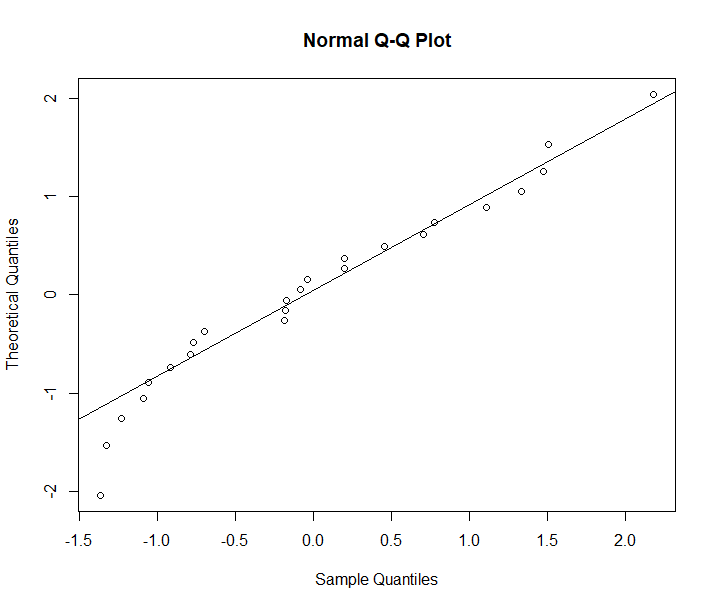
**1 (a)**

hseprice <- read.csv("houseprice.csv")

modelh <- lm(y ~ x1, hseprice)

qqnorm(rstandard(modelh),datax = TRUE)

qqline(rstandard(modelh),datax = TRUE)

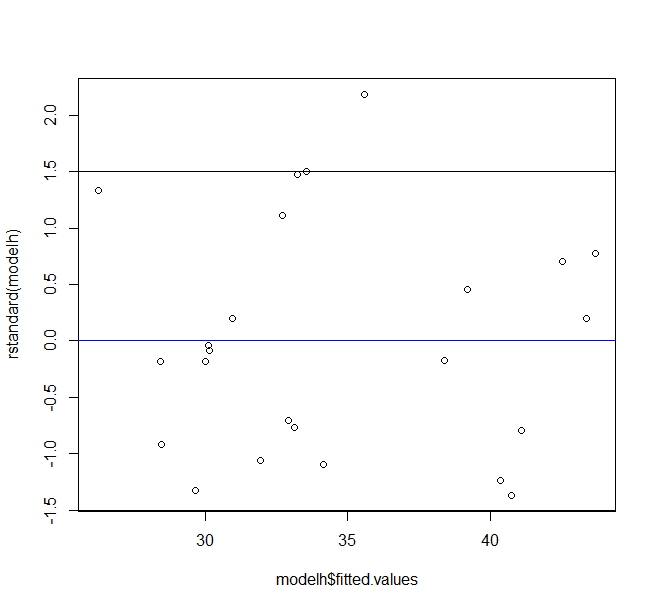


The distribution of the samples is approximately normal.  
Right tail heavier than normal – overprediction, slight deviation from normal (Right-skewed)

e = (yi – y\_hat)

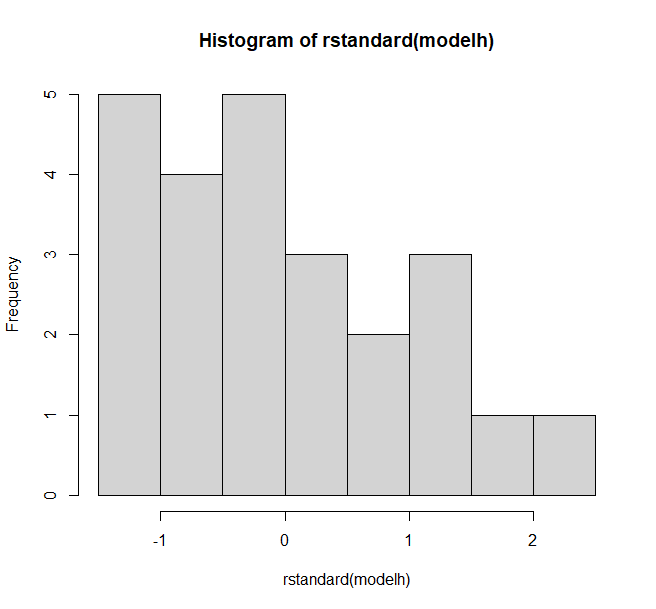
**1 ( b )**

plot(modelh$residuals, modelh$fitted.values)



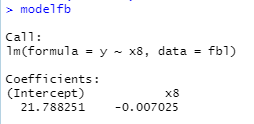
The variance of points is approximately constant, no discernible pattern to residuals.

