

How to perform White's Test in SAS?

Authored by
stats writer

November 19, 2025

RECOMMENDED CITATION

stats writer (2025). *How to perform White's Test in SAS?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=97016>

White's Test is a fundamental diagnostic procedure in statistics, particularly within the SAS environment, utilized to detect the presence of heteroskedasticity in a linear regression model. This critical test ensures that the necessary assumptions underlying ordinary least squares (OLS) regression are not violated, which is essential for deriving reliable statistical inferences and accurate hypothesis testing.

Conceptually, the procedure involves running an auxiliary regression where the squared residuals from the primary model are regressed against the independent variables, their squared terms, and their cross-products. By examining the overall fit or specific coefficients in this auxiliary regression--often summarized by a Chi-Square statistic--we determine if the variance of the residuals is non-constant. If the resulting test statistic is statistically significant, it signals that heteroskedasticity is present, necessitating immediate corrective measures to validate the model's standard errors.

In SAS, the implementation of this diagnostic is streamlined. While classical methods might involve manual calculation of residuals and auxiliary regressions, modern SAS procedures like

PROC REG

allow the test to be performed automatically using specialized options such as **SPEC**. This tutorial provides a comprehensive guide on executing and interpreting White's Test to guarantee the robustness of your regression analysis.

Introduction to White's Test and Heteroskedasticity

The main purpose of the White's Test, developed by Halbert White, is to assess the validity of the homoskedasticity assumption. If the variance of the error terms is not constant--a condition known as heteroskedasticity--the ordinary least squares estimator remains unbiased and consistent, but it loses its efficiency. More importantly, the calculated standard errors become biased, leading to unreliable t-statistics and potentially misleading conclusions about the significance of the predictor variables.

Understanding Heteroskedasticity in Regression

Heteroskedasticity refers to the unequal scatter of the error variance at different levels of the independent variables within a regression model. This situation violates a foundational assumption of the Classical Linear Model (CLM) that the residuals are equally scattered (homoskedastic) across the entire range of predicted values. Recognizing and addressing this violation is critical for producing scientifically sound research and accurate predictive models.

When homoskedasticity is violated, the reliability of hypothesis tests, such as those determining if a coefficient is significantly different from zero, is compromised because the reported standard errors

are incorrect. The White's Test serves as a powerful diagnostic tool, helping the analyst rigorously confirm whether this assumption holds true for their specific dataset before proceeding to interpret the final parameter estimates.

Step-by-Step Example Setup in SAS

To demonstrate the application of White's Test, we will establish a multiple linear regression model designed to predict student performance. Our response variable is the final exam score, predicted by two explanatory variables: the number of hours spent studying and the number of preparatory exams taken. The formalized model used is:

$$\text{Exam Score} = \beta_0 + \beta_1(\text{hours}) + \beta_2(\text{prep exams}) + \text{Residual Error}$$

The steps below detail how to structure the data and execute the necessary SAS commands to fit this model and simultaneously perform the required diagnostic test for unequal variance.

Generating the Sample Data in SAS

Initially, we must create a dataset in SAS that contains the observations for 20 students, detailing their study time, prep exams taken, and final scores. The following code uses the **DATA** and **DATALINES** statements to input this information, followed by **PROC PRINT** to display the resulting structure.

```
/*create dataset*/
data exam_data;
input hours prep_exams score;
datalines;
1 1 76
2 3 78
2 3 85
4 5 88
2 2 72
1 2 69
5 1 94
4 1 94
2 0 88
4 3 92
4 4 90
3 3 75
6 2 90
5 4 90
```

```
3 4 82
4 4 85
6 5 90
2 1 83
1 0 62
2 1 76
;
```

```
run;

/*view dataset*/
proc print data=exam_data;
```

The structured data is confirmed by the output, which provides a clean visualization of the variables used in the regression analysis.

Obs	hours	prep_exams	score
1	1	1	76
2	2	3	78
3	2	3	85
4	4	5	88
5	2	2	72
6	1	2	69
7	5	1	94
8	4	1	94
9	2	0	88
10	4	3	92
11	4	4	90
12	3	3	75
13	6	2	90
14	5	4	90
15	3	4	82
16	4	4	85
17	6	5	90
18	2	1	83
19	1	0	62
20	2	1	76

Executing White's Test using PROC REG

The multiple linear regression model is fitted using the **PROC REG** procedure. To automatically execute White's Test, we append the **SPEC** option to the **MODEL** statement. The **SPEC** option in **PROC REG** produces several specification tests, including a general test for heteroskedasticity based on the White methodology.

