CHECKING HOMOSCEDASTICITY WITH SAS

In a linear regression model, there should be **homogeneity of variance** of the residuals. In other words, the variance of residuals are approximately equal for all predicted dependent variable values.

**Example**

The Variation in income increases with years of work experience.

Income with work experience 4 years: 30,40,60 with absolute difference 10, 30 and relative difference 33%,100% and log difference 0.29, 0.69.

Income at work experience 8 years: 90,120, 180 with absolute difference 30, 90 and relative difference 33%, 100% and log difference 0.29, 0.69

**Note :**Often after log transformation of dependent variable makes variance constant.

**Consequences of Heteroscedasticity**

*The regression prediction remains unbiased and consistent but inefficient. It is inefficient because the estimators are no longer the Best Linear Unbiased Estimators (BLUE). The hypothesis tests (t-test and F-test) are no longer valid.*

**How to check Homoscedasticity**

1. **White Test -**This statistic is asymptotically distributed as chi-square with k-1 degrees of freedom, where k is the number of regressors, excluding the constant term.

2.**Breusch-Pagan test**  
3. **Lagrange multiplier (LM) test**  
 **With PROC AUTOREG (LM Test and Supports CLASS Statement)**

*proc autoreg data= bhalla.GLMSELECT;  
model crime = yr\_rnd mealcat some\_col /****archtest****;  
output out=r r=yresid;  
run;*

**Note : Check P-value of Q statistics and LM tests. P-value greater than .05 indicates homoscedasticity.**  
 **With PROC MODEL (White and PAGAN Test)**

*proc model data= bhalla.GLMSELECT;  
parms a1 b1 b2 b3;  
api00 = a1 + b1\*yr\_rnd + b2\*mealcat + b3\*some\_col;  
fit api00 / white pagan=(1 yr\_rnd mealcat some\_col)  
out=resid1 outresid;  
run;  
quit;*

**If the p-value of white test and Breusch-Pagan test is greater than .05, the homogenity of variance of residual has been met (Homoscedasticity).**  
 **Note : PROC AUTOREG supports CLASS statement.**  
 **Remedy :**  
 **1. Box-Cox transformations of the dependent variable**

Box-Cox transformations are used to find potentially nonlinear transformations of a dependent variable.

***PROC TRANSREG****DATA = bhalla.GLMSELECT  TEST;  
MODEL BOXCOX(api00) = IDENTITY(yr\_rnd mealcat some\_col);  
RUN;*

**Note :**Categorical variables can be used with **CLASS**statement instead of **IDENTITY**.

*Check****Lambda score****generated from****PROC TRANSREG***

|  |  |
| --- | --- |
| **Transformation** | **Best Lambda** |
| Square | 1.5 to 2.5 |
| None | 0.75 to 1.5 |
| Square-root | 0.25 to 0.75 |
| Natural log | -0.25 to 0.25 |
| Inverse square-root | -0.75 to -0.25 |
| Reciprocal | -1.5 to -0.75 |
| Inverse square | -2.5 to -1.5 |

**2. Weighted Least Squares**

*If variable transformation does not solve the problem, we can use****weighted least squares****.*

**How to construct weights :**

1. Compute the absolute and squared residuals
2. Find the absolute and squared residuals vs. independent variables to get the estimated standard deviation and variance
3. Compute the weights using the estimated standard deviations and variance.