

**PROJECT FOR STATISTICS ON LIVING STANDARDS AND DEVELOPMENT  
SURVEY 1993**

(SALDRU, University of Cape Town)

**SAMPLING FRAME: NUMBER OF CLUSTERS PER REGION**

Below is the number of clusters chosen per region. No great detail will be given here as to the whole process. An adjusted figure for the census population (which includes the self-governing territories, but excludes the TBVC), was used and supplemented by de facto population figures for the TBVC in order to get a total figure for the whole of South Africa. This total figure was used to calculate the first stage sampling interval. Unfortunately, due to discrepancies in the data (i.e. some are given as the number of individuals, others as the number of households per specified area), a separate sampling frame had to be developed for the five different regions. These were for the RSA (including the self-governing territories, but excluding the TBVC) and for the Transkei, Bophuthatswana, Ciskei, and Venda. To calculate the allocation of clusters nationally, the same interval was used for all five areas. This was necessary to get a self-weighted sample.

We have opted for approximately twenty-five households per cluster. Note: At this point in time, this figure should not be mentioned to any of the survey organizations. This figure was chosen given past experience in other surveys. It is arrived at by weighting the trade-off between efficiency of the data versus minimizing the cost of the survey. Efficiency requires as small a number of households per cluster as possible. Cost minimization requires as large a number of households per cluster as possible. There is some literature available on this issue, but we will refrain from going into any further detail here.

We know that our survey will cover roughly 9, 000 households. Our number of area sampling units or clusters selected is thus:

$$\alpha = \frac{m}{b_i}$$

where  $m$  is the fixed total sample size, and  $b_i$  is the number of households to be selected in the  $i$ -th ESD. In our case we are aiming for an average value of  $b_i$  which will be twenty-five households per cluster. The value of  $a$  is thus 360. Without loss of generality we can skip the first stage probability of selecting ESDs. Suffice to say that we will be selecting ESDs with probability proportional to size (PPS). Our first stage sampling interval is thus simply:

$$I_1 = \frac{N_{i,91}}{a}$$

where  $N_{i,91}$  is the total South African population in the last census (including TBVC). We obtained a rounded figure of 105, 800 for  $I_1$ . Given this as our fixed interval of selection, and our random start of 23, 237 we can easily calculate the number of clusters throughout South Africa (including the TBVC). The number of clusters per region is given in Table 1 below. Racial breakdown by province (excluding self-governing territories) is given in Table 2.

**TABLE 1.**

Area	Population	Number of Clusters
South Africa	38, 120, 853	360
1) Census (adjusted)	30, 986, 920	293
a) Provinces (excl. self-gov.)	20, 240, 416	192
i) Cape Province	6, 125, 335	58
ii) Natal	2, 430, 753	23
iii) Transvaal	9, 491, 265	90
iv) O.F.S	2, 193, 062	21
b) Self-Governing Territories	10, 746, 504	101
i) KwaZulu	5, 524, 774	52
ii) Kangwane	779, 240	7
iii) Qwa-Qwa	342, 886	3
iv) Gazankulu	954, 771	9
v) Lebowa	2, 740, 587	26
vi) Kwandebele	404, 246	4
2) Transkei	3, 457, 136	33
a) Urban	182, 534	2
b) Rural	3, 292, 602	31
3) Bophuthatswana	2, 250, 000	21
4) Venda	558, 797	5
5) Ciskei	850, 000	8
a) Urban	320, 000	3
b) Rural	530, 000	5
Provinces (incl. Self-Governing)		
1) Cape Province	6, 125, 335	58
2) Natal	7, 955, 527	75
3) Transvaal	14, 370, 109	136
4) O.F.S	2, 535, 948	24

Source: Own calculations, Population figures obtained from: CSS Census publications, consultations with various homeland governments, and Professor Stoker at the University of Pretoria

Note: The self-governing territories and provinces were amalgamated as follows:

- 1) Cape Province -- nothing added
- 2) Natal -- includes KwaZulu
- 3) Transvaal -- includes Kwandebele, Kangwane, Lebowa, and Gazankulu
- 4) Orange Free State -- includes Qwa-Qwa

**Table 2**

Area	Population	Number of Clusters
1) Census (inc. self-governing)	30, 986, 920	293
i. Whites	5, 068, 110	48
ii. Coloureds	3, 285, 718	31
iii. Asians	986, 620	9
iv. Africans	21, 646, 471	205
2) Provinces (exc. self-governing)	20, 240, 416	192
a) Cape Province	6, 125, 335	58
i. Whites	1, 348, 859	13
ii. Coloureds	2, 795, 415	27
iii. Asians	45, 465	0
iv. Africans	1, 935, 597	18
b) Natal	2, 430, 753	23
i. Whites	594, 865	6
ii. Coloureds	104, 006	1
iii. Asians	768, 695	7
iv. Africans	963, 186	9
c) Transvaal	9, 491, 265	90
i. Whites	2, 753, 076	26
ii. Coloureds	310, 897	3
iii. Asians	169, 340	2
iv. Africans	6, 257, 952	59
c) O.F.S	2, 193, 062	21
i. Whites	364, 679	3
ii. Coloureds	70, 594	1
iii. Asians	892	0
iv. Africans	1, 756, 879	17

Source: CSS, Population Census 1991: Adjustment for Undercount.

To get the number of clusters including the self-governing territories simply add the numbers from Table 1 onto the African population. The reason for this is that other population groups in these areas are rather small, and as we are selecting areas using the PPS method, it does not affect the efficiency of our sample in any way. See notes after Table 1 to see which self-governing territories is added to which provinces.

Please note that other documents exist in which full details regarding the methodology, and theory underlying our sampling frame is discussed. These are available upon request.

Peter D. Ellis  
University of Cape Town  
March 18, 1993