Assumptions in Linear regression:

1. Linearity between Independent and dependent variables

It is very important to check the linear condition before applying linear regression

Scatter plots will give that linear observations

1. There is no multicollinearity between the independent variables

* There should be no relation between the input variables
* Multicollinearity is the term related to co-relation

Tv np sales

100 200 2000

200 400 3000

300 600 6000

NP=2\*TV

Y= bo+b1\*TV+b2\*NP

Y= bo+b1\*TV+b2(2TV)

Y= bo+TV(b1+2b2)

B1+2b2=0

One equation we have but two unknown parameters are there, the equation never converge

B1 ===========

B2 ==========

B1 =k

B2= -k/2

If both variables are having relation, then why we need to include two variables

We can include only one variable, which one?

1. Related error

Errors are no auto correlated

There should be no auto correlation in errors

X y y\_pre error

1 2 3 1

2 3 5 2

3 4 8 4

1 2 4

Durbon - watson test :

1. Errors should follow normal distribution:

Q-Q plot: Quantile -Quantile plot

Task for you: Khan academy in youtube

1. Error follows homo scadasticity

Homo scadasticity vs hetero scadasticity

Homo= same

Scadasticity= spread

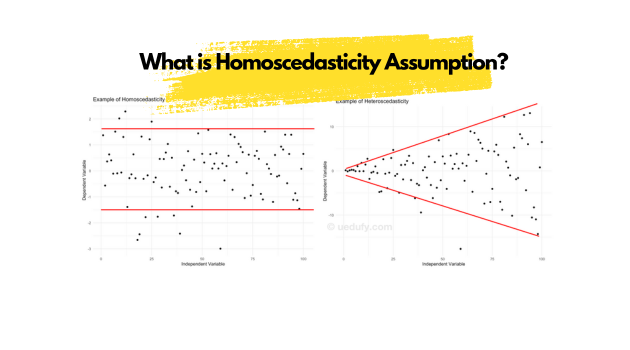
If we have a population of data has 10k observations

You divide into 10 parts each part has 1k observations

Then we apply LR model then we get the errors 1k

If you plot the distribution over that 1k observations ==== variance

And that variance should be same for all parts



All the parts of variance distribution should be same

Q1) what are LR assumptions

Q2) Multicollinearity

Q3) VIF

Q4) DW test

Q5) Homo scadasticity