Appears to be underpredicted for middle fitted values of .



Histogram confirms right-skewed ( -ve values in higher frequency)  
(do we use histogram to confirm? OR only focus on discernible pattern and deviance from qqline)

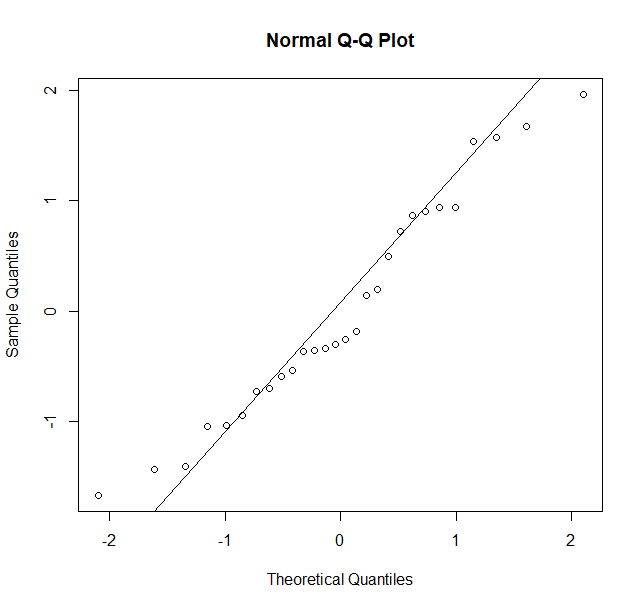
**2 (a)**



(b)

qqnorm(rstandard(modelfb), dataX = TRUE)

qqline(rstandard(modelfb))

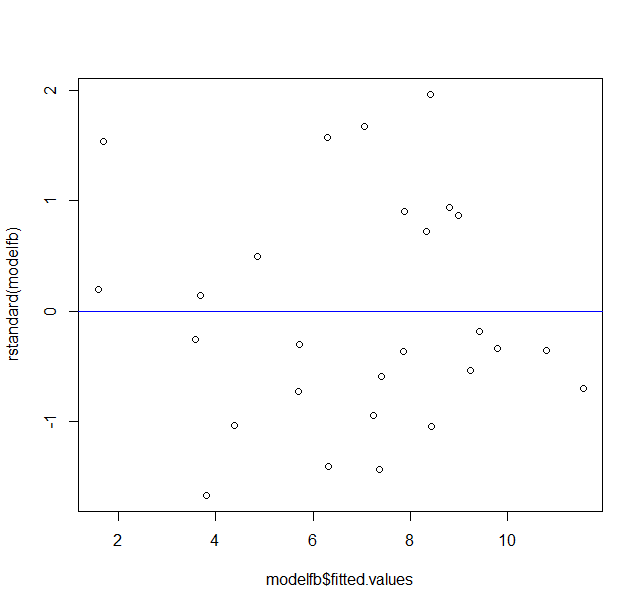


It is a heavy-tailed distribution, skewed both left and right

(b)

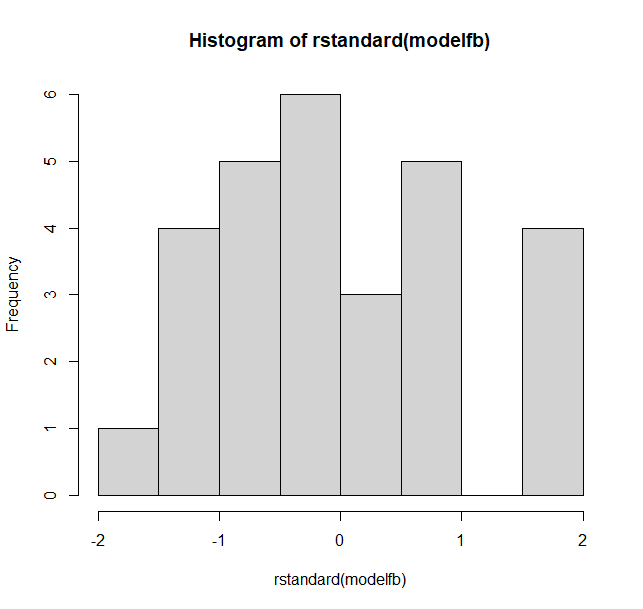
plot(modelfb$fitted.values, rstandard(modelfb))

abline(h = 0, col = "blue")

