How can White’s Test be performed in SAS?

Authored by

stats writer

June 23, 2024

White's Test is a statistical method used to assess the presence of heteroscedasticity, or unequal variance, in a regression model. This test can be performed in SAS (Statistical Analysis System) by using the built-in "proc model" procedure. First, the data must be loaded into SAS and the regression model must be specified. Then, the "proc model" procedure can be used to run White's Test by including the "white" option in the model statement. This will generate a test statistic and p-value, which can be used to determine the presence of heteroscedasticity. Additionally, SAS provides options to adjust for potential autocorrelation and to perform post-estimation diagnostics. The results of White's Test in SAS can be used to evaluate the assumptions of the regression model and make necessary adjustments to improve the accuracy of the analysis.

Perform White's Test in SAS

White's test is used to determine if is present in a regression model.

Heteroscedasticity refers to the unequal scatter of at different levels of a in a regression model, which violates one of the key that the residuals are equally scattered at each level of the response variable.

This tutorial explains how to perform White's test in SAS to determine whether or not heteroscedasticity is a problem in a given regression model.

Example: White's Test in SAS

Suppose we want to fit a multiple linear regression model that uses number of hours spent studying and number of prep exams taken to predict the final exam

score of students:

Exam Score = β0 + β1(hours) +β2(prep exams)

First, we'll use the following code to create a dataset that contains this information for 20 students:

```
/*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 printdata=exam_data;
```

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

Next, we'll use proc reg to fit this multiple linear regression model along with the spec option to perform White's test test for heteroscedasticity:

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

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 last table in the output shows the results of White's test.

From this table we can see that the Chi-Square test statistic is 3.54 and the corresponding p-value is 0.6175.

White's test uses the following null and alternative hypotheses:

Null (H0): Heteroscedasticity is not present.Alternative (HA): Heteroscedasticity is present.

Since the p-value is not less than 0.05, we fail to reject the null hypothesis.

Thus, it's safe to interpret the standard errors of the coefficient estimates in the regression summary table.

What To Do Next

If you fail to reject the null hypothesis of White's test, then heteroscedasticity is not present and you can proceed to interpret the output of the original regression.

However, if you reject the null hypothesis, this means heteroscedasticity is present in the data. In this case, the standard errors that are shown in the output table of the regression may be unreliable.

There are a couple common ways that you can fix this issue, including:

1. Transform the response variable. You can try performing a transformation on the response variable.

For example, you could use the log of the response variable instead of the original response variable.

Typically taking the log of the response variable is an effective way of making heteroscedasticity go away.

Another common transformation is to use the square root of the response variable.

2. Use weighted regression. This type of regression assigns a weight to each data point based on the variance of its fitted value.

This gives small weights to data points that have higher variances, which shrinks their squared residuals.

When the proper weights are used, this can eliminate the problem of heteroscedasticity.