South African Census 1996:
Data Quality Comments

DataFirst

August 10, 2011

1 Some Eastern Cape Observations Internally Inconsistent

Certain observations within the Eastern Cape are designated either as households or institutions, but other items disagree for that observation. For example, the "hhrecord" item on certain observations is "H" (which means that the household questionnaire was administered), but other variables for that observation capture the response N/A: Institution, which directly contradicts "hhrecord". Furthermore, certain respondents who listed themselves as either being from an Institution or Hostel without being represented as such in other fields (i.e. the relationship to head of household variable should be N/A: Institution).

There are 136 observations of the former, and 177 of the latter. All of these observations are in the Eastern Cape, which could suggest a problem with some individual enumerator or enumerator training for the broader area given the proximity of these anomalous data entries. Please note that this speculation is forwarded tentatively, and DataFirst remains agnostic on the cause of these seemingly contradictory data entries.

Figure 1: Geolocality of Internally Inconsistent Observations

2 'Occupation' and 'Occupation last worked' variables

There are two observations for which the "occupat" variable (Occupation) and one observation for "lstwrk" variable (Occupation last employed) in which the value (960) has no corresponding code, neither in the data file nor the meta data provided by Statistics South Africa.

3 Unemployed in Industries

The "industr" (Industry) variable has all unemployed respondents lumped in the same category (Transport, communication and storage). Furthermore, all respondents from institutions are also coded as being in the same, arbitrary category (Financial intermediation, except insurance and pension). Most of the latter observations are excluded from the merged dataset (which only includes household questionnaire respondents), with the observations from the Eastern Cape detailed in the first data quality issue being the exception for the reasons outlined therein.

	Industry		
Employment status	Transport	Financial	Total
Unemployed, looking for work	416,125	0	416,125
Not working - not looking for work	76,762	0	76,762
Not working - housewife/home-maker	176,114	0	176,114
Not working - scholar/full-time student	436,142	0	436,142
Not working - pensioner/retired person	234,363	0	234,363
Not working - disabled person	27,441	0	27,441
Not working - not wishing to work	27,188	0	27,188
Not working - none of the above	101,426	0	101,426
N/A: Aged <15	1,216,291	0	1,216,291
N/A: Institution	0	136	136
Unspecified	6,863	0	6,863
Total	2,718,715	136	2,718,851

Source: South African 1996 Census - 10 % Sample (v01.2)

4 Studying Status Variable

Some responses to this question contradicted responses given to a later question. That is, some persons recorded as not studying for question 16.3 were recorded as being scholar/full-time student for question 18.1 on usual activity. The cause of this problem is not clear but respondents could have mistakenly thought that the question on studying applied only to post-school institutions as it appeared after the question on post-school qualifications. Where responses contradicted, the response to the second question was considered to be correct and the variable Studying status was amended accordingly.

However, if users would prefer the unedited data, this is available from Stats SA as a special request. This information on data quality was taken directly from the original metadata made available by Stats SA.