

Test for Multicollinearity in SPSS, how?

Authored by
stats writer

December 26, 2025

RECOMMENDED CITATION

stats writer (2025). *Test for Multicollinearity in SPSS, how?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=108870>

The reliability of statistical models, particularly those derived from regression analysis, hinges on the independence of their components. When examining relationships between variables, a critical assumption often violated is the absence of multicollinearity. This pervasive statistical issue arises when two or more independent, or predictor variables, exhibit a high degree of linear correlation among themselves. If present and ignored, high correlation can severely compromise the accuracy, stability, and interpretability of the model's estimated coefficients, making it difficult to discern the unique contribution of each predictor.

To effectively diagnose this problem within statistical software like SPSS (Statistical Package for the Social Sciences), researchers rely on quantitative metrics. The most widely accepted and robust diagnostic tool for quantifying the severity of collinearity is the Variance Inflation Factor (VIF). By calculating the VIF for each predictor in the model, we can systematically assess how much the variance of an estimated regression coefficient is increased due to collinearity with the other predictors.

This comprehensive guide is designed for expert users and aspiring analysts alike, detailing the precise methodology for calculating and interpreting the VIF within SPSS. We will walk through the conceptual underpinnings of multicollinearity, establish clear interpretation criteria, and provide a step-by-step tutorial demonstrating how to execute the necessary diagnostic tests to ensure the robustness of your regression analysis.

Understanding the Challenge of Multicollinearity

In the context of regression analysis, the fundamental goal is to isolate the unique influence of each predictor variable on the dependent variable. Multicollinearity fundamentally undermines this goal. It occurs when two or more predictor variables are so closely related--highly correlated--that they essentially convey redundant information to the statistical model. While some degree of correlation is expected in real-world data, excessively high correlation means the predictors are not providing unique or independent insights into the response variable.

The presence of severe collinearity leads to several problematic outcomes. Firstly, it inflates the standard errors of the regression coefficients, which in turn reduces the statistical power of the model. This inflation makes it difficult to declare certain predictors statistically significant, even if they truly influence the outcome. Secondly, the coefficient estimates themselves become highly sensitive to minor changes in the dataset, leading to unstable and unreliable model results. If you were to collect a slightly different sample, the signs and magnitudes of your regression coefficients might dramatically shift, hindering practical interpretation and theoretical advancement.

For instance, if you are modeling academic success using both "hours spent studying" and "number of practice tests taken," and students who study more always take more practice tests,

these two predictor variables are highly collinear. The model struggles to determine whether the increase in score is attributable to studying duration or practice tests, because they move in near-perfect lockstep. Diagnosing this interdependence is the crucial first step toward building a reliable predictive model, and this diagnosis is precisely where the Variance Inflation Factor (VIF) becomes indispensable.

The Critical Role of the Variance Inflation Factor (VIF)

The Variance Inflation Factor (VIF) is the primary metric used to quantify the extent of multicollinearity in a regression analysis. Conceptually, the VIF measures how much the variance of an estimated regression coefficient is inflated compared to what it would be if that specific predictor variable were uncorrelated with the other predictor variables in the model. Technically, the VIF for a predictor variable (X_i) is calculated using the coefficient of determination (R^2) obtained from regressing X_i against all other predictor variables in the model.

The formula for the VIF is defined as: $VIF_i = 1 / (1 - R^2_i)$. Here, R^2_i represents the R-squared value achieved when the i -th predictor variable is designated as the dependent variable and is regressed on all the remaining predictor variables. If a predictor variable has no linear relationship with any of the others, R^2_i would be zero, resulting in a VIF of exactly 1. As the correlation between the i -th predictor and the others increases, R^2_i approaches 1, causing the denominator ($1 - R^2_i$) to approach zero, and thus the VIF value spirals upward, signifying extreme variance inflation. The VIF measures the correlation and strength of correlation between the predictor variables in a regression model.

Therefore, the higher the VIF value, the greater the degree of correlation between that specific predictor variables and the other predictors, and the more unreliable its corresponding regression coefficient estimate becomes. Understanding this mathematical relationship is key to appreciating why VIF is such a robust and necessary diagnostic tool. It provides a numerical measure that directly translates the conceptual problem of shared variance into a quantifiable measure of estimation instability.

