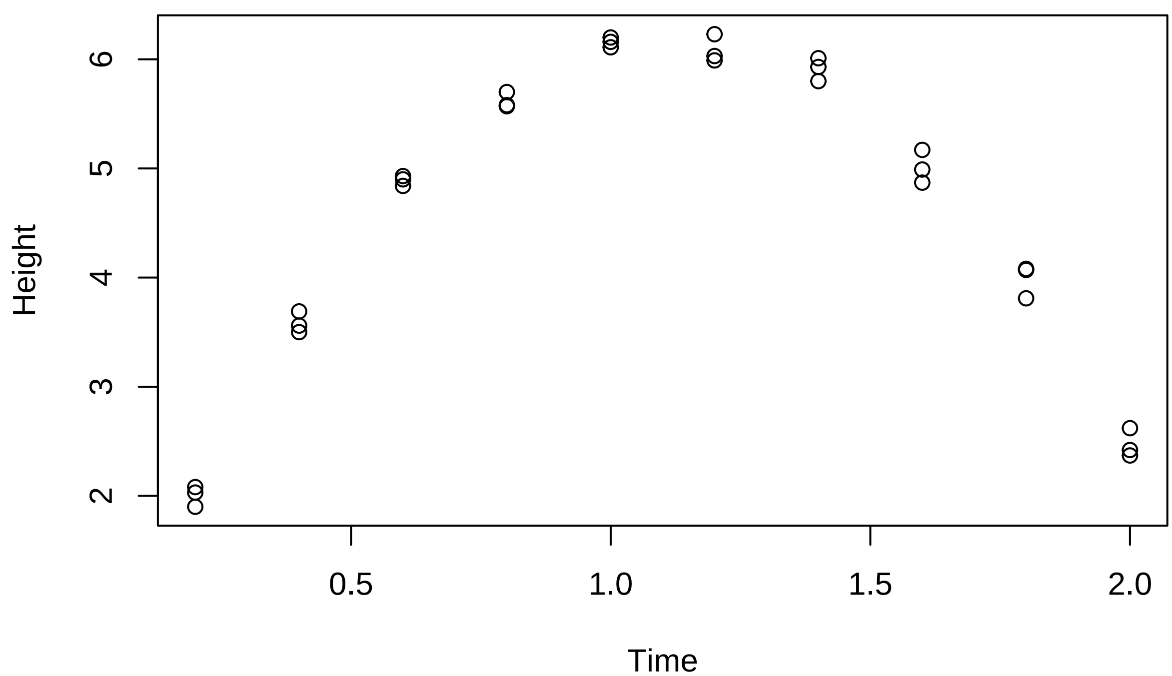
STATS 201 Lab Class 1

Liu Siyuan, 2019210173

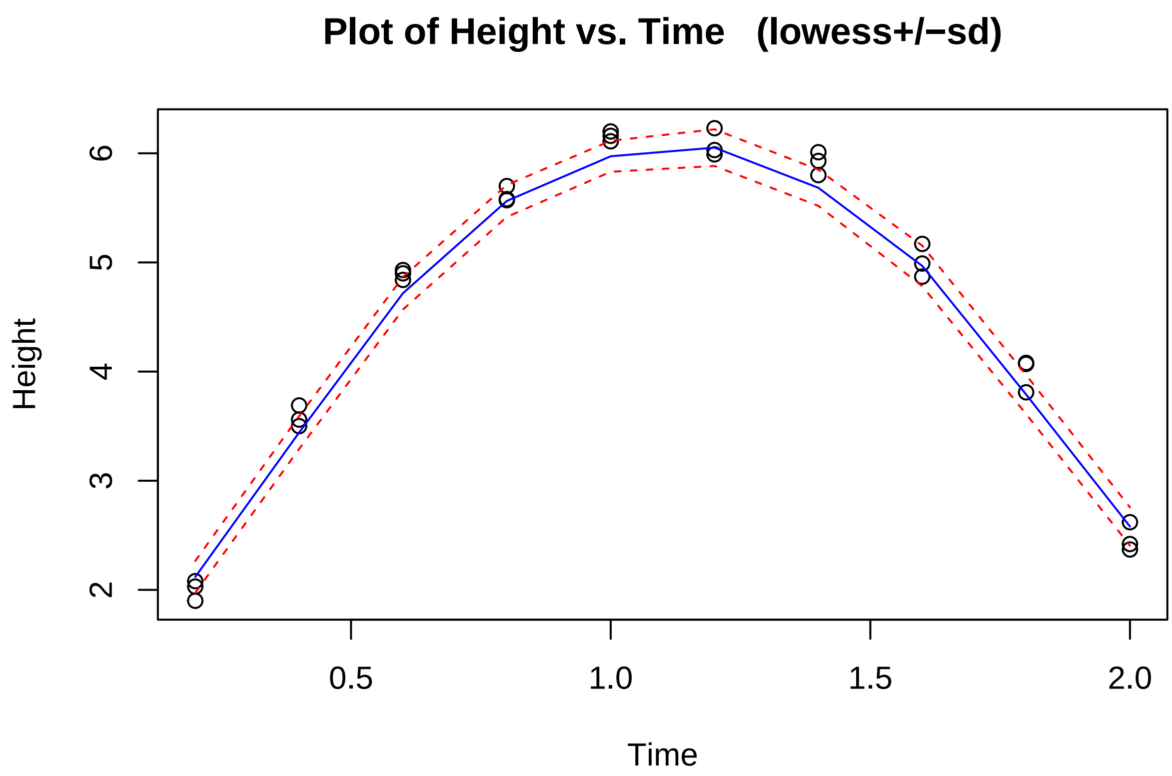
#Gravity experiment

Code and output

```
library(s20x)
## Loading in and inspect the data.
Gravity.df = read.table("GravityExpt.txt", header = TRUE)
plot(Height ~ Time, data = Gravity.df)
```



```r
trendscatter(Height ~ Time, data = Gravity.df)
```

