Diabetes Data:

Attribute Information:

- age age in years

- bmi body mass index

- bp average blood pressure

- s1 tc, T-Cells (a type of white blood cells)

- s2 ldl, low-density lipoproteins

- s3 hdl, high-density lipoproteins

- s4 tch, thyroid stimulating hormone

- s5 ltg, lamotrigine

- s6 glu, blood sugar level

Target:

Y is a quantitative measure of disease progression one year after baseline

1. Import Data “ Diabetes”
2. Check Data Dimensions
3. Perform MLR using Y as Dependent Variable and others as independent variable using 100% Data.
4. Check summary. Comment on R Squared.
5. Check VIF.
6. If VIFs > 5 for many its obvious that we need to resolve multocillinearity and one mtehod is to use Principal Component Regression
7. Divide the data into Train and Test data and perform MLR on Train Data. Apply the model on Test data and calculate RMSE.
8. On the same Train data apply PCR. Check PCR Model on Test Data and calculate RMSE.
9. Compare RMSE for Test Data using MLR and PCR.
10. Comment.