

---

Subject: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [drsanmis](#) on Wed, 07 May 2014 14:45:19 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

I am looking at a categorical outcome in the NFHS 3 data. Do I need to use PROC LOGISTIC or PROC SURVEY LOGISTIC in SAS. (Published papers have used either of them).

If proc survey logistic is used the syntax requires using STRATA and CLUSTER statements. I am not sure which variables in the dataset signify strata and cluster.

It will be very helpful if someone can guide me on this.

---

---

Subject: Re: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [Bridgette-DHS](#) on Thu, 15 May 2014 12:58:18 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Our SAS Specialist is reviewing your posting, and we will get back to you soonest.

Thanks.

---

---

Subject: Re: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [Bridgette-DHS](#) on Thu, 15 May 2014 13:21:18 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from our Senior SAS Specialist, Ruilin Ren.

For the use of the logistic regression, it really depends on your purpose of analysis. If you are interested in exploring patterns of the survey data themselves, but do not intend to extrapolate the entire survey population, you can use proc logistic. But if your intention is to explore patterns among the entire survey population, you need to use proc survey logistic. In the later case, you need to declare CLUSTER (variable HV001 at household level, V001 at individual level), STRATA (HV022 at household level, V022 at individual level). You also need to declare the sampling weight WEIGHT (HV005 divided by 1000000 at household level, and V005 divided by 1000000 at individual level).

---

---

Subject: Re: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [ABLR](#) on Sun, 03 Aug 2014 02:22:22 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Just to doublecheck, the strata variable would be v022 and not v023 or a combination of v024 and

v025? Isn't it unusual in DHS data for v022 to be the strata variable. [I'm using Stata, but I assume this wouldn't differ across software.] Thanks for checking!

---

---

Subject: Re: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [Bridgette-DHS](#) on Tue, 19 Aug 2014 15:45:45 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Another response from Ruilin Ren:

As for stratification variables for NFHS-3, the stratification should be HV024 (V024) crossing HV025 (V025). HV022 (V022) is not the stratification variable.

---

---

Subject: Re: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [Kathryn Kershaw](#) on Thu, 09 Jul 2015 10:20:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

My query is clarification of this one. In Stata I used the following commands to survey set my data:

```
gen wgt = v005/1000000  
svyset v021 [pweight = wgt], strata(v023)
```

But I found this thread and thought maybe I should create a new strata variable:

```
egen strata = group(v024 v025)
```

This produces a different variable from v023 with separate groups for the rural and urban aspects of each state, which v023 doesn't have.

Could you clarify which one I should use please?

Many thanks!

Kathryn

---

---

Subject: Re: NFHS 3 data - what variables define STRATA and CLUSTER for SAS analysis

Posted by [Reduced-For\(u\)m](#) on Thu, 09 Jul 2015 21:15:37 GMT

---

Check out the country report for the year/country you are using. It should tell you how the sampling was done (see the intro or the appendix). From that, you should be able to determine which the correct stratification is - it is not the same for every country, but my experience is that the stratification by regionXurban [group(v024 v025)] is the most commonly used.

---