Name | Label | Question |
|----|------|-------|----------|
|----|------|-------|----------|

| | | | |
|-----|----------|------------------------|-------------------------------------|
| V27 | CASE\$ID | | |
| V28 | Q031_ROW | S/No | |
| V29 | Q031C02 | Relationship to Head | Relationship to head |
| V30 | Q031C03 | Sex | Sex |
| V31 | Q031C04 | Age | Age |
| V32 | Q031C05 | Survival of Mother | Survival of parents:Mother |
| V33 | Q031C06 | Survival of Father | Survival of parents:Father |
| V34 | Q031C07 | Read & Write | Read and Write |
| V35 | Q031C08 | Education Status | Education status |
| V36 | Q031C09 | Education Level | Education level reached |
| V37 | Q031C10 | Involvement in Farming | Involvement in farming |
| V38 | Q031C11 | Main Activity | Main activity(for aged 5 and above) |
| V39 | Q031C12 | Off Farm Income | Off-farm income |
| V40 | Hecta | Size of Holding (Ha) | |
| V41 | Class | Planted Area Class | |
| V42 | Region | Region | |
| V43 | District | District | |
| V44 | Wt | Village weight | |