19 January 2018
18:48

TYPES OF RESIDUALS

There are 4 types of Residuals.

1. Standardized Residual
2. Studentized Residual
3. Deleted Residual
4. R-Student Residual

<u>Standardized Residuals:</u> It stands for di $= \frac{e_i}{\sqrt{Ms\,Re\,s}}$

<u>Studentized Residuals:</u> It stands for $r_i = \frac{e_i}{\sqrt{v(e_i)}}$

Note:

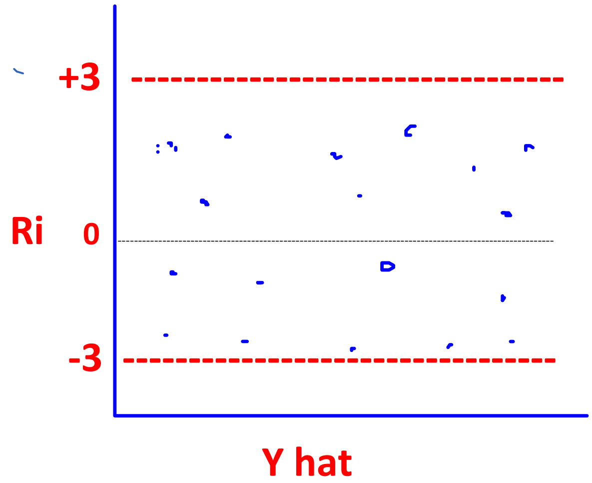
Comparing in both the residuals, studentized is more accurate.

These residuals helps us to find out the weather our data is following model assumptions are not

Let us work on the way to evaluate the model assumptions using residuals.

Residual Plots :

Ri vs Y hat



- If all the data points falling under + or - 3 which is called as band and we can assume that data is following homoscedasticity i.e $v(\epsilon) = \sigma^2$. Since it is having constant variance between point to point with in the range.

If my model is good fit model, then it could be like this
Ex: Y = B0 + B1X1 + B2X2 + ∈

If I my model is reduced to with only X1 then the model will be like this
Ex: Y = B0 + B1X1 + e

If I compare between e and ∈$\longrightarrow$ $e > \in$ that means, $e = \in + e0$, So if I fit the X2 in model e0 will go away. Right? . That means e and are ∈ going to be equal.

In the above, ∈ which is not reducible and e is reducible thing.

Case 1
Now, If I plot the graph between reduced model Y and missing variable X2 then the their should me some relation ship exists.

Case 2
Now, If I plot the graph between good fit model Y and variable X2 then the their should not be any relation ship will be seen.

So, whenever if you want to check a new variable to add or not to the existing model,
You need to plot the graph between existing model Ri vs X* (New variable).
If you find any relation ship exists between those then you need to add those variable to the model or else it is not required.

Note:

- Some times, we fitted a model using x1 in the model and we don't know what else to be done to that model, so I plotted the graph Ri vs Y hat and seen some curvature kind of graph is missing then we need to update the model as Y = Bo + B1x1 + B2 x_1^2
This kind of regression is called Polynomial regression.

So, the question is while using polynomial regression, is that multicollinearity going to exist or not?

Yes, to manage this issue we modify the fitted model in to the below method.

Y = Bo + B1(x1-x1 bar) + B2 (x1-x1 bar)^2

- Sometimes, we fit a model with proper variables and still we found some homoscedasticity issues and we cannot under stand how to handle the situation.

Ex: Actual model is like this Y = Bo + B1x1 + B2 x_2^2
We fitted the model like Y = Bo + B1x1 + B2x2

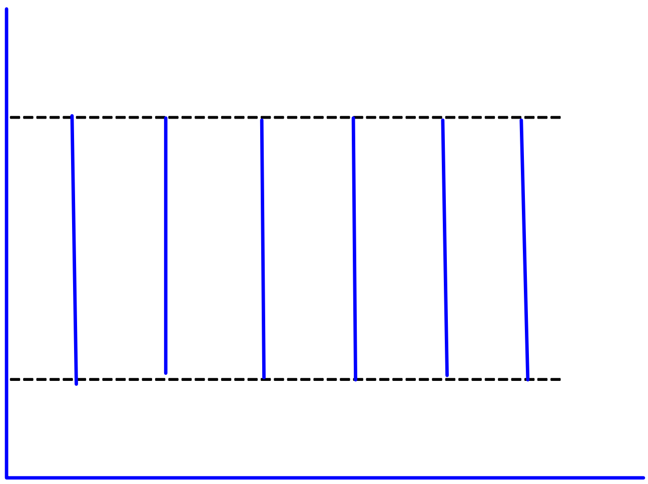
That means we came to know that X2 to be used in our model fitting but it was not used in the proper order. Because we still see homoscedasticity problem existing.

