Confounding vs Collinearity

They are both related to the relationship between the **predictors** in the statistical models

Confounding

its a variable *predictor* X that influences both the **Independent variables**(Other predictor)X_i and the **Dependent variable**(Response)Y.
Effect :

- It effect our estimates of the relationship between a *predictor* and a *response* making it **biased**

Example
Studying effect of balance X_1 on default Y

- **Confounder** Variable here is the **income** cause
  - Low income people might higher balances X_1 and thus higher default risk Y
  - If **income** is unmeasured $X_1 \to Y$ estimate is **biased**
  - The **income** is effects both the predictor X_1 **balance** and the response Y **default**

How to handle Confounding

- Include the **Confounder** as a control variable in the regression model
- **Stratified analysis** analyzing subgroups separately (**high income - low income** groups)
- Randomizing generally decreases confounder variable effects

Collinearity (Multicollinearity)

Previously explained in Other Considerations in the Regression Model
Collinearity occurs when two or more predictors variables are highly **correlated** making it hard to isolate the **indivudual** effect of each

Effect :

- increases **variance of coefficient estimates**
- Causes **unstable** model estimates , the smallest changes in the data will result in large swings in the coefficients estimates
- Makes interpreting coefficients harder

Conclusion

- Confounding makes the estimates biased
- Collinearity **inflates** the variance of the coefficients