

USER GUIDE

SOUTH AFRICAN SOCIAL ATTITUDES SURVEY (SASAS) 2011

HSRC Study ID
SASAS 2011 Q1

Human Sciences Research Council
2014

Data Curation



HSRC
Human Sciences
Research Council

Table of Contents

1. Introduction.....	1
2. Available data set.....	1
2.1. Program code to generate the different statistical data sets mentioned above:	1
3. Conversion of data	2
3.1. Additional notes on SPSS.....	2
3.2. Additional notes on SAS.....	3
3.3. Additional notes on STATA.....	4
3.4. Other formats:	4
4. Data and documentation notes	4
4.1. Deleted / recoded / derived variables.....	4
4.2. Missing values within the data sets	4
4.3. Variables and labels within the data set	4
4.4. Weighting of the data	4
4.5. Coding of occupation	4
APPENDIX A.....	5
APPENDIX B.....	7

South African Social Attitudes Survey (SASAS) 2011

1. Introduction

The primary objective of SASAS is to design, develop and implement a conceptually and methodologically robust study of changing social attitudes and values in South Africa. In meeting this objective, the HSRC intends to position itself so as to be able to carefully and consistently monitor and explain changes in attitudes amongst various socio-demographic groupings.

It is intended that SASAS will explore a wide range of value changes, including the distribution and shape of racial attitudes and aspirations, attitudes towards democratic and constitutional issues, and the redistribution of resources and power.

Moreover, there is also an explicit interest in mapping changing attitudes towards some of the great moral issues that confront and are fiercely debated in South Africa, for instance abortion, gender issues, AIDS, crime and punishment, governance, service delivery, etc.

In sum, then, SASAS is intended to provide a unique long-term account of the social fabric of modern South Africa, and of how its changing political and institutional structures interact over time with changing social attitudes and values.

The core module will remain constant from round to round, with the aim of monitoring change and continuity in a variety of socio-economic and socio-political variables. In addition, a number of themes will be accommodated on a rotational basis. This rotating element of the survey consists of two or more topic-specific modules in each round of interviewing and is directed at measuring a range of policy and academic concerns and issues that require more detailed examination at a specific point in time than the multi-topic core module would permit. Fieldworkers were required to complete a paper-based instrument while interviews were conducted face-to-face.

Thematic content:

Questionnaire 1: Topics covered in the questionnaire are: democracy and governance, national identity and pride, intergroup relations, education, moral issues, crime and safety, personal wellbeing index, poverty, police confidence, tolerance, corruption, Batho Pele, voting, respondent characteristics, household characteristics, personal and household income variables.

Please refer to the [Study information](#) document on the web site for a detailed description of the study.

2. Available data set

SPSS data set:

SASAS2011_Q1.sav (n=3057); 447 variables, 'benchwgt' (weight variable)

STATA data set:

SASAS2011_Q1.dta (n=3057); 447 variables, 'benchwgt' (weight variable)

SAS data set and catalog:

SASAS2011_Q1.sas7bdat (n=3057); 447 variables, 'benchwgt' (weight variable),
SASAS2011_Q1.sas7bcatalog

ASCII data sets:

Fixed format

SASAS2011_Q1.dat (n=3057)

Comma Separated Values data file (csv)

SASAS2011_Q1.csv (n=3057); 447 variables, 'benchwgt' (weight variable)

2.1. Program code to generate the different statistical data sets mentioned above:

SPSS:

SASAS2011_Q1.sps

SAS:
SASAS2011_Q1.sas
STATA:
SASAS2011_Q1.do, SASAS2011_Q1.dct (dictionary)

3. Conversion of data

Additional comment about each data format is given below, please read the wording corresponding to the data format you have downloaded:

SPSS:
If SPSS has not been the ingest format this format has been created according to StatTransfer standards. Three files are created:
ASCII data file (.dat), SPSS syntax file (.sps) and an SPSS data set (.sav).
Use the SPSS syntax file to convert data to an SPSS data set (if not provided). Be sure to change the path in the DATA LIST FILE= statement to point to where you stored the ASCII data file.

SAS:
If SAS is not the ingest format, SAS files are created according to StatTransfer standards. Three files are created:
ASCII data file (.dat), SAS program file (.sas) and SAS data set (.sas7bdat). The formats.sas7bcat file contains the formats of the variables and must be used with the data set as illustrated under the heading: 'Additional notes on SAS'.
In the SAS program file, change the infile statement to point to where you stored the ASCII data file. To create a SAS data set (if not provided) open the SAS program in the SAS program editor and submit. **If you do not want the value labels permanently associated with the variable values in the data set, comment out the format statement in the data step and rather use it in the PROCS.**

STATA:
If STATA is not the ingest format, STATA files are created according to StatTransfer standards. Four files are created:
ASCII data file (.dat), Stata program file (.do), dictionary file (.dct) and a Stata data file (.dta).
To create a Stata data set (if not provided) open the .do program in (e.g. Stata editor) and change the path name to point to where you stored the ASCII data and the .dct (dictionary) files. Also indicate the

The data set contains a mix of ordinal, nominal and scale (continuous) variables. (See Appendix B for an explanation.)

