

Enterprise-based Sample Design

Naume Malepe and Mandla Masemula

Division: Methodology and Evaluation (M & E)

09 June 2008

Presentation Outline

1. Overview of Business Sampling Frame (BSF)
2. Example of sampling specifications
3. Sampling Methodology
 - 3.1. Stratification
 - 3.2. Sample allocation
 - 3.3. Drawing of sample (JALES Technique)

1. Overview of Business Sampling Frame (BSF)

- M & E receives BSF from Business Register (BR). BSF is a set of **ENTERPRISES** that serves as a common sampling frame for all business surveys within Statistics South Africa.
- An **ENTERPRISE** is characterized by a unique identifier which is an Enterprise Number, a unit identification name (*legal enterprise name*), and attributes describing the enterprise (*e.g. turnover, Industrial-classification*); and it is available for sampling.
- The sources of BSF are South African Receiver of Revenue (SARS), Department of Trade and Industry (DTI) and Department of Labor (DoL).
- Only enterprises that are classified as *birthed, alive* and *reactivated* form part of the BSF; *dead* and *deactivated* enterprises do not form part of BSF.
- The enterprises on the BSF are classified according to Standard Industrial Classification (SIC '93) codes, ranging from 1 to 5 digit-level.

Table 1: Standard Industrial Classification (SIC '93) codes

SIC	Description of the SIC
1	Agriculture, Hunting, Forestry and Fishing
2	Mining and Quarrying
3	Manufacturing
4	Electricity, gas and water supply
5	Construction
6	Trade Industries (61-Wholesale ,62 – Retail, 63 – Motor, 64 - Hotels and restaurants)
7	Transport, Storage and Communication
8	Financial intermediation
9	Community, social and personal services

Overview of BSF – (cont...)

- 1 digit-level is at overall industrial level and 5 digit-level is the most detailed level per industry. **Example:**

ENTERPRISE_NBR	SIC 1 = 6	SIC 2 = 64	SIC 3 = 641	SIC 4 = 6410	SIC 5 = 64101
EN12345678	Wholesale and retail trade of motor vehicles, motor cycles and personal and household goods; hotels and restaurants.	Hotel and Restaurants	Hotel, Camping sites and other provision of short-stay accommodation.	Hotel, Camping sites and other provision of short-stay accommodation.	Hotels, motels, botels and inns registered with the SA Tourism Board

Overview of BSF – (cont...)

- Measure of size on the BSF is an *Enterprise* **TURNOVER**
- **Turnover** of BSF is derived as follows:

For VAT registered enterprises registered → Turnover = Vat Turnover

For IT registered enterprises below VAT-threshold → Turnover = IT Turnover
(enterprises with turnover < R1 Million)

- Enterprises are size grouped based on Turnover as a Measure of size using the DTI cut-off points as stipulated in the National Small Business Amendment Bill of 2003; for the purpose of stratification (*more on slide 10*).

Overview of BSF – (cont...)

Table 2: DTI cut-off points as stipulated in the National Small Business Amendment Bill of 2003

Column 1	Column 2	Column 3	Column 4	Column 5
Sector or subsector in accordance with the Standard Industrial Classification	Size of class	The total full-time equivalent of paid employees	Total turn-over	Total gross asset value (fixed property excluded)
Agriculture	Medium	100	R5m	R5m
	Small	50	R3m	R3m
	Very Small	10	R0.50m	R0.50m
	Micro	5	R0.20m	R0.10m
Mining and Quarrying	Medium	200	R39m	R23m
	Small	50	R10m	R6m
	Very Small	20	R4m	R2m
	Micro	5	R0.20m	R0.10m
Manufacturing	Medium	200	R51m	R19m
	Small	50	R13m	R5m
	Very Small	20	R5m	R2m
	Micro	5	R0.20m	R0.10m

Size group 1=Medium, 2=Small, 3=Very small or 4=Micro

- **Example of Enterprises size grouping:**

Table 3: Distribution of enterprises on the Hotel Accommodation by SIZE GROUP

Enterprises on SIC 3=641(Hotel-Accommodation) per SIZE GROUP

Size Group	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Total turnover	Turnover Percentage contribution
1	100	0.05	100	3.70	15861560316.25	70.52
2	100	0.05	200	7.30	1895927409.57	8.43
3	100	0.07	300	8.37	356170729.75	1.58
4	100	3.74	400	100.00	4376891667.63	19.46

Note: For confidentiality purpose the figures above have been distorted

2. Example of sampling specifications

Table 4: Hotel Accommodation Survey sampling specifications for 2008

Survey title	Proposed sampling SIC level	Collection SIC level	Proposed sample size
Accommodation	5-digit	5-digit	900

Special requirement:

- Group sizes **1, 2 and 3** for Accommodation (SIC 641) to be fully enumerated.

