Subject: Help-Seeking Behaviour of Domestic Violence_Myanmar DHS (2015-16)
Posted by Anonymous on Wed, 31 Jul 2024 11:26:48 GMT
View Forum Message <> Reply to Message

Dear Moderator,

May I discuss two questions regarding "Weighting"?

(1) I am doing DHS data analysis using "R". Before going data analysis, I did weight the selected variables after managing missing values and mutating some variables. Then, I compared the results of unweighted and weighted frequencies. The unusual differences in 7 categories were found in one variable "v101" which had 15 categories. For example, (n = 60) before weight and (n = 5.11) after weight. I checked again and again. Also, I checked in Stata. It showed similar. Is it possible? Should I weight or not? I used R codes as below:

IRdata$wt<- IRdata$v005/1000000
mysurvey <- svydesign(id = IRdata$v021, data = IRdata, strata = IRdata$v022,
            weights = IRdata$wt, nest = T)
options (survey.lonely.psu = "adjust")
svytable(~ v101, mysurvey)

(2) In relation to weighting logistic regression analysis. Some researchers say that it should be weighted all analysis (descriptive, bivariate, and multivariate analysis). However, some researchers say that it should be considered for descriptive data analysis only. Please kindly give me your suggestion.

Thank you,
Soe

Subject: Re: Help-Seeking Behaviour of Domestic Violence_Myanmar DHS (2015-16)
Posted by Janet-DHS on Wed, 31 Jul 2024 18:33:34 GMT
View Forum Message <> Reply to Message

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

You are using the weights correctly. v101 is a duplicate of v024--both are "region"--and region x urban/rural defines the strata in the sample design. In the design of the sample, the small strata were over-sampled and the large strata were under-sampled, relatively speaking. If you look at the unweighted distribution of v101, the number of cases is nearly the same in each region, varying between about 751 and 1039. However, the weighted distribution, which is proportional to the actual size in the population, ranges from about 65 to 1649. By sampling approximately the same number of cases in each region, the sample is made more efficient. The standard errors of estimates are relatively equal across regions or strata. I repeat that what you observe is ok. (For most other variables, weighted and unweighted frequencies are typically within 10% to 20% of each other.)

We recommend using weights for all estimates. Indeed, we recommend a full adjustment for the sampling design, taking into account clustering, stratification, and weights, for everything except some initial exploratory analysis. If you don't use weights, it becomes almost impossible to understand changes and differences between surveys and sub-populations.

The argument against weights comes from some econometricians who believe that their models are correctly specified and whose main interest is in test statistics and p-values. I believe that statisticians and demographers uniformly advocate the use of weights to compensate for variation in sampling fractions in complex surveys. With cross-sectional DHS data, our models cannot be correctly specified and we are usually more interested in estimates (including confidence intervals) than in tests.

Subject: Re: Help-Seeking Behaviour of Domestic Violence_Myanmar DHS (2015-16)
Posted by Anonymous on Thu, 01 Aug 2024 12:37:02 GMT
View Forum Message <> Reply to Message

Thank you, for your prompt response in detail.

Subject: Re: Help-Seeking Behaviour of Domestic Violence_Myanmar DHS (2015-16)
Posted by Anonymous on Thu, 01 Aug 2024 14:32:58 GMT
View Forum Message <> Reply to Message

Let me continue follow up questions.

As I am working a sub-sample, the sample size for my study, 811 before weighting, it has changed 690 after weighting. Is it required to use weight for multivariate analysis?

Subject: Re: Help-Seeking Behaviour of Domestic Violence_Myanmar DHS (2015-16)
Posted by Janet-DHS on Wed, 07 Aug 2024 14:28:25 GMT
View Forum Message <> Reply to Message

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

Yes, you should use weights, as well as the other adjustments in svyset for clustering and stratification. There has been much discussion of weights on the forum.

The change from 811 to 690 is plausible, but I cannot confirm it without knowing how you defined your subsample. Use d005, not v005, when analyzing the DV data.