Assumptions ~~for~~ Linear Regression:

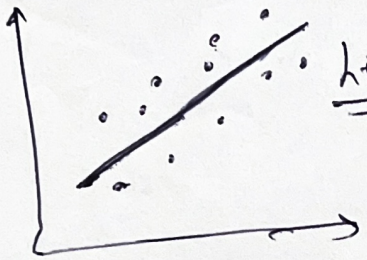
- Linear Relationship
- Multivariate Normality
- No or little Multicollinearity
- No auto-correlation
- Homoscedasticity

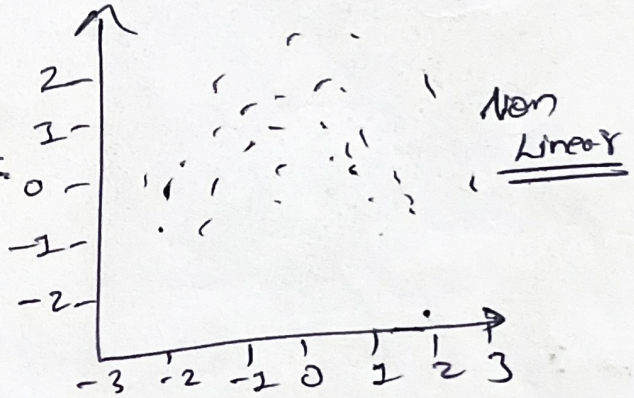
I st Linear Relationship

① Linear regression needs the relationship b/w (dependent & independent variables) variable to be linear.

② It is also important to check outliers since linear regression is sensitive to outliers effect.

Test: Can be tested with scatter plots





II nd Multivariate Normal

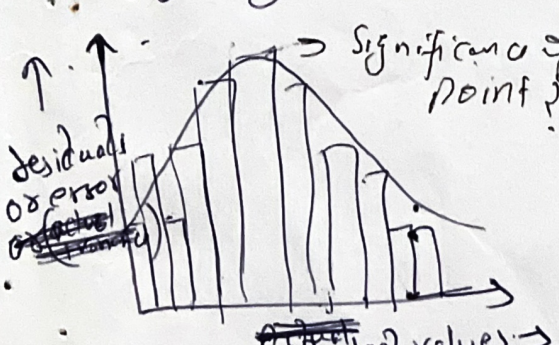
It assumes that the residuals are normally distributed. A plot of standardized residuals versus predicted values can show whether points are equally distributed across all values of independent variables.

Test: Can be checked with a histogram or a Q-Q plot.

Fix: When data is not normally distributed a non-linear transformation (eg. log transformation) might fix this issue.

III rd Homoscedasticity

This assumption means that variance around the regression line is same for all values of the predictor variable (X). The plot shows violations of this assumption. For lower values on X-axis.



→ Significance of this point?

Diff. b/w residual & Variance?