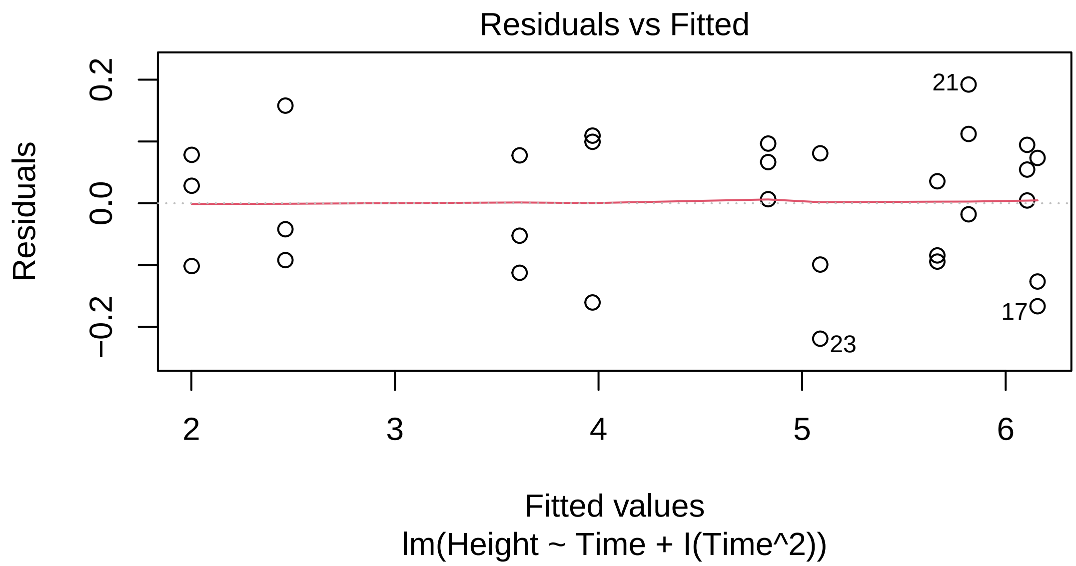


Comment

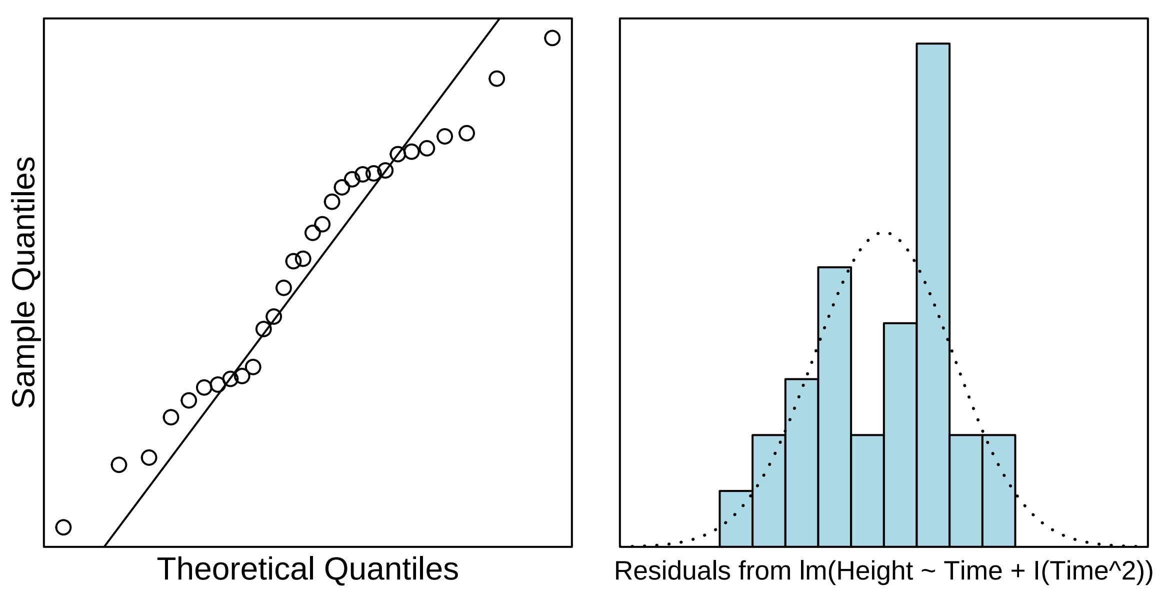
We plot the data and the trend scatter. We find that the Height and Time do not have a linear relationship. It's like quadratic. So we try to fit it with a quadratic regression model.

