Stata vs SUDAAN: Which does what?

Stata (unlike SUDAAN) is a general purpose statistical package that performs data management, graphics and statistical analysis, including many advanced modeling features and post estimation commands. Stata is particularly adept at creating dummy variables and interaction terms within the model statement rather than requiring a separate data step. Stata also does not require the re-coding of 0/1 variables or the specification of the number of levels for tables. In addition, Stata handles survey data from complex multi-stage designs, including sub-populations and DEFF using SRS within the sub-population. The specification of the sample design in Stata is somewhat different than in SUDAAN, requiring a finite population correction factor (fpc) rather than specifying a without replacement design. This requires knowing either the sampling proportion or the population total (within strata) to enter the fpc.

Survey Analysis Features in SUDAAN and Stata

Feature	SUDAAN	Stata
Means	Descript	svymean
Proportions	Crosstab	svyprop
Ratios	Ratio	svyratio
Totals	Crosstab	svytotal
Linear Regression	Regress	svyreg
Logistic Regression	Logistic	svylogit
Probit		svyprobt
Hypothesis tests	Hyptest	svytest
Linear combinations of coefficients		svylc
Proportional Hazards	Survival	
Repeated measures	Multilog	svyolog, svyoprob, (xtgee?)
Contingency Tables	Catan	

Stata traits:
Stata has a number of distinctive advantages over SUDAAN, but also a few disadvantages. The advantages are a probit model, the calculation of a within sub-population design effect, and not having to recode 0/1 variables for tables. It also does not require the user to specify the number of levels a variable has. In addition, svylc will calculate the difference between two estimate by class and perform the appropriate significance test. The syntax for defining the sample design is also simpler, and doesn't require sorting. Some sample designs may not be exactly replicable, however. Stata is also extremely fast for data sets that fit in real memory (as opposed to virtual memory), and Stata allows several levels of data typing including byte, int, long and float, thereby conserving space. Stata also has an easily extensible design, allowing users to create their own likelihood functions, optimization routines, and even graphics routines. This makes Stata very versatile for programming new features. In fact, there is a Stata listserv where users (and people at Stata Corporation) post new features on a regular basis that can be downloaded and added to the

user's program.

The major disadvantage is lack of control over the format of output. To a certain extent, output can be saved as a Stata data set and printed, but by and large one has to accept the output as Stata produces it.