ASSUMPTIONS FOR LOGISTIC REGRESSION

1. Linearity of Log-Odds:

The relationship between the independent variables and the log-odds of the dependent variable should be linear. This implies that changes in the independent variables are associated with a constant change in the log-odds of the event occurring.

2. Independence of Observations:

Observations should be independent of each other. This means that the value of the dependent variable for one observation should not be influenced by the values of the dependent variable for other observations.

3. Absence of Multicollinearity:

There should be little to no multicollinearity among the independent variables. Multicollinearity occurs when two or more independent variables are highly correlated, making it difficult to separate their individual effects on the dependent variable.

5. No Endogeneity:

There should be no endogeneity, meaning that the independent variables are not influenced by the dependent variable. Violation of this assumption can lead to biased estimates.

6. No Perfect Separation:

Perfect separation occurs when the logistic regression model can perfectly predict the outcome variable for certain values of the independent variables. This can lead to estimation problems, and adjustments may be needed (e.g., Firth's correction).

7. Adequate Sample Size:

Logistic Regression tends to work well with larger sample sizes. A rule of thumb is to have at least 10-15 events (cases where the dependent variable is 1) per independent variable to ensure stable estimates.

8. Binary or Ordinal Outcome:

Logistic Regression is designed for binary or ordinal dependent variables. For a binary outcome, the dependent variable should have two categories (0 and 1).