Fit an appropriate quadratic model, including model checks.

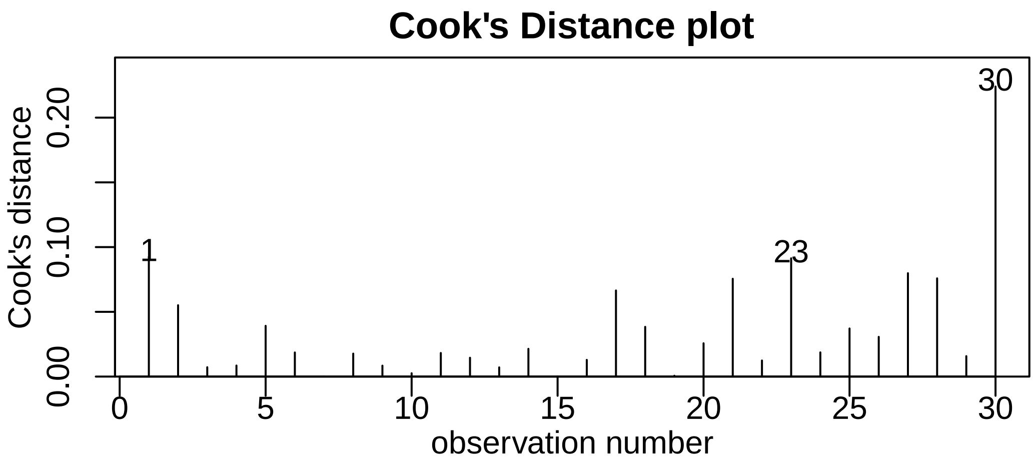
```r
Gravity.fit = lm(Height ~ Time + I(Time ^ 2), data = Gravity.df)
plot(Gravity.fit, which = 1)
```




```
normcheck(Gravity.fit)
```