Interpreting VIF Thresholds: Guidelines for Detection

The VIF is bounded by 1 on the lower end, indicating zero collinearity, and has no theoretical upper limit. Because there is no single universally accepted threshold for defining problematic multicollinearity, researchers rely on established rules of thumb developed through extensive statistical practice. These guidelines help determine when the inflation of variance is severe enough to warrant corrective action, such as removing a variable, combining variables, or employing advanced techniques like principal component regression.

The general interpretation framework for VIF values is categorized 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.

A value greater than 10 is commonly cited as the critical threshold, signifying **definite high multicollinearity**. This indicates a high degree of multicollinearity where results should be treated with extreme caution.

While the $VIF > 10$ rule is a conservative standard, some strict statistical fields may prefer the $VIF > 5$ threshold for more rigorous models. The choice often depends on the research context and the intended application of the model. However, regardless of the specific threshold chosen, the VIF provides the objective data necessary for making an informed decision about model diagnostics.

Setting Up the Multicollinearity Test in SPSS

To demonstrate the practical application of VIF diagnostics, we will utilize a typical educational dataset within SPSS. Our objective is to perform a multiple linear regression aimed at predicting a student's final exam score based on several potential determinants. Specifically, we are interested in whether the independent variables are sufficiently independent of one another to provide reliable coefficient estimates.

Suppose we have a dataset showing the exam score of 10 students along with the number of hours they spent studying, the number of prep exams they took, and their current grade in the course. This dataset structure is commonly used to explore the relationship between study inputs and academic outputs:

Exam **Score** (Dependent variable).

Number of study **Hours** (Predictor variable 1).

Number of **Prep Exams** taken (Predictor variable 2).

Current Grade in the course (Predictor variable 3).

The initial requirement, before interpreting the predictive power of these variables, is to ensure that **hours**, **prep_exams**, and **current_grade** are not severely correlated. The visual representation of this dataset structure in the SPSS Data View is presented below:



27 :

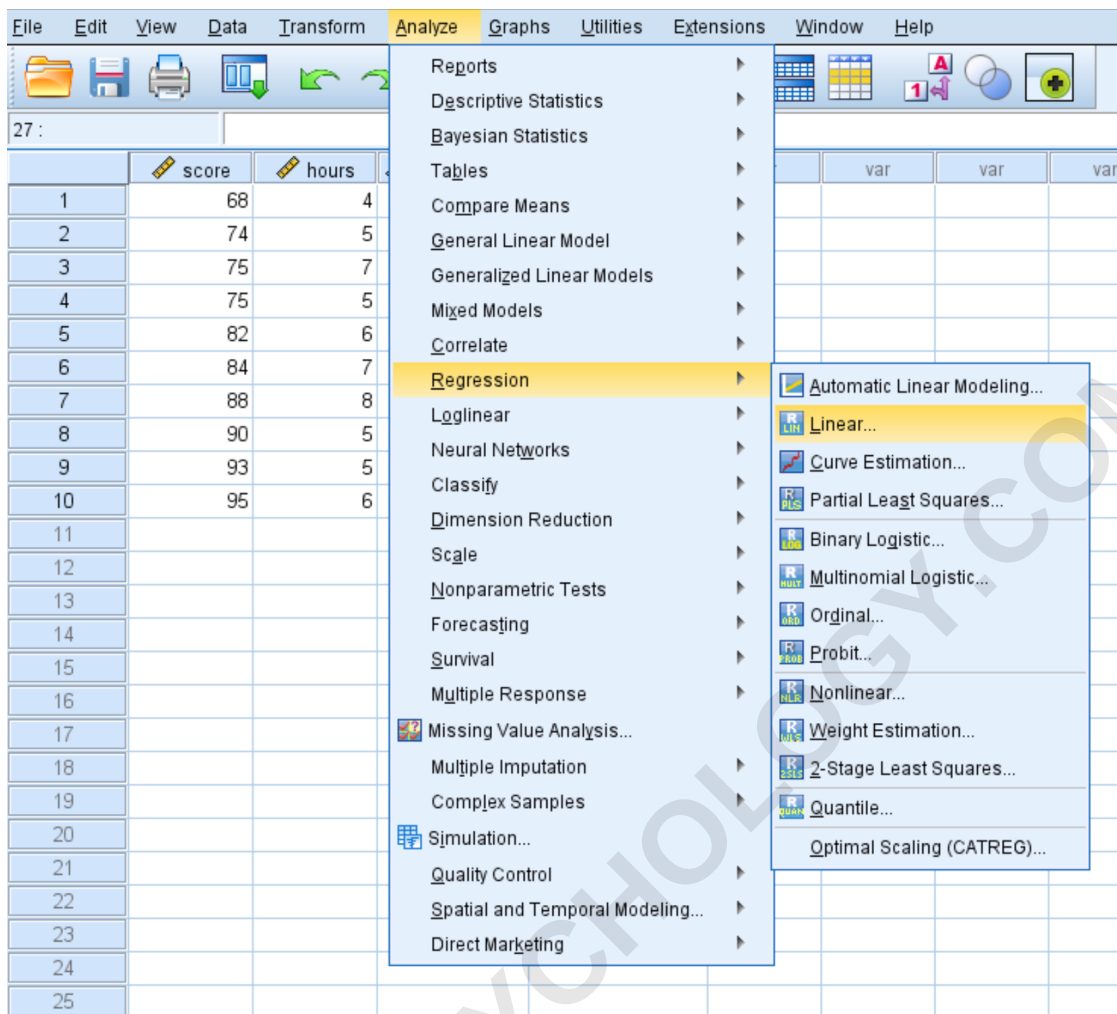
	score	hours	prep_exams	current_grade
1	68	4	2	75
2	74	5	2	77
3	75	7	3	78
4	75	5	3	80
5	82	6	1	75
6	84	7	2	90
7	88	8	3	88
8	90	5	4	84
9	93	5	4	96
10	95	6	2	90
11				
12				
13				

