

```
# TRAIN/DEV/TEST dataset
```

```
Small Data: 60,20,20
```

```
Big Data: 98,1,1      # go with No of records in DEV and TEST not the %
```

```
# TRAIN/DEV/TEST should come from same Data Distribution
```

```
## Linear Regression
```

A linear regression is used if there is relationship or significant association between the variables. This can be checked by scatterplots.

If no association appears between the variables, fitting a linear regression model to the data will not provide useful model.

```
## Logistic Regression
```