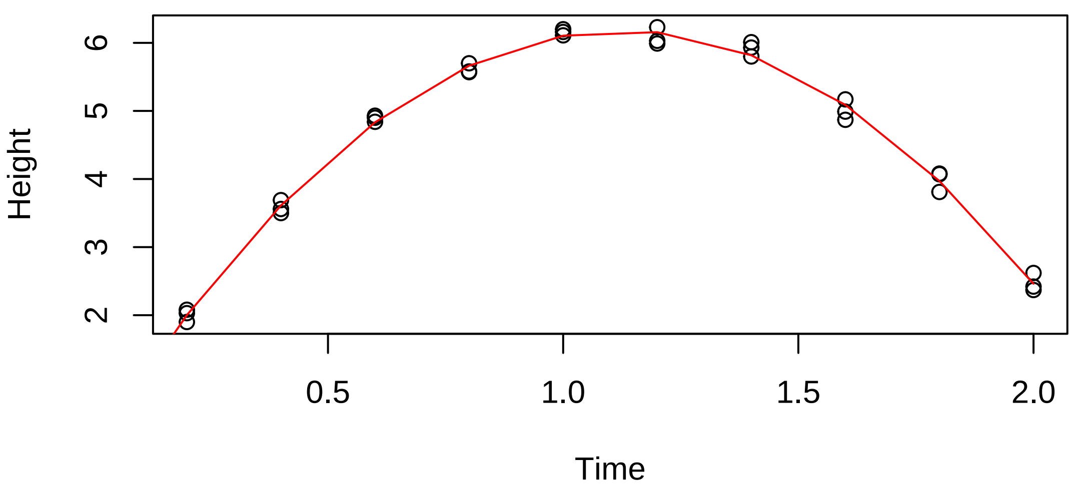



```
cooks20x(Gravity.fit)
```

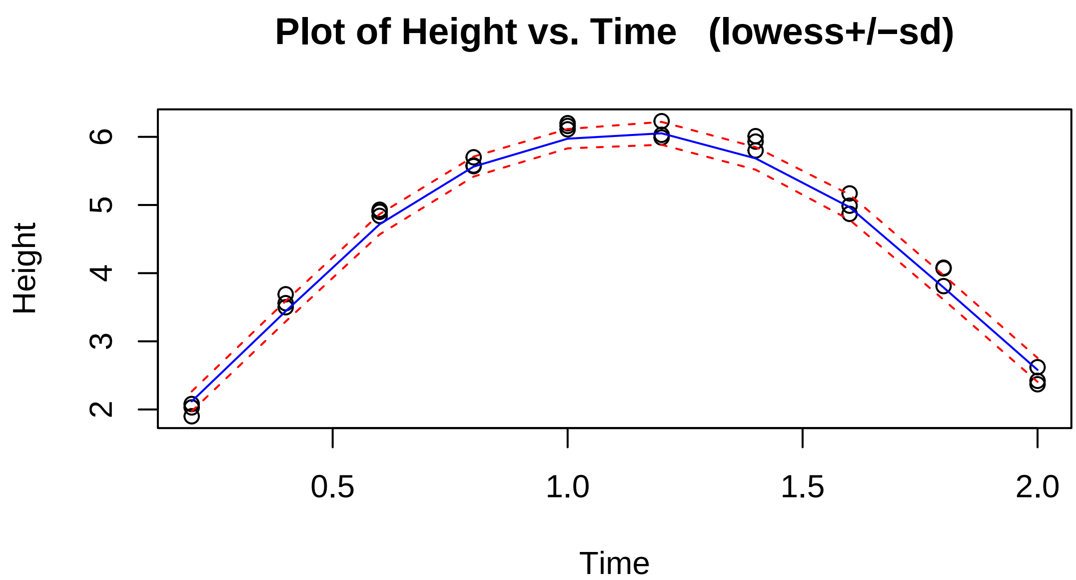


Plot the data with quadratic model superimposed over it.

```
plot(Height ~ Time, data = Gravity.df)
x = seq(0, 2.0, 0.2)
lines(x, predict(Gravity.fit, data.frame(Time = x)), col="red")
```



```r
trendscatter(Height ~ Time, data = Gravity.df)
```