3.1. Additional notes on SPSS

When using SPSS please note the following: if statistical analysis, other than descriptive statistics is done, the measurement level for variables should be set to the appropriate level as per the example below.

Example: Measurement Level syntax (SPSS):

```
VARIABLE LEVEL  
V1 V3 V8  
(SCALE)  
/ V4 V7 V11 V13  
(NOMINAL)  
/ V15 to V21  
(ORDINAL)
```

3.2. Additional notes on SAS

In the downloadable SAS program the formats were stored in a SAS catalog: **SASAS2011_Q1.sas7bcat**. This was the result of using the **library** option in proc format and putting the format statement to associate the value labels with the variables, in the data step. Whenever the SAS data set is used, a libname statement must indicate where the permanent SAS data set and the format catalog are located. Store them in the same folder.

Example: Referencing the permanent SAS data set and the format catalog

```
Libname library 'full path where the data set and catalog are stored e.g. c:/myfolder';
PROC FREQ DATA=library.sasas2011_q1;
Run;
```

*Please note that there is no need to specify the formats catalog if it is **formats.sas7bcat** (default) as long as it is in the same folder as the data set.*

*If however the format has a name such as **SASAS2011_q1.sas7bcat**, you need to put in additional statements (see red below).*

```
libname library 'c:/myfolder' ;
OPTIONS FMTSEARCH=(library.SASAS2011_q1); (This is to locate the formats)
PROC FREQ DATA=library.SASAS2011_q1; (This is for the SAS data set SASAS2011_q1.sas7bdat)
```

Note:

When searching for a format or informat, SAS always searches in WORK.FORMATS first, and then LIBRARY.FORMATS, unless one of them appears in the FMTSEARCH= list. SAS searches the catalogs in the FMTSEARCH= list in the order that they are listed until the format or informat is found.

If you want the values and not the formats in the data set, then use the following statements to disassociate the formats from the variables in the data set. You can then run the proc format separately and associate the variables with the formats in a PROC.

WARNING: If the value labels are in the data set, SAS will do the analysis on the value labels which are text instead of the numeric values.

Example: Getting rid of the formats in the data set

The following code will remove all formats associated with the variables in the SAS data set.

```
Libname library 'full path where the data set and catalog are stored e.g. c:\myfolder';
DATA library.newdsn;
SET library.dsn;
FORMAT _all_;
RUN;
```

- *library* refers to name you used in the libname statement.
- *dsn* is the name of the SAS data set, whose formats need to be removed.
- *newdsn* is the name of the new SAS data set with the formats removed.
- *format _all_* will delete all the formats in the SAS data set.

The following code will remove only the formats of specific variables in the SAS data set. Specify the variables without their formats in the FORMAT statement. This is also applicable to PROCs.

```
DATA library.newdsn;
SET library.dsn;
FORMAT q1_1 q2 q3;
RUN;
```

3.3. Additional notes on STATA

Q1:

In the analysis of the data, the multilevel complex design of the survey should be taken into account. The following design variables should be specified when analysing these data:

- the weight variable is **benchwgt**
- the strata are **province, geotype, race (q254)**
- the primary sampling unit is **ea**

3.4. Other formats:

If you do not have any of the above mentioned statistical software you can use the Comma Separated Values data file (.csv) as it is compatible with most software. The variable and value labels are provided in any of the program files (.sas, .sps or .do) which can be opened in any text editor. The program files can also be used to locate the positions of the variables if you want to use the .dat fixed format file.

4. Data and documentation notes

4.1. Deleted / recoded / derived variables

Certain variables that were not relevant to the research were deleted from the data sets.

4.2. Missing values within the data sets

Missing values in the data are indicated as 'System missing' denoted by a dot (.).

Inconsistencies and missing values in the data are due to questions that had to be skipped depending on the answer provided at a previous question and follow-up question(s) were left blank.

Missing values also occurred when respondents refused to answer particular questions or due to fieldworker inaccuracy.

4.3. Variables and labels within the data set

The following code book may be consulted for the variables and variable and value labels within the data set:

SASAS 2011 Q1 Code book_b.pdf

4.4. Weighting of the data

The metadata record provides a detailed description of the weighting procedure that was followed for the SASAS 2011 survey.

