Assumptions of L.R.

① Assumption of Linearity: We assume that the data can be predicted using a straight line (hyperplane). It means the independent variables (features) & the target variable should have linear relationship.

② No multicoleaniarity:

What is colinearity?

Ans: Suppose we have two features f_1 & f_2 and if

$$f_1 = a_2 f_2 + a_1$$ then f_1 & f_2 are colinear

Multicolinearity: Multiple features are colinear:

$$f_1 = a_1 + a_2 \cdot f_2 + a_3 \cdot f_3 \Rightarrow$$ f_1, f_2 & f_3 are multi-colinear.

How is it a problem?

Suppose we found out $\vec{w}$ as $[1, 2, 3]$ & $w_0 = 5$

$$\therefore \hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_0$$

$$\therefore \hat{y} = x_1 + 2x_2 + 3x_3 + 5$$

Now suppose x_1 & x_2 are colinear and their colinearity is described as: $x_2 = 1.5\, x_1$

$$\therefore \hat{y} = x_1 + 2(1.5\, x_1) + 3x_3 + 5$$

$$\hat{y} = 4x_1 + 3x_3 + 5$$ $\therefore$ our classifier $(\vec{w}) = [4, 0, 3]$

will be same as $\vec{w} = [1, 2, 3]$

But we know that higher the value of w_i, more important

But we know that higher the value of w_i, more important the feature is. ∴ According to original $\vec{w}$ [1, 2, 3], feature x_1 was least imp. but with the new $\vec{w}$ [4, 0, 3], it becomes the most imp. feature!

∴ We will not be able to identify feature importance

How to deal will multicolinearity?

Ans: VIF (Variance Inflation Factor)

① To calculate VIF, we first consider one of the factors as 'y' and the others as 'x'

f_1 f_2 f_3 - - -	f_d
⟵ X ⟶	⟵ y ⟶

② Then we train a Linear Regression model for these new X & y.

③ After training the model, we compute the R^2 score of the model. We will call this as R_j^2 (R^2 score of feature f_j)

④ Then we calculate VIF as:

$$VIF = \frac{1}{1 - R_j^2}$$

Range of VIF: $[0, \infty]$

if $R^2 = -\infty \Rightarrow VIF = \frac{1}{1-(-\infty)} = 0$

if $R^2 = 1 \Rightarrow VIF = \frac{1}{0} = \infty$

But In most cases, values of R^2 will be between 0 to 1

$\therefore$ case-1: $R_j^2 \approx 1$

$\Rightarrow VIF \approx \infty$

→ High R_j^2 means the Feature is highly colinear

→ $\therefore$ we can drop this feature

case-2: $R_j^2 \approx 0$

$\Rightarrow VIF \approx 1$

→ Low R_j^2 means the feature is not highly colinear

→ $\therefore$ Don't remove this feature

We do this process for each feature. (calculate the VIF of each feature & based on VIF we will either drop that feature or keep it.)

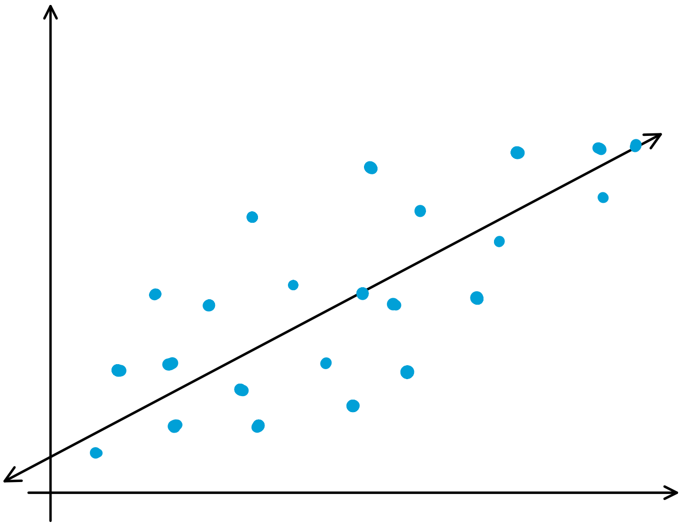
→ **Practically,**

$VIF > 10$: Highly colinear feature (drop)