#

Summay and Confint

```r
summary(Gravity.fit)
```

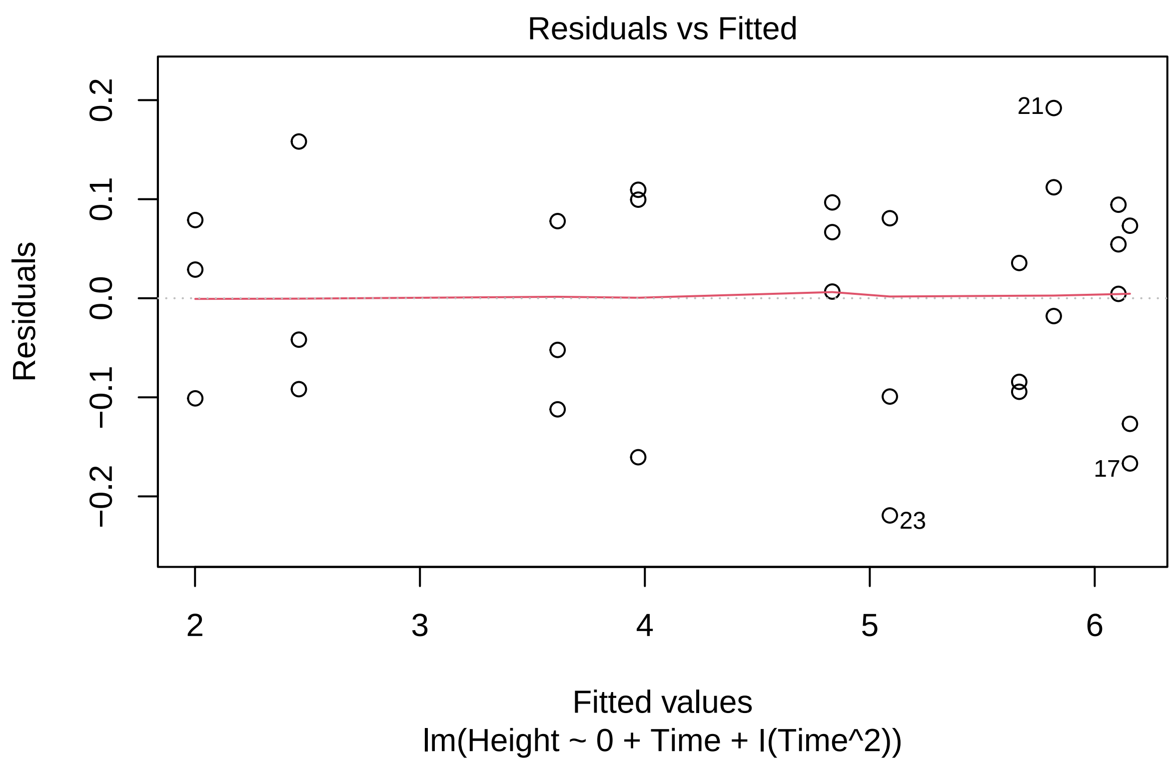
```
##
## Call:
## lm(formula = Height ~ Time + I(Time^2), data = Gravity.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21914 -0.09378  0.01751  0.08025  0.19217
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0008333  0.0748403   0.011    0.991
## Time        10.9786742  0.1562823  70.249   <2e-16 ***
## I(Time^2)   -4.8740530  0.0692301 -70.404   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1102 on 27 degrees of freedom
## Multiple R-squared:  0.9946, Adjusted R-squared:  0.9942
## F-statistic:  2505 on 2 and 27 DF,  p-value: < 2.2e-16
```

```r
confint(Gravity.fit)
```