The weighting variable to be used in analysis: 'benchwgt' for the data set.

The sample is weighted to represent the South African population. With certain exceptions analysis should be performed on weighted data.

4.5. Coding of occupation

The occupation was coded to two digits on the basis of the International Standard Classification of Occupations (ISCO 88), URL: www.ilo.org/public/english/bureau/stat/isco/isco88/major.htm

APPENDIX A

Analysing multiple response questions

(Please note that this is an example from a different data set)

If you want to analyse multiple response questions to get an overview of similar responses, you can use SPSS multiple response set functionality under *Analyze, Multiple response, Define Variable Sets*:

The screenshot shows the IBM SPSS Statistics Data Editor with the 'Define Variable Sets' dialog box open. The dialog box has two main sections: 'Set Definition' and 'Multiple Response Sets:'. In the 'Set Definition' section, the 'Variables in Set:' box contains three variables: 'Q1.1 The 3 MOST important challenges facing SA', 'Q1.2 The 3 MOST important challenges facing SA', and 'Q1.3 The 3 MOST important challenges facing SA'. The 'Variables Are Coded As' section has 'Categories' selected, with a range from 1 to 98. The 'Name' field is 'q1_mr' and the 'Label' is 'Q1 Most important challenges facing SA'. The 'Multiple Response Sets:' box on the right contains the variable '\$q1_mr'. The background shows a data table with columns for Name, Type, Label, Values, Missing, Columns, Align, and Measure.

	Name	Type	Label	Values	Missing	Columns	Align	Measure
1	uniqueid	Numeric	Questionnaire Number	None	None	8	Right	Scale
2	ea	Numeric	EA number	None	None	8	Right	Scale
3	prov	Numeric	Province	{1, WC}...	None	8	Right	Nominal
4	hhnum	Numeric	Number of households at visiting point	None	None	8	Right	Scale
5	pers16	Numeric	Number of persons 16 years and older at ...	None	None	8	Right	Scale
6	hhpers	Numeric	Number of persons in this household	None	None	8	Right	Scale
7	hhpers16	Numeric	Number of persons 16 years and older in t...	None	None	8	Right	Scale
8	q1_1	Numeric	Q1.1 The 3 MOST important challenges fa...	{1, HIV or AI...	None	8	Right	Nominal
9	q1_2	Numeric	Q1.2 The 3 MOST important challenges fa...	{1, HIV or AI...	None	8	Right	Nominal
10	q1_3	Numeric	Q1.3 The 3 MOST important challenges fa...	{1, HIV or AI...	None	8	Right	Nominal
11	q1_rec	Numeric	Q1 The 3 MOST important challenges faci...	{2, ABIDING...	None	10	Right	Nominal
12	q2	Numeric	Q2 In last 5 years, has life improved or no...	{1, Improved...	None	8	Right	Ordinal
13	q3	Numeric	Q3 Do you think tha...	{1, Improve}...	None	8	Right	Ordinal
14	q4	Numeric	Q4 How satisfied are you w...	{1, Very sati...	None	8	Right	Ordinal
15	q5	Numeric	Q5 How satisfied a...	{1, Very sati...	None	8	Right	Ordinal
16	q6	Numeric	Q6 How satisfied a...	{1, Strongly ...	None	8	Right	Ordinal
17	q7	Numeric	Q7 Indicate the extent you trust the follow...	{1, Strongly ...	None	8	Right	Ordinal
18	q8	Numeric	Q8 Indicate the extent you trust the follow...	{1, Strongly ...	None	8	Right	Ordinal

The screenshot shows the 'Define Multiple Response Sets' dialog box. The 'Set Definition' section has a list of variables on the left and a 'Variables in Set:' box on the right. The 'Variables in Set:' box contains three variables: 'Q1.1 The 3 MOST important challenges facing SA', 'Q1.2 The 3 MOST important challenges facing SA', and 'Q1.3 The 3 MOST important challenges facing SA'. The 'Variables Are Coded As' section has 'Categories' selected, with a range from 1 to 98. The 'Name' field is 'q1_mr' and the 'Label' is 'Q1 Most important challenges facing SA'. The 'Multiple Response Sets:' box on the right contains the variable '\$q1_mr'. The 'Add', 'Change', and 'Remove' buttons are visible. A note at the bottom states: 'Note: Sets defined here are only available in the Multiple Response Frequencies and Crosstabs procedures. Use Define Multiple Response Sets on the Data menu for sets used elsewhere.'