We intend to execute a linear regression using **score** as the response variable and the three established variables--**hours**, **prep_exams**, and **current_grade**--as the predictor variables. The subsequent steps will detail the exact process within SPSS required to generate the necessary VIF metrics.

Step-by-Step Procedure: Running Linear Regression in SPSS

The calculation of the VIF is automatically integrated into the standard regression procedures within SPSS, specifically under the Collinearity Diagnostics options. To begin the analysis, navigate through the primary menus using the following sequence:

1. Click on the **Analyze** tab located in the main menu bar.
2. Hover over **Regression** in the dropdown menu.
3. Select **Linear** from the subsequent options, which initiates the dialogue box necessary for setting up the model. This is the crucial starting point for any standard linear regression or collinearity check.



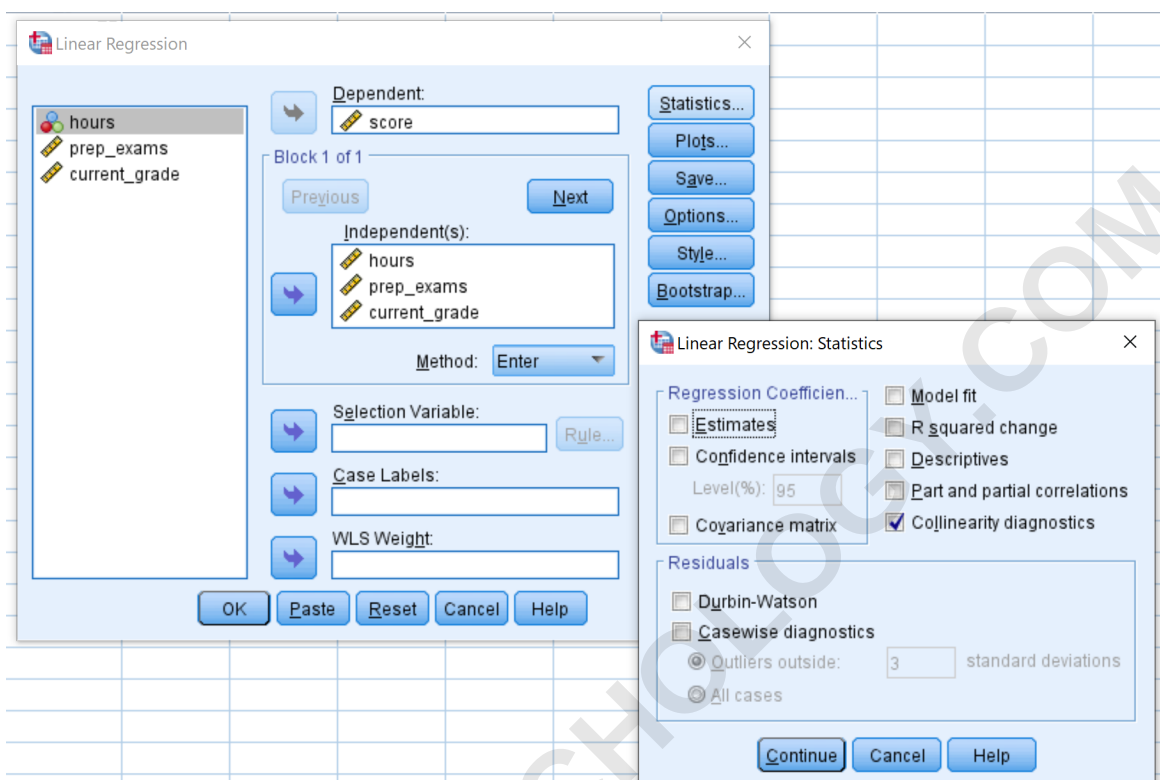
Once the Linear Regression dialog box appears, the variables must be assigned their appropriate roles. Drag the dependent variable, **score**, into the box labelled "Dependent." Next, select the three predictor variables--**hours**, **prep_exams**, and **current_grade**--and move them into the box labelled "Independent(s)." Note that the order of the independent variables does not affect the calculation of VIF, as this metric is derived from their collective intercorrelation structure.

After correctly placing the variables, the essential step is to activate the diagnostic statistics. Click the **Statistics** button within the Linear Regression dialog box. This opens a new window where specialized metrics can be requested. Ensure that the checkbox next to **Collinearity diagnostics** is selected. Then click **Continue** to return to the main regression dialogue. Then click **OK** to execute the analysis.

Accessing and Utilizing Collinearity Diagnostics

The request for **Collinearity diagnostics** ensures that the VIF values are generated in the output viewer. This step is critical because standard regression output only provides coefficients and R-

squared information, not the intercorrelation diagnostics needed to assess model stability. By activating this option, we compel SPSS to perform the internal regressions required to derive the VIF for each predictor variable.