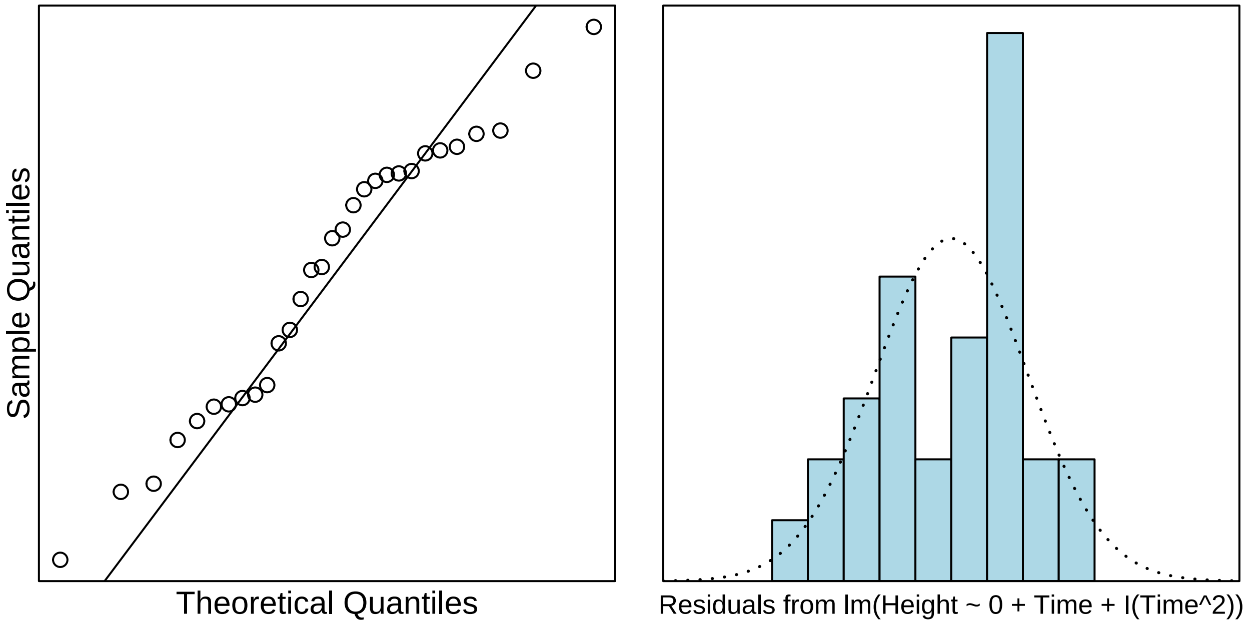
```
##                   2.5 %     97.5 %
## (Intercept) -0.1527263  0.1543929
## Time        10.6580094 11.2993390
## I(Time^2)   -5.0161015 -4.7320046
```