So if the model assumptions are violating i.e (Ri vs Y hat) is violating in this case we need to use either transformations or add new variables

<u>Possible Residual Graphs:</u>

a.

Ri



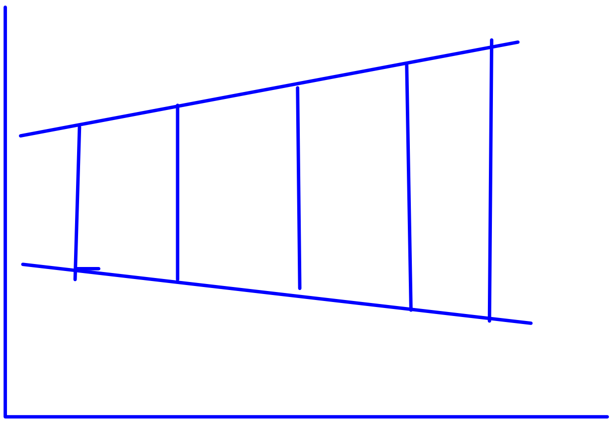
-------> No Problem

Y hat, xi

Below are some Homoscedastic problem graphs

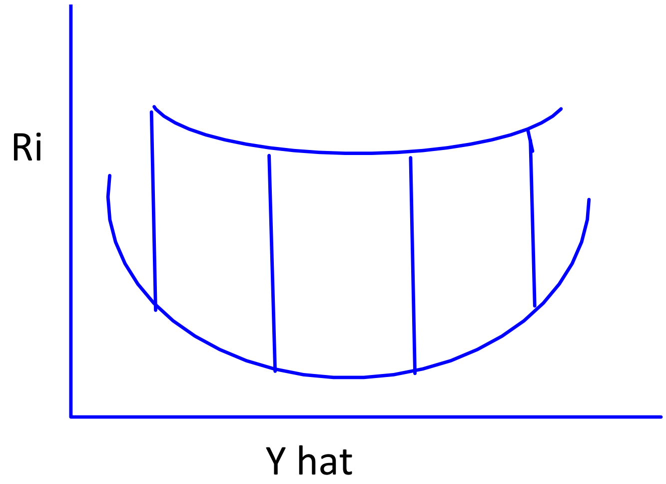
b.

Ri

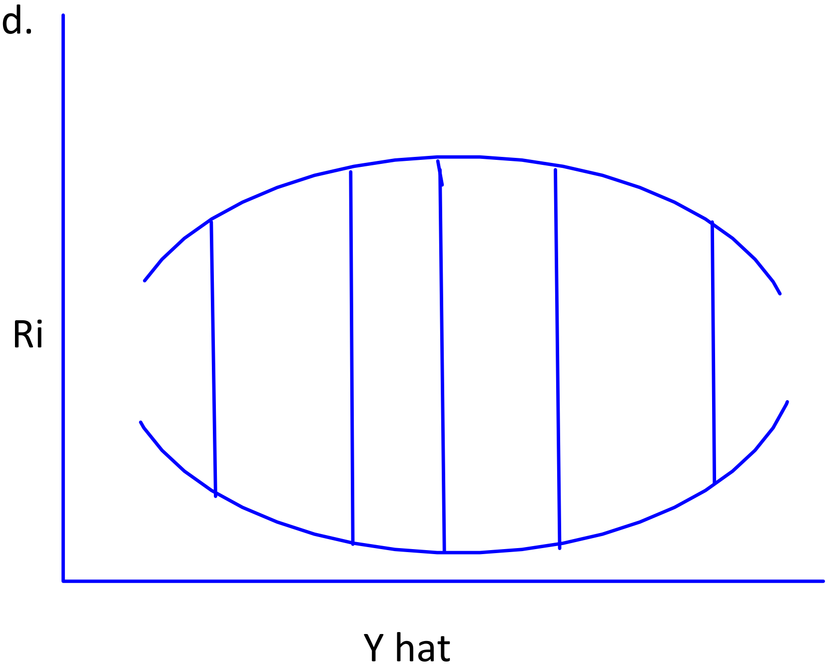


Y hat

c.



d.



----> Double Bow Problem