
Subject: Variance Inflation Factor

Posted by [nwegbus](#) on Tue, 31 Jan 2017 00:21:45 GMT

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

I'm working with the domestic violence module of Nigeria DHS (2013) using Stata 13.

I'm trying to examine my predictor variables for collinearity using the VIF score. Please see the attached document for all the output (you might want to zoom to 125%).

Here;s a summary of what I did:

I weighted my data using the following command:

```
generate wgt5 = d005/1000000  
svyset [pweight = wgt5],psu(v021) strata(v022) singleunit(centered)
```

Then I ran a regression model of my predictor variables using this command:

```
svy: logit exipv agemarr4 religion weduc say attwb wipv polyg hseek chcomm
```

Then I issued the VIF command:

```
display "tolerance = " 1-e(r2) " VIF = " 1/(1-e(r2))
```

However this is the output I got:

```
tolerance = . VIF = .
```

It seems that the tolerance and VIF are missing? I'm wondering what I'm doing wrong, as I don't know what to make of a missing VIF score. Thanks in advance for your help.

SN

File Attachments

1) [Stata output.docx](#), downloaded 533 times

Subject: Re: Variance Inflation Factor

Posted by [Bridgette-DHS](#) on Tue, 31 Jan 2017 22:16:29 GMT

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

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

Sorry, but this question is not about DHS data or even about Stata. I am not familiar with this procedure and don't know what it does.

I do notice that your logit regression seems to treat all variables as interval-level, whereas most of them are categorical and should be preceded by "i.". And if they are all categorical, you will have

at least $2^9=512$ combinations. You will have many combinations that are empty for one or both categories of the outcome and the regression will be unstable. The solution for this is to use fewer variables on the right hand side of the equation.

Subject: Re: Variance Inflation Factor
Posted by [nwegbus](#) on Fri, 03 Feb 2017 08:24:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for your reply.

According to the UCLA Stata webpage here:

<http://www.ats.ucla.edu/stat/stata/faq/svycollin.htm>

the VIF and tolerance values are the way to check for multicollinearity (i.e., redundancy) among predictors for survey data.

You did point out that I had too many predictors in my model. I agree. I'm trying to use the VIF and tolerance values to see which ones are best to drop from my model.

The problem I'm having is that, even though I followed the directions on the UCLA webpage step by step, I end up with an output of:

Tolerance=. and VIF=.

Could this have anything to do with the way I weighted the data? i.e.,
generate wgt = d005/1000000
svyset [pweight = wgt],psu(v021) strata(v022) singleunit(centered)

Thank you once again for your help and patience.
Som

File Attachments

1) [Stata output for VIF command.docx](#), downloaded 545 times

Subject: Re: Variance Inflation Factor
Posted by [Bridgette-DHS](#) on Fri, 03 Feb 2017 14:27:04 GMT
[View Forum Message](#) <> [Reply to Message](#)

Another response from Tom Pullum:

This question or problem goes beyond what we would normally give advice on. If I were in your

situation, I would try to do the backward selection with successive reductions of the svy adjustments, until (hopefully!) it worked. That is, first remove the stratum adjustment, and try it. If it still doesn't work, remove the cluster adjustment, and try it. If it still doesn't work, remove the weight adjustment, and try it. I HOPE that at one stage of this process, at least the final one (with no svy adjustments at all) the procedure will work.

When you include all of the svy adjustments, the models are not fitted with maximum-likelihood methods. The measures of fit and the optimization of fit are not as solid as with ML (at least that's my understanding). However, I think that model selection procedures are fairly robust with respect to inclusion/omission of the svy adjustments. Plus, my general strategy when running into a complex problem is to simplify, simplify, simplify, until I get a strategy or a solution! That's all I can suggest.

Subject: Re: Variance Inflation Factor
Posted by [nwegbus](#) on Fri, 03 Feb 2017 15:23:56 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you Tom. I will try this and see where it gets me.

Som

Subject: Re: Variance Inflation Factor
Posted by [bakerchowdhury](#) on Thu, 13 Apr 2017 19:25:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,
I see VIF for SVY command gives us an overall tolerance for the model. However, I am wondering if there a way of calculating VIF for each predictor variable in SVY command (i.e. more like the VIF command under normal regression)?
Thank you
Baker

Subject: Re: Variance Inflation Factor
Posted by [Bridgette-DHS](#) on Fri, 14 Apr 2017 11:46:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

Response from Senior DHS Stata Specialist, Tom Pullum:

This is not a question that DHS staff can answer. Perhaps other users can help you.

Subject: Re: Variance Inflation Factor

Posted by [AnvitaDixit](#) on Wed, 28 Jun 2017 18:03:42 GMT

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

Hi Baker, did you find a solution? Im having the same problem and would really appreciate if you could share how you got to individual VIFs for your independent variable in the model. Thanks!
