How can I calculate the Variance Inflation Factor (VIF) using SAS software?

Authored by

stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the Variance Inflation Factor (VIF) using SAS software?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=148334

The Variance Inflation Factor (VIF) is a measure used to assess the severity of multicollinearity in a regression model. It is calculated by dividing the variance of a regression coefficient by the variance of the same coefficient in a model without that variable. To calculate the VIF using SAS software, one can use the PROC REG procedure and specify the VIF option. This will provide a table with the VIF values for each variable in the model. The VIF values can then be interpreted to identify any variables with high levels of multicollinearity, which can affect the accuracy and reliability of the regression results. By using SAS software, users can easily and efficiently assess and address multicollinearity in their regression models.

Calculate Variance Inflation Factor (VIF) in SAS

In regression analysis, occurs when two or more predictor variables are highly correlated with each other, such that they do not provide unique or independent information in the regression model.

If the degree of correlation is high enough between variables, it can cause problems when fitting and interpreting the regression model.

One way to detect multicollinearity is by using a metric known as the variance inflation factor (VIF), which measures the correlation and strength of correlation between the explanatory variables in a .

This tutorial explains how to calculate VIF in SAS.

Example: Calculating VIF in SAS

For this example we'll create a dataset that describes the attributes of 10 basketball players:

```
/*create dataset*/
data my_data;
input rating points assists rebounds;
datalines;
90 25 5 11
85 20 7 8
82 14 7 10
88 16 8 6
94 27 5 6
90 20 7 9
76 12 6 6
75 15 9 10
87 14 9 10
86 19 5 7
;
run;

/*view dataset*/
proc printdata=my_data;
```

Obs	rating	points	assists	rebounds
1	90	25	5	11
2	85	20	7	8
3	82	14	7	10
4	88	16	8	6
5	94	27	5	6
6	90	20	7	9
7	76	12	6	6
8	75	15	9	10
9	87	14	9	10
10	86	19	5	7

Suppose we would like to fit a multiple linear regression model using rating as the response variable and points, assists, and rebounds as the predictor variables.

We can use to fit this regression model along with the VIF option to calculate the VIF values for each predictor variable in the model:

```
/*fit regression model and calculate VIF values*/
proc regdata=my_data;
model rating = points assists rebounds / vif;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: rating

Number of Observations Read	10
Number of Observations Used	10

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	207.99697	69.33232	3.30	0.0995
Error	6	126.10303	21.01717		
Corrected Total	9	334.10000			

Root MSE	4.58445	R-Square	0.6226
Dependent Mean	85.30000	Adj R-Sq	0.4338
Coeff Var	5.37450		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	62.47163	14.58822	4.28	0.0052	0
points	1	1.11933	0.41088	2.72	0.0345	1.76398
assists	1	0.88340	1.38067	0.64	0.5459	1.95910
rebounds	1	-0.42777	0.85101	-0.50	0.6331	1.17503

From the Parameter Estimates table we can see the VIF values for each of the predictor variables:

points: 1.76398assists: 1.96591rebounds: 1.17503

Note: Ignore the VIF for the "Intercept" in the model since this value is irrelevant.

The value for VIF starts at 1 and has no upper limit. A

rule of thumb for interpreting VIFs is as follows:

A value of 1 indicates there is no correlation between a given predictor variable and any other predictor variables in the model.A value between 1 and 5 indicates moderate correlation between a given predictor variable and other predictor variables in the model, but this is often not severe enough to require attention.A value greater than 5 indicates potentially severe correlation between a given predictor variable and other predictor variables in the model. In this case, the coefficient estimates and p-values in the regression output are likely unreliable.

How to Deal with Multicollinearity

If you determine that multicollinearity is a problem in your regression model, there are a few common ways to deal with it:

1. Remove one or more of the highly correlated variables.

This is the quickest fix in most cases and is often an acceptable solution because the variables you're removing are redundant anyway and add little unique or

independent information the model.

2. Linearly combine the predictor variables in some way, such as adding or subtracting them from one way.

By doing so, you can create one new variables that encompasses the information from both variables and you no longer have an issue of multicollinearity.

3. Perform an analysis that is designed to account for highly correlated variables such as principal component analysis or partial least squares (PLS) regression.

These techniques are specifically designed to handle highly correlated predictor variables.

The following tutorials explain how to perform other common tasks in SAS: