Generic Regression Model

Data

Leverage Points

Data Cleaning (NA, Numeric, sd ==0)

Influential Points

Outlier analysis(??)

If there is one explanatory variable

No

Yes

Response

Sensitivity analysis

Discrete

**(C)** Continuous Response

Count data non negative integer value (0, 1, 2,) and these integers are generated from counting rather than ranking

Binary (0,1)

Count Data

**Binary Response (0, 1)**

Glm Poisson

Logistic

Classification **(??)**

By histogram/density plot

**(C)** Continuous Response

Using chi-square goodness of fit test

Normal Distribution

No

Yes

Use transformation to make it normal (Box Cox )

Parametric Regression

No

Yes

**(B)** Nonparametric Regression

**(A)**  Parametric Regression

Compute correlation and dump into a data frame and then check individual correlation if it is greater than 0.7 then drop one. Do it for all explanatory variables

Handle multicollinearity manually

**(B)** Nonparametric Regression

**Nonparametric spline method**

1 Vif is not defined

**Nonparametric kernel method**

1 Vif is not defined

2 Data size should be small

**Nonparametric Gam**

1 Vif is not defined

1. Parametric Regression

**Linear Relationship**

No

Yes

**Polynomial Regression**

Quantile Regression

**Bayesian Regression**, if additional information is known about parameter

**Linear Regression**

**Glm Gamma**

**Glm inverse Gaussian**

1. Reliability Studies
2. Limited to one predictor variable

**If multicollinearity is present in the data, we can use the following two models separately**

**Lasso**

**Ridge**