3.1. Sampling Methodology-Stratification

- The main objective of stratification is to form homogeneous classes within the sampling frame.
- Strata are formed to reduce variability within the sampling frame, by concatenating the SIC (classification) and Size Group variables, viz. SAS statement **Stratum=SIC||size_Group.**

Sampling Methodology – Stratification (Cont...)

- Table 5 below shows example of 1st six strata formed within the Hotel Accommodation frame

Table 5: Enterprises on Hotel Accommodation frame per Stratum

Enterprises on SIC 3=641(Hotel-Accommodation) per STRATUM						
stratum	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Total turnover	Turnover Percentage contribution
6410 1	100	0.05	100	0.05	364222633.08	1.63
6410 2	100	0.05	200	0.09	21010298.47	0.09
6410 3	100	0.07	300	0.16	16970657.03	0.07
641011	100	3.74	400	3.90	13384865356.68	60.13
641012	100	2.84	500	6.75	1084905863.07	4.87
641013	100	0.85	600	7.59	199640267.70	0.89

Note: For confidentiality purpose the figures above have been distorted

3.2. Sampling Methodology – Sample allocation

- Neyman Allocation was used to allocate sample size within each stratum.
- The purpose of the method is to maximize survey precision, given a fixed sample size, and by assuming that the cost to sample an enterprise is equal across strata.
- Based on Neyman Allocation the best sample size for stratum h would be:
- $$n_h = n \times (N_h \times \sigma_h) / [\sum (N_i \times \sigma_i)],$$

where n_h = sample size for stratum h ,
 n = total sample size,
 N_h = population size for stratum h , and
 σ_h = standard deviation of stratum h .
- Turnover variable is used for Sample Allocation.
- Precision is the driving factor behind the sample size and allocation of sample size to various strata.

3.2. Sampling Methodology – Sample allocation

- Example of application of Neyman Allocation:

Suppose we have population of 140 enterprises distributed with σ (turnover standard deviation per stratum) as follows:

Stratum	N	σ	<i>n</i>
A	80	10,000	9
B	60	30,000	21

The required sample size is **30** enterprises,

The Neyman allocation will allocate sample sizes to the two strata as follows:

Stratum A sample size:

$$\begin{aligned}
 &= 30 N_A \times \sigma_A / (N_A \times \sigma_A + N_B \times \sigma_B) \\
 &= 30 \times 80 \times 10,000 / (80 \times 10,000 + 60 \times 30,000) \\
 &= 9.23 \approx \mathbf{9}
 \end{aligned}$$

Stratum B sample size:

$$\begin{aligned}
 &= 30 \times N_B \times \sigma_B / (N_A \times \sigma_A + N_B \times \sigma_B) \\
 &= 30 \times 60 \times 30,000 / (80 \times 10,000 + 60 \times 30,000) \\
 &= 20.77 \approx \mathbf{21}
 \end{aligned}$$

3.3. Sampling Methodology – Drawing of sample (JALES Technique)

- Samples are drawn using **Stratified Simple Random Sampling (SRS) which applies JALES technique** within each stratum; this takes place by using a unique random number that is available for each enterprise.
- **Unique number** is a random number uniformly distributed over the interval $(0,1)$ to every registered enterprise in the BSF.
- The **JALES technique** uses:
 - **starting point** is any number between $(0,1)$ selected to be different across surveys, to avoid an overlap of sampled enterprises,
 - and **sampling direction** chosen across surveys is to the right.

Sampling Methodology – Drawing of sample (Cont...)

- Every sampled enterprise carries a **design weight** = number of enterprises it represents in that stratum.
- The **design weight** of an enterprise within a stratum is calculated by dividing the population size by sample size of that particular stratum.

Example:

Stratum	N	σ	n	Design Weight
A	80	10,000	9	9
B	60	30,000	21	3

$$\text{Design Weight}_{(A)} = N/n = 80/9 = 8.89 \approx 9$$

$$\text{Design Weight}_{(B)} = N/n = 60/21 = 2.86 \approx 3$$

Explanation: Each sampled enterprise in Stratum A represents 9 enterprises; and each sampled enterprise in Stratum B represents 3 enterprises.



Thank you