Subject: Re: Performing a manual backward stepwise logistic regression in Stata
Posted by Bridgette-DHS on Tue, 06 Sep 2016 16:44:53 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I have revised some of your Stata code below. Rather than manual stepwise selection I would use the Stata command "stepwise". There are limitations with this command. It does not work with "i." for categorical variables, so you must form sets of dummies with "xi". Then if one of the dummies is dropped from the model, it means that the category has been consolidated with the reference category. It also does not work with svyset, but it DOES work with the pweight and cluster specifications within an estimation command, so all you are missing is the stratum adjustment and subpop, which you can add after you have selected the final model. In the example, I am setting a p value of .1 as the threshold for retention in the model. This number can be changed. I would repeat the procedure with various values ranging from, say, .5 to .05. I personally prefer logit to logistic, but have used logistic below since that's what you had.

```
use e:\DHS\DHS_data\PR_files\UGPR60FL.dta

generate overfive=.
replace overfive = 1 if hv105>=5 & hv105<96 & hv103==1
replace overfive = 0 if overfive!=1

* construct onedisability
describe sh24-sh29
tab1 sh24-sh29,m

gen sh24r=0
replace sh24r=1 if sh24>=2 & sh24<=4
replace sh24r=. if sh24==.

gen sh25r=0
replace sh25r=1 if sh25>=2 & sh25<=4
replace sh25r=. if sh25==.

gen sh26r=0
replace sh26r=1 if sh26>=2 & sh26<=4
replace sh26r=. if sh26==.

gen sh27r=0
replace sh27r=1 if sh27>=2 & sh27<=4
replace sh27r=. if sh27==.

gen sh28r=0
replace sh28r=1 if sh28>=2 & sh28<=4
replace sh28r=. if sh28==.

gen sh29r=0
```

```
replace sh29r=1 if sh29>=2 & sh29<=4
replace sh29r=. if sh29==.

egen onedisability=rowtotal(sh24r-sh29r), missing
replace onedisability=1 if onedisability>1 & onedisability<.

tab onedisability,m

*CONDUCTING MULTIVARIATE LOGISTICS REGRESSION
*Where 'onedisability' is the variable for difficulty in at least one functionalarea

* specify svyset, but stepwise cannot work with svyset.  Use pweight and cluster, and do not
* include strata and subpop until you get to the final model.

* stepwise cannot handle i. for categorical variables; must form sets of dummies

xi, prefix(v_) i.hv024 i.hv025 i.hv104 i.hv106 i.hv270

logistic onedisability v_* [pweight=hv005], cluster(hv021)

stepwise, pr(.1): logistic onedisability v_* [pweight=hv005], cluster(hv021)
```