The following code instructs SAS to model the **score** using **hours** and **prep_exams** as predictors, and then run the required specification checks:

```
/*fit regression model and perform White's test*/  
proc reg data=exam_data;  
model score = hours prep_exams / spec;  
run;  
quit;
```

Upon execution, the output will include all standard regression metrics as well as the specialized table detailing the results of the specification tests performed due to the **SPEC** option.

Interpreting the White's Test Output

The key to interpreting White's Test lies in the specific output table containing the Chi-Square test statistic and the associated p-value. This table summarizes the auxiliary regression used to detect variance instability.

From the visual output provided below, we can isolate the necessary values. The calculated **Chi-Square test statistic** is found to be **3.54**, and the corresponding p-value is **0.6175**. These figures are crucial for performing the final hypothesis test regarding the presence of heteroskedasticity.

The REG Procedure
Model: MODEL1
Dependent Variable: score

Number of Observations Read	20
Number of Observations Used	20

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	937.81311	468.90656	13.62	0.0003
Error	17	585.13689	34.41982		
Corrected Total	19	1522.95000			

Root MSE	5.86684	R-Square	0.6158
Dependent Mean	82.95000	Adj R-Sq	0.5706
Coeff Var	7.07274		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	69.66126	3.07879	22.63	<.0001
hours	1	4.65376	0.98317	4.73	0.0002
prep_exams	1	-0.55943	0.99979	-0.56	0.5831

The REG Procedure
Model: MODEL1
Dependent Variable: score

Test of First and Second Moment Specification		
DF	Chi-Square	Pr > ChiSq
5	3.54	0.6175

The Null and Alternative Hypotheses

To formally conclude the diagnostic test, we must evaluate the results against the predefined null and alternative hypotheses:

Null Hypothesis (H₀): The variance of the errors is constant across all observations

(Homoskedasticity is present).

Alternative Hypothesis (HA): The variance of the errors is not constant (Heteroskedasticity is present).

Using the standard significance level of $\alpha = 0.05$, we compare our observed p-value (0.6175) to α . Since 0.6175 is substantially greater than 0.05 , we fail to reject the null hypothesis.

The conclusion is clear: we do not have sufficient evidence to suggest that heteroskedasticity is an issue in this specific regression model. Therefore, it is appropriate and safe to interpret the standard errors and subsequent t-tests derived from the original regression summary table.

Addressing Heteroskedasticity: Next Steps

Had we rejected the null hypothesis, meaning the p-value was less than 0.05 , this would confirm the presence of unequal variance. In such a scenario, the output from the original OLS regression is compromised, specifically regarding the calculation of the standard errors, making the inference invalid. When this occurs, analysts must implement corrective strategies to ensure statistically robust results.

The two most common approaches to mitigate the effects of heteroskedasticity involve either modifying the data through transformation or adjusting the estimation technique itself, such as using robust variance estimators. The choice depends on the severity of the violation and the analyst's goals.

Methods for Correcting Heteroskedasticity

If heteroskedasticity is confirmed by White's Test, two highly effective corrective methods are recommended:

Transform the Response Variable. This involves applying a mathematical function, such as the natural logarithm (log) or the square root, to the dependent variable. Log transformation is particularly effective because it often compresses the scale of the response, which tends to stabilize the residual variance and restore homoskedasticity.

Use Heteroskedasticity-Consistent (HC) Standard Errors. This method, often called "robust standard errors," recalculates the standard errors to be valid even in the presence of unequal variance. In SAS, this is achieved by specifying the **VCE(ROBUST)** option in the **PROC REG** statement. This approach is frequently preferred as it modifies the standard errors without changing the estimated regression coefficients.

Alternatively, analysts can employ **Weighted Least Squares (WLS)**. This technique explicitly

assigns smaller weights to observations associated with larger error variances and greater weights to observations with smaller variances. By assigning appropriate weights, WLS can effectively eliminate the problem of heteroskedasticity and produce efficient, unbiased estimators.

ARABPSYCHOLOGY.COM