**EDA of Calories Consumed**

Y – weight gain

X – calories consumed

**Mean** of X **==** 2340.714

**Median** of X **==** 2250

**Variance** of X**==** 565668.7

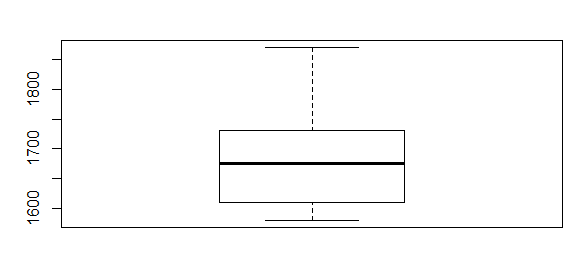
**Standard Deviation** **==** 752.10

**Range** of X**==** [1400 3900]

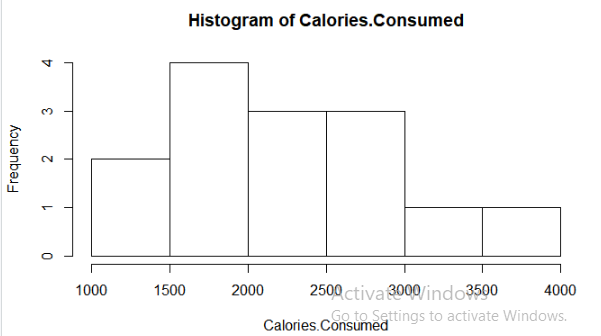
**Skewness** of X == 0.5825

**Kurtosis** of X ==2.4033

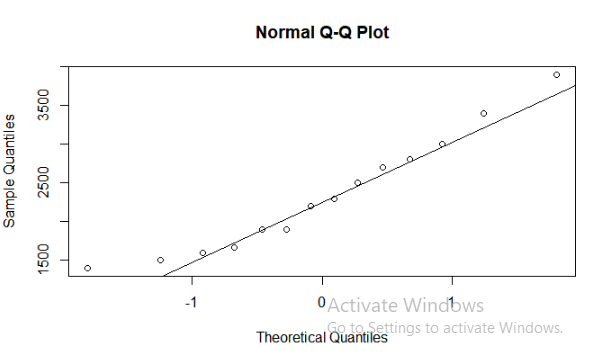
**Boxplot** of X



**Histogram** of X :



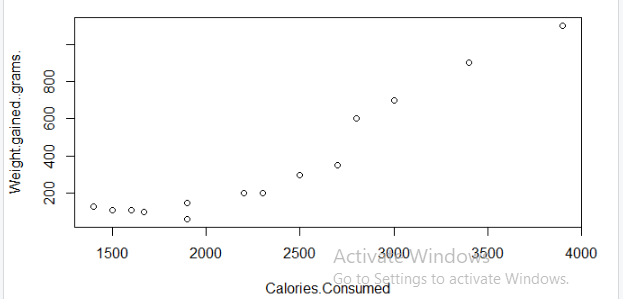
**Normal distribution** plot qqnorm and qqline of X



The mean of the input variable i.e. calories consumed lie at 2340.714 and the median lies at 2250. Since the mean of the data is slightly greater than the median, so it can be said that the data is slightly positively skewed with value 0.78. The kurtosis value 2.40 and histogram indicates that the data does not have proper peakness for a particular range but is flat distributed for most of the starting points. From the qqplot we can consider the data as normally distributed as most of the points lie on or near the qqline. The p-value from shapiro test also indicates that data is normally distributed as 0.50> 0.05

**Model Building**

**Scatter plot**  of X and Y



Plot represents a linear, positive and a moderate relationship

correlation coefficient = -0.94,so that we can say it is a good measure to predict the output ad correlation coefficient is 0.94

After building the model the best fit line equation is

Y = c + mx .

y = -625.75 + (0.4201)(x)

**Weight Gained = -625.75+ (0.4201)(calorie consumed)**

p-value<0.05

r-squared value is 0.83

95% confidence interval of

y-intercept (c) - (406.078 845.4266)

slope (m) - (0.330506 0.50980)