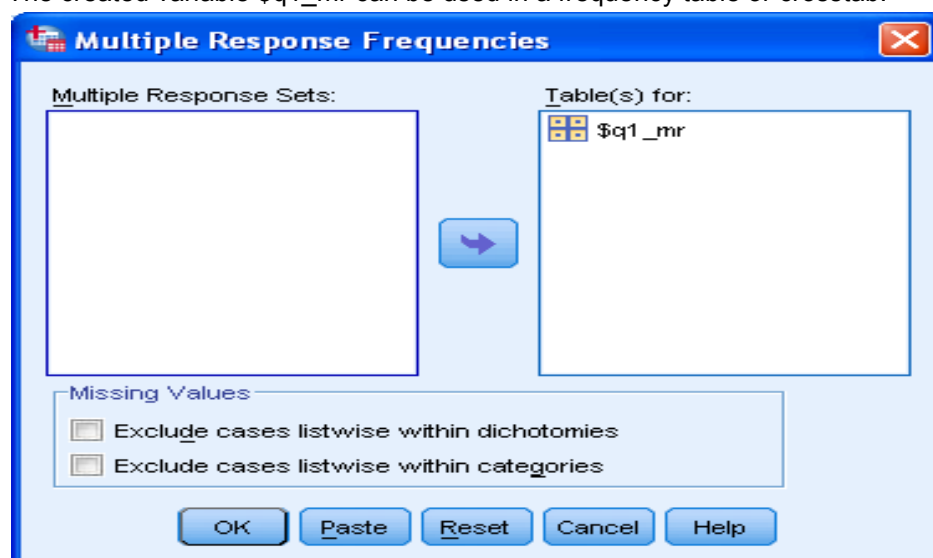
Select the variables you will need for the multiple response set: e.g. Q1_1, Q1_2, Q1_3 and move to *Variables in set:* box by ticking the ► arrow.

In this illustration *Categories* will be selected, as answer options in Q1.1 to Q1.3 vary from code 1 to code 19 and code 98. (In the case where variables are coded as 'Yes/No', *Dichotomies* will be selected and *Counted value* will be assigned in the box.)

Assign a variable name in the *Name* box (q1_mr) and a description in the *Label* box (Q1 Most important challenges facing SA). Select *Add*, *Close*.

To run a frequency or cross tabulation: Analyze, Multiple response, select Frequencies....or Crosstabs....

The created variable \$q1_mr can be used in a frequency table or crosstab.



Example of output for multiple response questions:

\$q1_mr Frequencies				
Q1 Most important challenges facing SA ^a		Responses		Percent of Cases
		N	Percent	
	HIV-AIDS	1437	14.7%	43.5%
	Unemployment	2361	24.2%	71.5%
	Racism	146	1.5%	4.4%
	Xenophobia	118	1.2%	3.6%
	Crime and safety	1651	16.9%	50.0%
	Service provision or delivery	454	4.7%	13.7%
	Affordable housing	324	3.3%	9.8%
	Land reform issues	67	0.7%	2.0%
	Human rights	98	1.0%	3.0%
	Education	289	3.0%	8.7%
	Economic and financial issues	291	3.0%	8.8%
	Work-related issues	78	0.8%	2.4%
	Family and youth issues	86	0.9%	2.6%
	Religion and culture issues	30	0.3%	0.9%
	Environmental issues	68	0.7%	2.1%
	Political issues	149	1.5%	4.5%
	Corruption	657	6.7%	19.9%
	Poverty	1311	13.5%	39.7%
	Other	109	1.1%	3.3%
	Do not know	22	0.2%	0.7%
Total		9746	100.0%	295.0%

a. Group

Note that percentages do not add up to 100.0% in the 'Percent of Cases' column as one person could select more than one answer option. (The above is an example of a previous SASAS survey and calculations were not performed on weighted data.)

APPENDIX B

Nominal

A variable can be treated as nominal when its values represent categories with no intrinsic ranking; for example, the department of the company in which an employee works. Examples of nominal variables include region, zip code, or religious affiliation.

Ordinal

A variable can be treated as ordinal when its values represent categories with some intrinsic ranking; for example, levels of service satisfaction from highly dissatisfied to highly satisfied. Examples of ordinal variables include attitude scores representing degree of satisfaction or confidence and preference rating scores. For ordinal string variables, the alphabetic order of string values is assumed to reflect the true order of the categories. For example, for a string variable with the values of low, medium, high, the order of the categories is interpreted as high, low, medium which is not the correct order. In general, it is more reliable to use numeric codes to represent ordinal data.

Scale

A variable can be treated as scale when its values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate. Examples of scale variables include age in years and income in thousands of dollars.