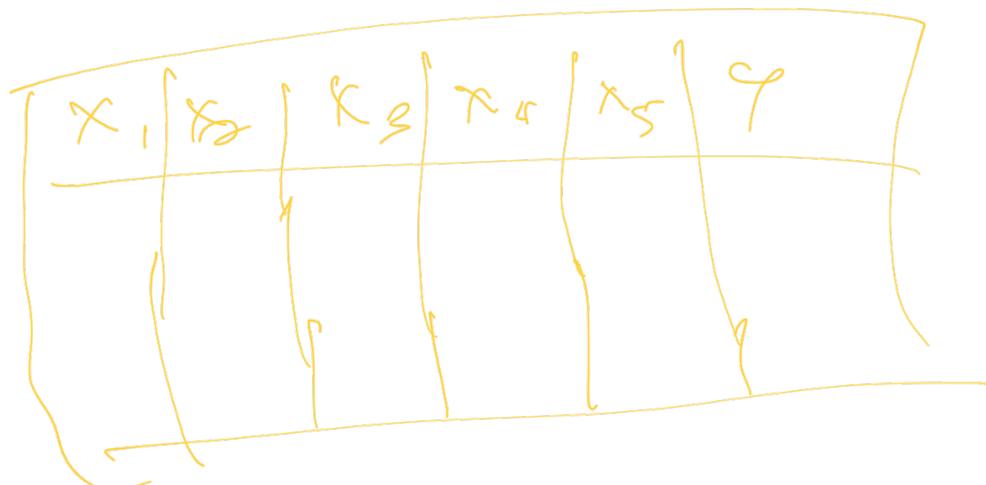


Assumptions of Linear regression

① Assumptions

- i) linear relationship between X & Y variables
- ✓ ii) Data should have little to no multicollinearity
- ✓ iii) Homoscedasticity \rightarrow Residuals should have homoscedasticity
- ✓ iv) The residuals should be normally distributed
- ① There should be linearity



X & Y
when X changes
 Y should
also change



② little to no multicollinearity



Salary & Experience

weakly depend on age

strongly dependent on experience

strong interrelationship with age

Indep. variable
Nishan's understanding of dependent variable
Person Q?

Irrelevant actual job felt.

This will affect the prediction

indirectly

{ Experience Coeff } { Non significant towards pred. }

{ Age Coeff } { towards pred. }

↳ Significant factor towards

age just little contribution

age salary

Slope $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

what are methods Ideals, Multicollinearity

① Correlation Matrix



② $VIF \rightarrow$ Factor \rightarrow Goodness of fit
 Variation Inflation Factor of Regn \rightarrow D.R. Total Var

$$Y = C + M_1 x_1 + M_2 x_2 + \epsilon$$

$$X_1 = C + M_1^* x_1 + M_2^* x_2 + \epsilon$$

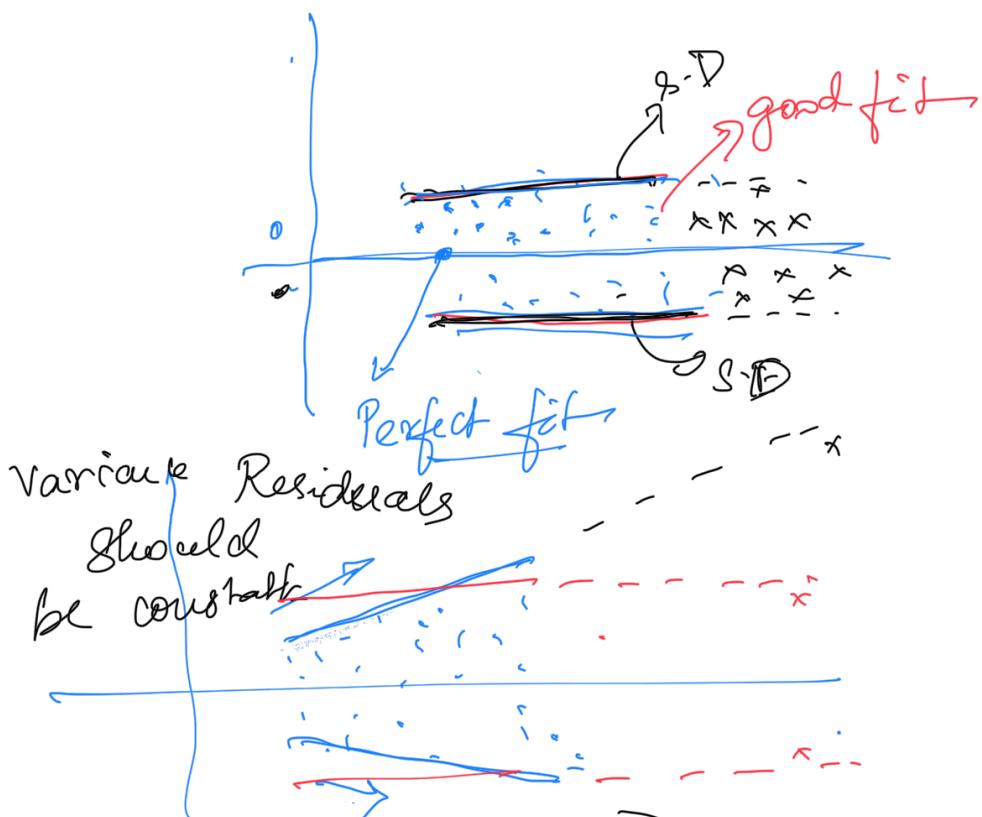
$x_2 =$ feature $\rightarrow [VIF > 10 \text{ should be}]$

