

---

Subject: Re: Balancing number of observations for the dependent and independent variables

Posted by [Janet-DHS](#) on Wed, 08 Mar 2023 21:47:58 GMT

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

---

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

I believe you are doing several different regressions, and are getting different sample sizes (n's). This can happen because the different variables may have different numbers of cases that are not applicable or are automatically excluded for different reasons. If you want all the models to have the same number of cases, then you have to define a variable "varsmismissing" (for example) that is coded "1" if a case is dropped from ANY of the models and "0" otherwise. Then you re-run the models with a line "if varsmismissing==0". There are alternative ways to do this, for example with "svy: subpop(X)". (If you do it with subpop, the variable X in parentheses should be 1 if you want to KEEP the case, the reverse of the coding I suggested for "varsmismissing".)

There are advantages to having the same n for several models, for example if you want to test one model against another. But if you lose a lot of cases from just one or two of your variables, it may be preferable to drop the variable and keep the cases.

---