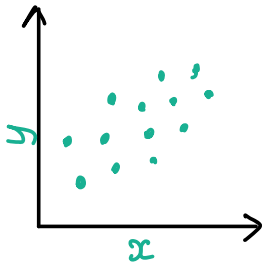# Linear Regression Assumptions

## ① Linearity → b/w target and attributes



**check**
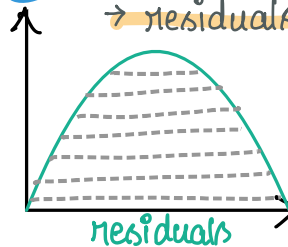* Scatter Plot
* Pearson correlation

**fix**
* Data transf. (eg: log)
* use non-linear model

**ensures**
* low biased model
* avoid modeling non-linearity

## ④ Multivariate Normality → residuals normaly distributed


residuals

**check**
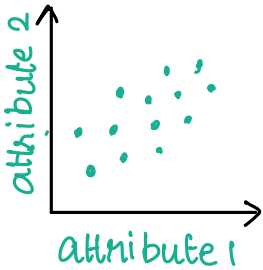* Histogram, QQ Plot
* Normality tests (eg: Kolmogrov smirnov)

**ensures**
* model can explain data
* Data comes from random dataset

**fix**
* use non-linear models

## ② Multi Collinearity → high correlation b/w attributes



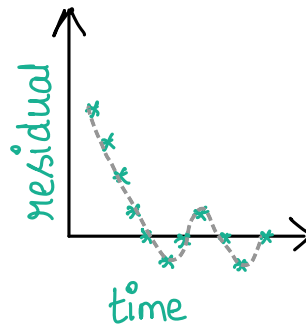**ensures**
* causal inference
* explainable model

**check**
* pairwise plot
* pearson corr  > 0.80
* Variance Inflation Factor
  $\approx 1 \rightarrow$ no corr
  $1 \leq VIF \leq 5 \rightarrow$ mod corr
  VIF > 5 → high corr

**fix**
* PCA
* Lasso, Ridge
* Recursive feature elemination

## ⑤ Auto correlation → current residual dependent on previous residual


time

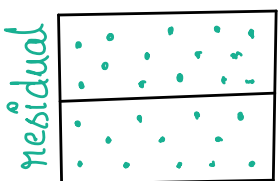**ensures**
* avoid modeling non-linearity

**check**
* residual / time plot
* Durbin-Watson test

**fix**
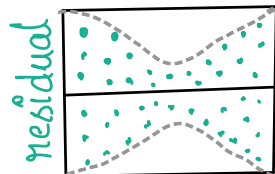* Try generalized least squares

## ③ Homoscadasticity → residuals with constant variance


Predicted
Homoscadasticity


Predicted


Predicted
Heteroscadasticity

**fix**
* data transformations
* weighted regression

**ensures**
* Outlier doesn't weigh down regression

**check**
* Breusch-Pegan test
* predicted-residual plot