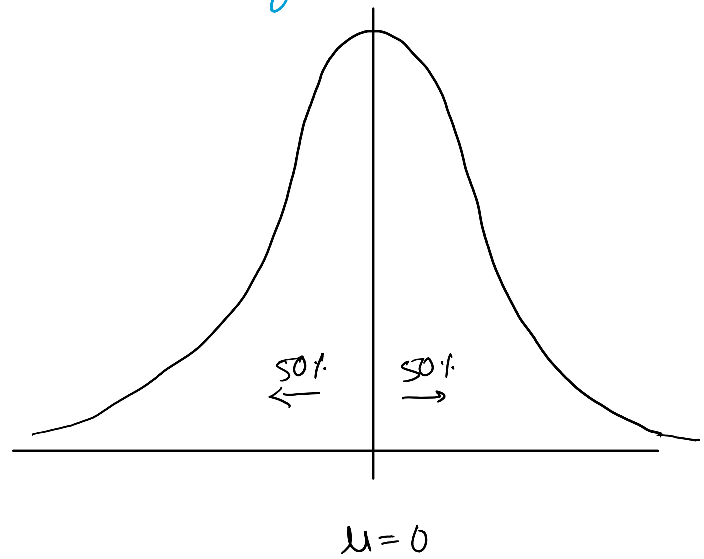
$5 \leq VIF \leq 10$: Highly colinear feature (Think about the other aspects and then decide whether to remove or to keep)

$VIF < 5$: Low multicolinearity (Don't remove it)

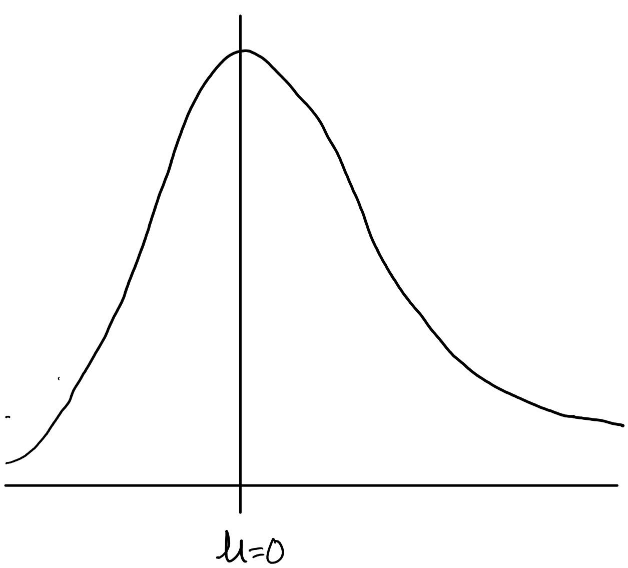
③ **Normality of Residuals:** The histogram of errors must exhibit normal distribution.

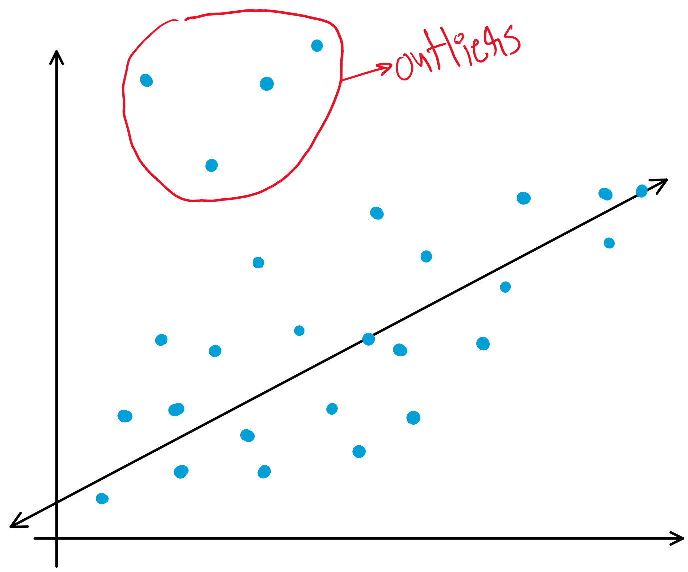


Histogram of errors

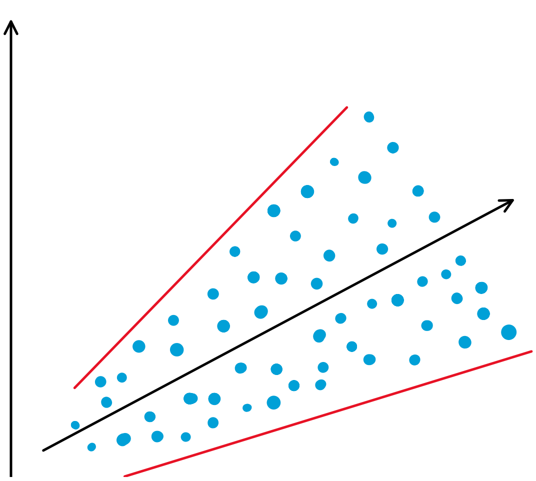


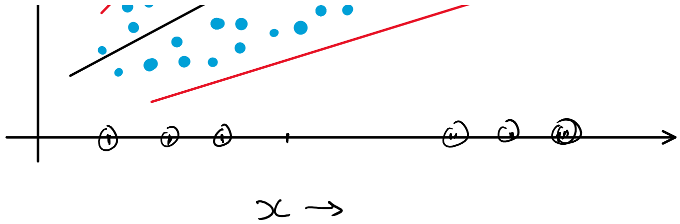
Case-2: Right skewed.



