

Activity #2
“Managing Survey Data”

This activity demonstrates many of the steps necessary to prepare a publicly available survey dataset for analysis. Many of the steps are also relevant to preparing other forms of data (e.g., administrative, population/census). There are three datasets in this activity which contain a subset of the variables from the SESRI Omnibus survey data:

- SESRI_Omnibus_2011_excerpt1.dta: the main dataset for this activity.
- SESRI_Omnibus_2011_excerpt2.dta: the dataset with activities completed.
- SESRI_Omnibus_2014_excerpt2.dta: additional data that will be combined with the main dataset.

1. Launch Stata
2. Create a log file
log using <filename.log> or use drop-down menu
3. Open SESRI_Omnibus_2011_excerpt1.dta
use <filename> or use drop-down menu
4. Open Data Browser
5. Inspect the data and compare to codebook
6. Identify variable names
7. Close Data Browser
8. Calculate summary statistics for each variable
summarize <varname>, detail
9. Tabulate the values of *campaign, yob, household, accident_cause, gender, saloons, suvs, trucks, accident*
tab <varname>
10. Recode missing values
 - a. Recode coding error on *campaign* to missing
gen <newvar>=<varname>
recode <newvar> (oldval=.)

b. Recode 'don't know'/'no response' on *yob* to missing
gen <newvar> = <varname>
replace <newvar> = . if <newvar>==(val)

11. Recode into binary variables

a. Reverse code (switch values on *household*)

gen <newvar> = <varname>
recode <newvar> (oldval=newval) (oldval=newval)

b. Transform categorical variable into binary variable (*accident_cause_roads*)

gen <newvar> = <varname>
recode <newvar> (oldval = 1) (oldval = 0)

c. Recode to '0' and '1'

gen <newvar> = <varname>
recode <newvar> (1=0) (2=1)

12. Add value labels for *accident_cause_roads*

label define yesno 1 "yes" 0 "no"
label values <varname> yesno

13. Generate new variable from two (or more) others (*number_cars = saloons + suvs + trucks*)

tab <varname1> <varname2> <varname3>
egen <newvar> = rowtotal (<varname1> <varname2> <varname3>)

14. Append SESRI_Omnibus_2014_excerpt2.dta

use <filename1>
append using <filename2>
summarize <varname>
save <filename_combined>

15. Collapse data across observations

collapse (statistic) <varname>, by(group1, group2)
summarize <varname>

16. Log off

close <filename.log> or use drop-down menu