Separate Colm	VIF	handled.
age	5.62	
Exp	6.83	
Second	0.03	Multicollin.
Third	10	
Handle		

- ① do nothing
- ② drop \rightarrow drop outliers / omitted bias.
- ③ Combine the outliers
make it a weighted



Heteroscedasticity of the Residuals



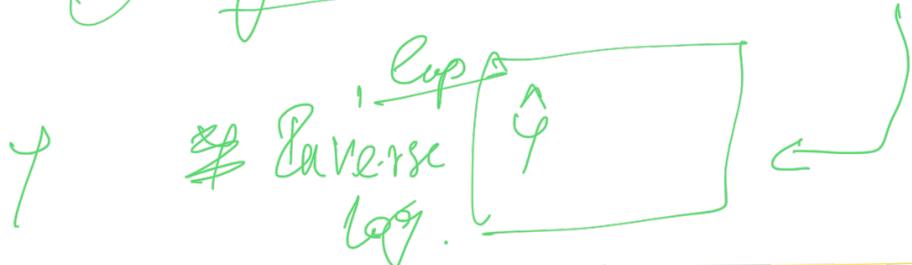
Methods :-

A Residual Plot !

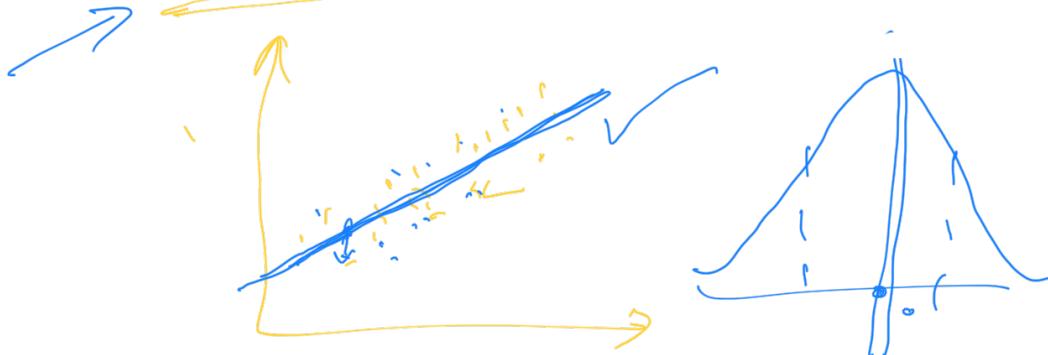
