Simple linear regression for calories consumed dataset

We have to predict weight gained using calories consumed.

Using qqnorm and qqline function to check for the data how it is distributed.

For calories consumed it is normal.

For weight gained the values are slightly away from the line.

To check for the outliers in the data we have to use boxplot function there are no outliers in the variables.

Using summary function for details.

Use scatter plot function to check the data how it is distributed for x and y values. X is calories consumed and y is weight gained.

Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **COR** | **R Squared** | **RMSE** | **Sum of errors** |
| **Model1 (Y~X)** | 0.94 | 0.89 | 103 | 0 |
| **Model2 (Y~log(X))** | 0.89 | 0.80 | 141 | 0 |
| **Model3 (log(Y)~X)** | 0.93 | 0.87 | 118 | 0 |
| **Model4 (log(Y)~log(X))** | 0.92 | 0.84 | - | - |
| **Model5 (Y~x+I(x2))** | 0.97 | 0.95 | 70 (Not significant) | 0 |

Best model is simple linear regression

Weightgained = -625.75 + 0.42(caloriesconsumed)