
Data Collection and Processing in South Africa

(Economic Large Scale Surveys)

Presenters: Itani Magwaba and Michael Manamela

Contents

1. Introduction
2. Background
3. Organisation
4. Data collection
 - 4.1 Dispatching questionnaires
 - 4.2 Follow-ups
 - 4.3 Treatment of responses
5. Data processing
6. Way forward

Introduction

- The **main aim** during data collection is to ensure that **quality** data is collected with a '**high**' **response rate** (i.e. sufficient quantity).
- Thus, the need to ensure **accurate data** capturing.
- The **editing** of data guarantees that will have quality data.

Background

- Data collection and processing is centralised at Head Office.
- Collection is mainly through self-enumeration.
- Personal visits are conducted after request or failure to respond despite repeated telephonic and written reminders. Mostly small and medium sized (in terms of turnover) enterprises are visited.

Organisation

- **The workload per enumerator is 450.**

The total sample size for the 2004 large sample surveys is 17 000 enterprises. The total number of enumerators is 38.

- **The supervisor to staff ration is presently 10.**

This will be reduced to 6 before starting with the collection of the 2005 large sample surveys.

- **The cost for enumerating one enterprise is R605,00 (approximately US\$90).**

Dispatching questionnaires

1. The sample is distributed among the enumerators.
2. The enumerators update and confirm the classification and contact details by telephone.
3. The contact details are updated mainly using phone books, internet, previous surveys and the business register.
4. The questionnaires are then dispatched by mail, fax or internet.

Follow-ups

1. There is a need for continuous follow-ups to remind respondents regarding outstanding questionnaires and to establish if any assistance is needed.
2. Follow-ups are made regularly via telephone and written reminders are sent by fax and/or registered mail.
3. Four written reminders are sent throughout the collection period. The final reminder mentions 'legal action' and is sent by registered mail.

Treatment of various responses

The following are the main response categories:

- Outstanding questionnaire (00).

A questionnaire can be outstanding because of non-response and no contact.

- Received completed questionnaire (04).
- Received financial statement (05).
- Investigation with contribution (12).
- Investigation with **zero** contribution (13).

Note: Investigation refers to an enterprise which has either 'merged', 'sold', 'split off', 'taken over', 'liquidated during or before survey', 'closed down during or before survey', 'untraceable', 'foreign business' or 'reclassified out of scope'.

Receiving questionnaires and finalising investigations

- Most of the questionnaires and/or investigations are received by fax, followed by mail and e-mail.
- An investigation is only finalised after receiving signed written confirmation from the respondent and approval by the supervisor or survey statistician or manager.
- In some cases, the business register is used to finalise investigation, e.g. an enterprise 'closing down during the survey period'.

Data collection

Preliminary checking

A **'completed' questionnaire** is checked to ensure the following:

- The information received is for the sampled enterprise by comparing company registration number and name.
- The classification is correct by checking main activity. If the classification is incorrect, the enterprise is considered to be 'out of scope'.
- The information is for survey period by checking enterprise's financial year.
- All the relevant questions have been answered.

Capturing System

1. The capturing system is developed using visual basic.
2. The 'editing instructions' are contained within the capturing system.
3. The capturing system is linked to an SQL database.
4. Capturing of questionnaires is done manually through the keyboard.
5. After capturing each questionnaire, a 'warning messages' reported is reflected.

Micro-editing

- This process detects errors in data through checking of individual enterprise information.
- Initial micro-editing is done through preliminary checking on receiving completed questionnaire.
- During capturing, mainly range and consistency data edits are done.
- After capturing, ratios are computed and plotted on scatter plot for each stratum (a combination of classification and size group). Outliers are identified and investigated.

Ratios

The following are examples of the computed ratios:

- Total income / 'Register turnover',
- Total expenses / Total income,
- Gross salaries and wages / Number of employees,
- Net profit before tax / Total income i.e. profitability ratio, and
- Sales of services / Sales of goods

Treatment of outliers

- **Large enterprises/ businesses**

- If a respondent fails to verify outlier, the enterprise is treated as non-response.
- If the outlier is verified, the information is accepted.

Note: The large enterprises are completely enumerated and thus have a design weight of one.

- **Other enterprises/ businesses**

- If an outlier (whether verified or not) has an impact on the estimates, it is treated a non-response.

Macro-editing

1. This involves detecting errors in data through the analysis of aggregate data (totals).
2. The aggregated data is compared with data from the following surveys:
 - 2.1 Annual business or household surveys
 - 2.2 Short-term business or household surveys (monthly or quarterly)

Unit imputation

(Compensation for unit non-response)

- **Large enterprises**

All non-responding large businesses or enterprises are imputed.

- **Small and medium enterprise**

The non-responding units are not imputed.

Dissemination of results

- There are two types of publications:
 - A '**statistical release**' which contains mainly financial information, and/or
 - A '**report**' which contains financial and 'products' information.
- The results are only published at **national** level. No regional figures are published at the moment.

IMPROVEMENTS

- De-centralise data collection to provinces. This will improve the response rate of small and medium enterprises as personal visits will be much easier.
- Dispatch questionnaires through registered mail. This will serve as proof that the questionnaire was received.
- Investigate the use of a more efficient capturing method, e.g. scanning.

THANK YOU