

---

Subject: Women's occupation variable NFHS 5  
Posted by [Dibyasree](#) on Thu, 30 Mar 2023 05:38:34 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I would like to know how we can use the variable v717 (respondent's occupation) in the analysis for NFHS-5. From the NFHS-5 report, it is clear that v717 has only been collected at the state level, a sub-sample of 15% of households were selected for the implementation. 85% of the response has been coded as missing ("."). Can you confirm how we can use v717 as an independent variable in the regression? "svy: reg dependentvar v717 controlvars if v717!=."; Otherwise, how to tackle the problem of large missing values in the independent variable?

Also, can you please confirm the use of weight for the same analysis? Is it correct to use svyset using v005 for the analysis [svyset [pw=v005], psu (v021) strata(v023)singleunit(centered)]?

---

---

Subject: Re: Women's occupation variable NFHS 5  
Posted by [Bridgette-DHS](#) on Thu, 30 Mar 2023 15:01:25 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS staff member, Tom Pullum:

A dot in Stata means that a response is Not Applicable (NA) because the respondent was not asked the question. If your table or regression includes a variable for which some cases are NA, those cases will be excluded. There is no way to get around that. The variable can be an outcome, a predictor, etc., but the cases that are NA on that variable will be dropped entirely.

Regarding your svyset command, v005 is the correct weight (for the IR, KR, or BR files). However, in the India surveys, clusters (either v021 or v001) are numbered 1, 2, etc., within states (v024). For the psu you have to construct a new variable such as "egen clusterID=group(v024 v021)" and then "svyset [pw=v005] psu (clusterID), strata(v023) singleunit(centered)". (The format for svyset is pretty flexible, but I believe strata is/are the only component that is expected after the comma.)

Sometimes I forget to include hv024 and re-number the clusters for the India surveys when using svyset--sorry if I have done that on the forum. This renumbering does not have to be done for the strata.

Correction: the clusters are NOT numbered 1, 2, etc., within states. Senior DHS staff member, Fred Arnold has pointed out that the original numbering of the clusters was 1, 2, etc. within districts rather than states. The cluster numbers in the data files were then constructed by concatenating the district code with that original cluster number. For example, looking at the PR file, state 3 (hv024==3) and district 36 (shdist=36), there are 42 clusters, numbered 3601 through 3642. These numbers are distinct within state, but will come up again in different states. If you use the "egen group" command as I gave it, cluster 3601 in state 3 (Punjab) would become 33601, and so on, so all the clusters in the survey will have a unique ID. "egen clusterID=group(v024 v021)" is correct (in the IR file) but the explanation I gave was a little off.

---

Subject: Re: Women's occupation variable NFHS 5  
Posted by [Dibyasree](#) on Fri, 31 Mar 2023 05:19:11 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you so much for your response. This is very helpful.

---