Curvilinear

Remedy :-

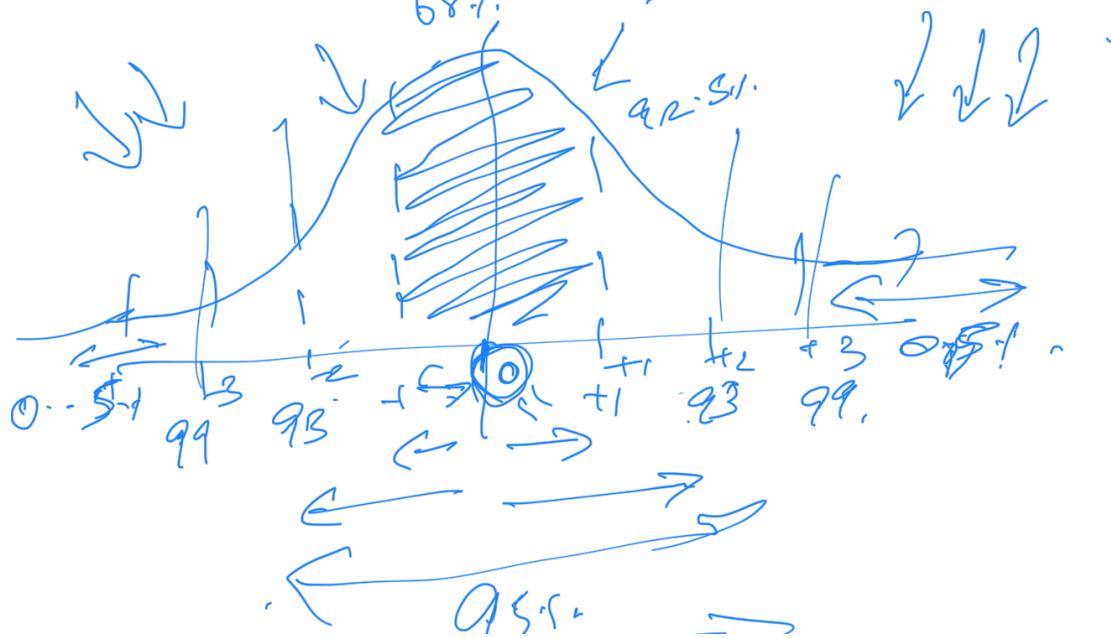
(2) log (transformation) Y

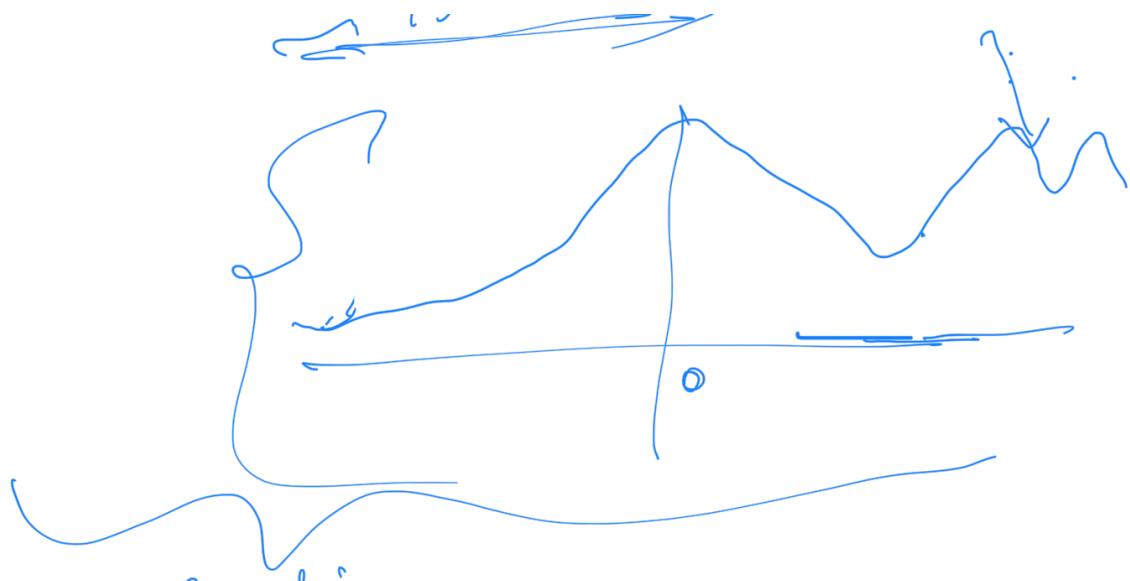


(4) The residuals must be randomly distributed



→ (1) Regression or the best fit line passes through the Centroid





① fmuch's
 ② fmuch's → fm libraries

Break till

11:20 am IST