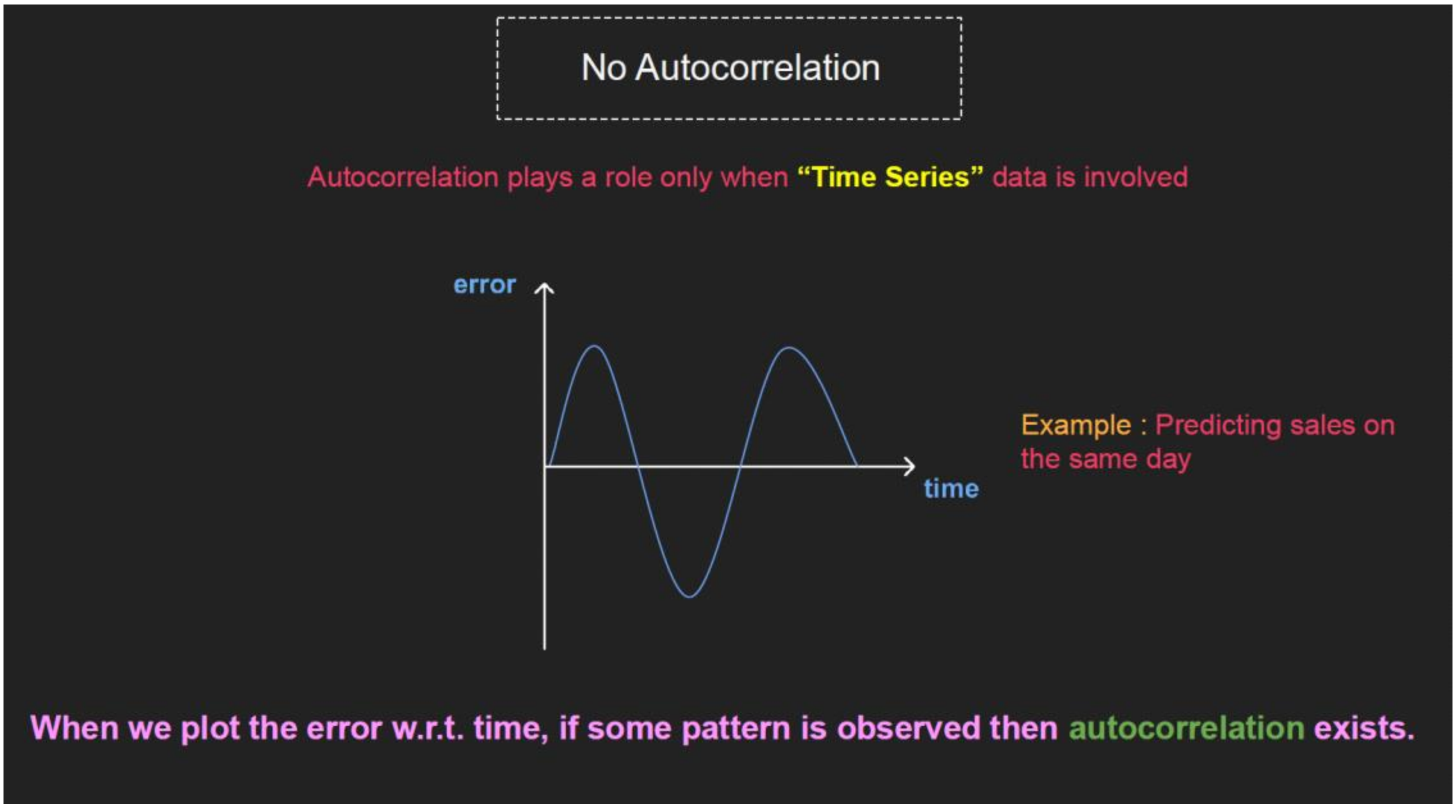


④ No Heteroskedasticity





③ No Autocorrelation (seasonality):



Examples:- ① Predicting evening sales of a restaurant from the data of morning sales & afternoon sales

② Predicting sales of umbrellas in January based on sales data from June to Dec.