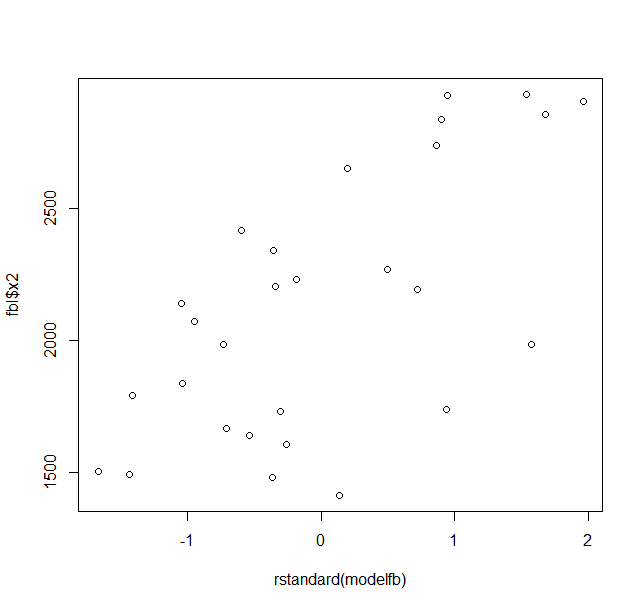


The variance appears to be constant across all fitted values at a quick glance.

There is no pattern to the std residuals.



Residuals are right skewed

plot(rstandard(modelfb), fbl$x2)  


Appears to have an upward linear trend (band), confirming a linear relationship

(e)

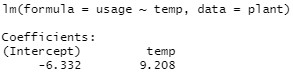
SSr(B2 | B8,B0) = 64.934

Test Statistic, F0 = = 19.33711 <

F1,25(0.05) = 4.241699

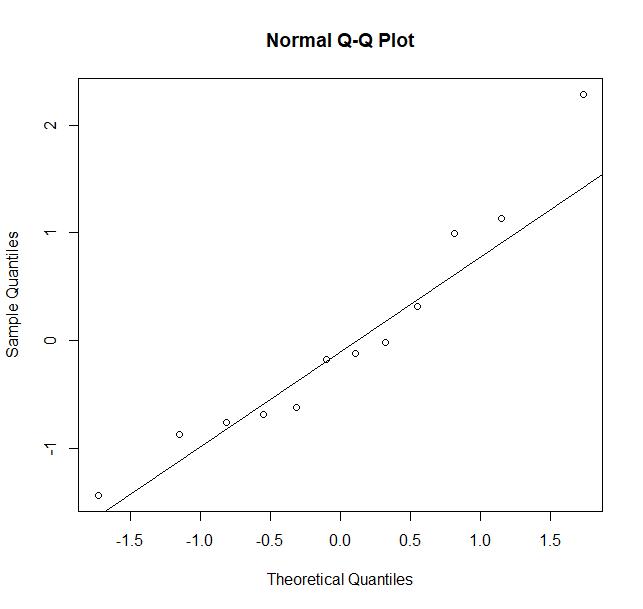
F0 < F1,25(0.05), H0 is rejected, X2 is not significant.