Upon execution, SPSS generates several tables. The crucial table for our current analysis is the one labeled **Coefficients**, which typically houses not only the unstandardized and standardized coefficients but also the two key collinearity diagnostics: Tolerance and VIF. The resulting table, detailing the collinearity statistics for our sample dataset, will appear in the output window as shown:

Coefficients^a

		Collinearity Statistics	
Model		Tolerance	VIF
1	hours	.856	1.169
	prep_exams	.713	1.403
	current_grade	.657	1.522

a. Dependent Variable: score

Analyzing the SPSS Output and Drawing Conclusions

The final step in the diagnostic process involves meticulously examining the generated VIF values against the established interpretation thresholds. Reviewing the 'Coefficients' table excerpt provided by SPSS, we extract the VIF values calculated for each of the three predictor variables:

hours: **1.169**

prep_exams: **1.403**

current_grade: **1.522**

Each of these values must be critically assessed to determine if they indicate a concerning level of multicollinearity that could compromise the final regression analysis model. We apply the standard interpretive rule that a VIF value significantly greater than 5, or more conservatively greater than 10, signals problematic collinearity requiring intervention.

In this particular example, all calculated VIF values are extremely low, falling well within the acceptable range of 1 to 5. The highest VIF, corresponding to **current_grade**, is only 1.522. This finding strongly suggests that the three predictor variables--hours studied, preparation exams taken, and current course grade--do not exhibit an unacceptably high degree of linear intercorrelation. Consequently, the variance inflation of the estimated regression coefficients is minimal.

The conclusion drawn is that multicollinearity is not a significant problem in this specific dataset and model structure. We can proceed with confidence to interpret the coefficient estimates and p-values generated by the linear regression model, knowing that these results are stable and reliable, reflecting the unique contribution of each predictor toward explaining the variance in the exam score.