3(a)

modelp <- lm(usage ~ temp, plant)  


(b)  
qqnorm(rstandard(modelp), dataX = TRUE)

qqline(rstandard(modelp))

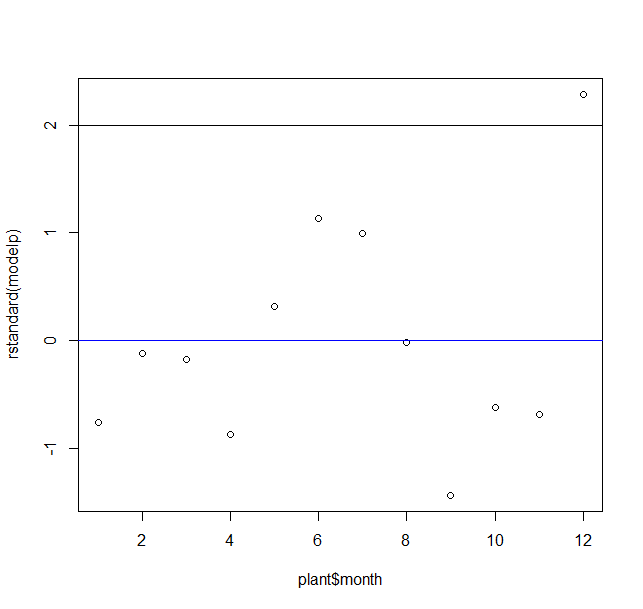


Normality assumption is ideal, with presence of outliers

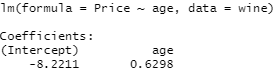
(c)

plant$month = seq(1,12)

plot(plant$month, rstandard(modelp))



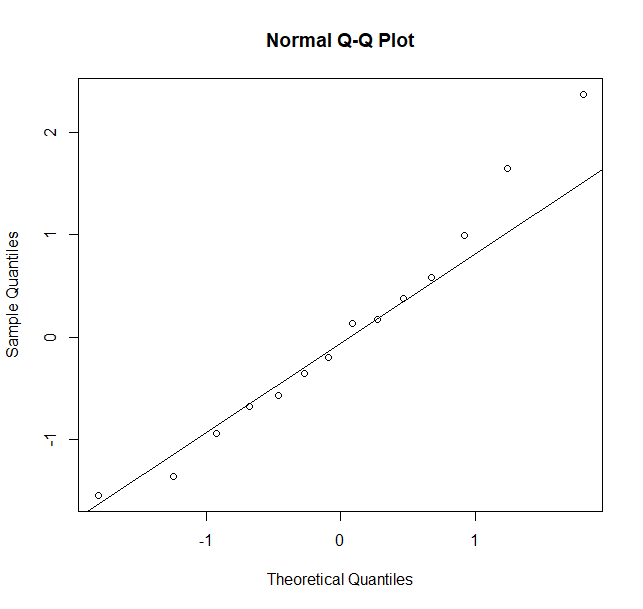
A curved scatter plot indicates nonlinearity, a higher order term on regressors is needed in the model.

4(a)  


(b)

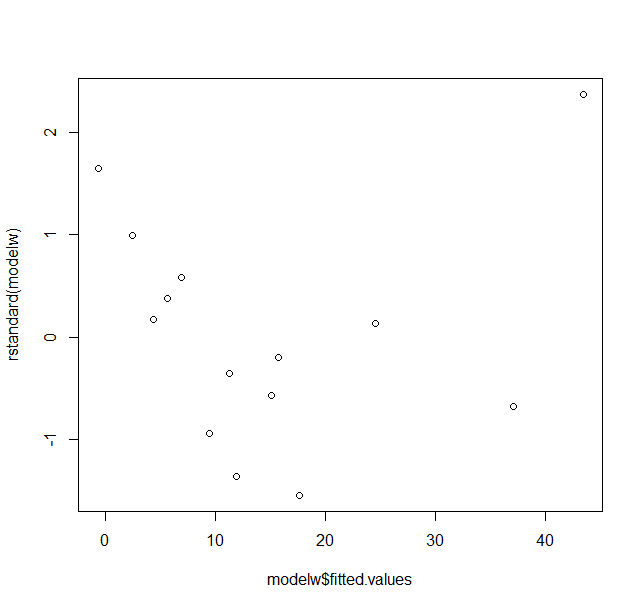
qqnorm(rstandard(modelw),dataX = TRUE)

qqline(rstandard(modelw))



* deviation from normality assumption as left tail is heavy (left-skewed graph)
* multiple outliers (over-prediction)

(c)



* Deviation from constant variance assumption( variance not in a constant band),
* Larger residuals at smaller response values, indicating underprediction
* Single > 40 has large residual ( anomaly )
* Non-linear parabolic pattern of residuals – higher order term of regressor variables are required.