Refit the Model without the Intercept

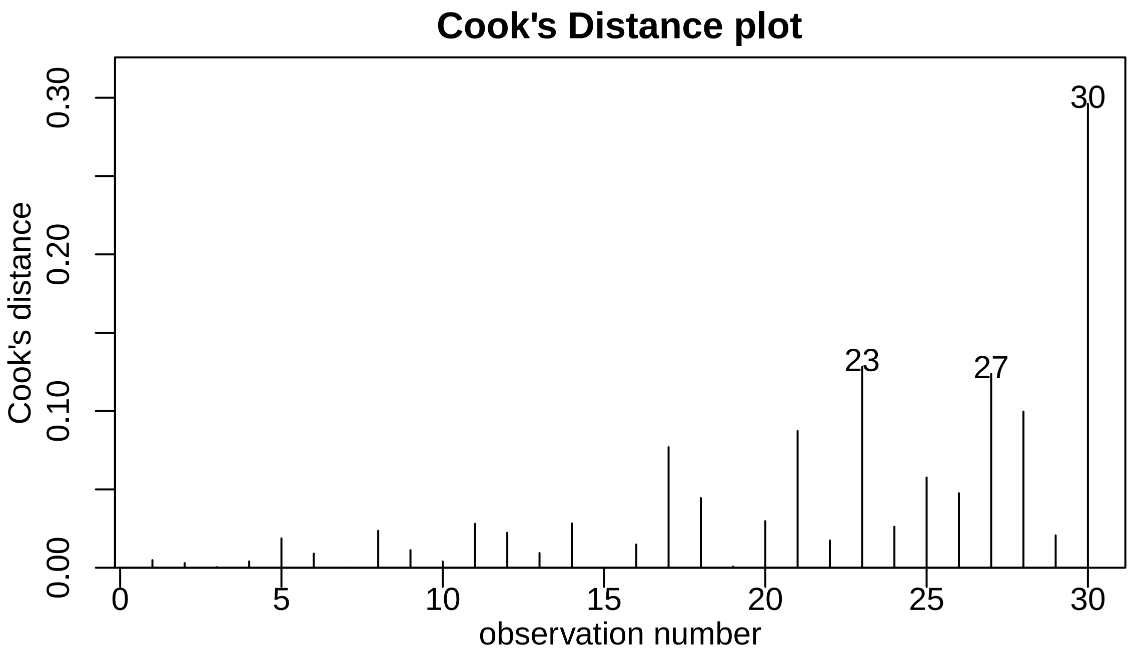
```r
Gravity1.fit = lm(Height ~ 0 + Time + I(Time ^ 2), data = Gravity.df)
plot(Gravity1.fit, which = 1)
```



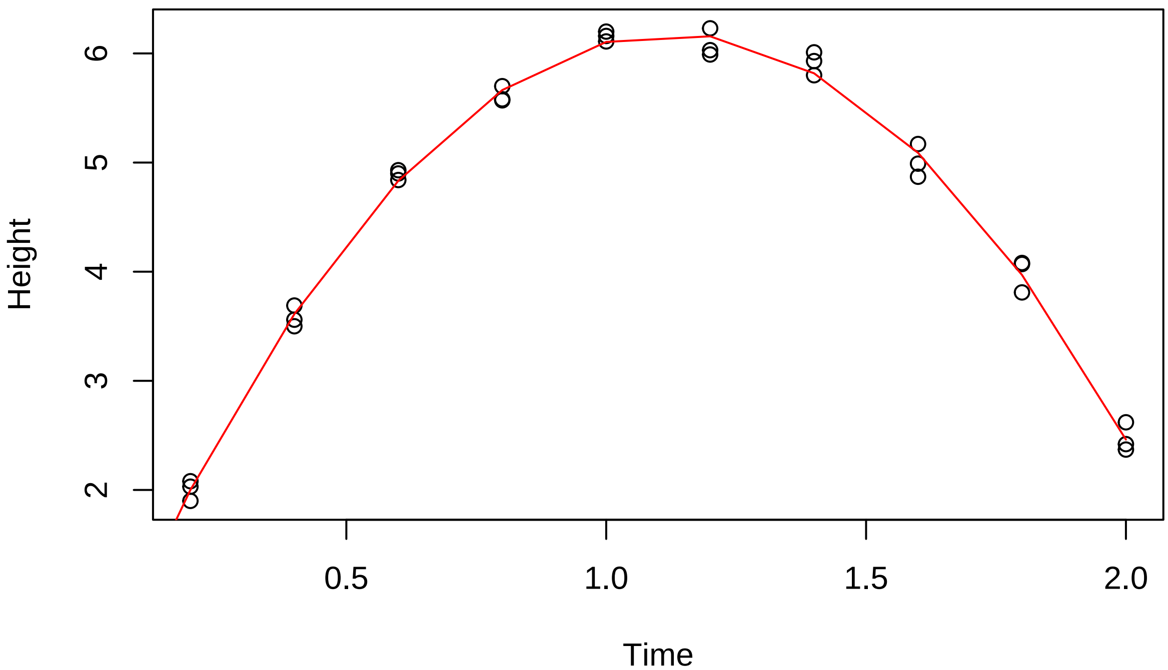

```r
normcheck(Gravity1.fit)
```



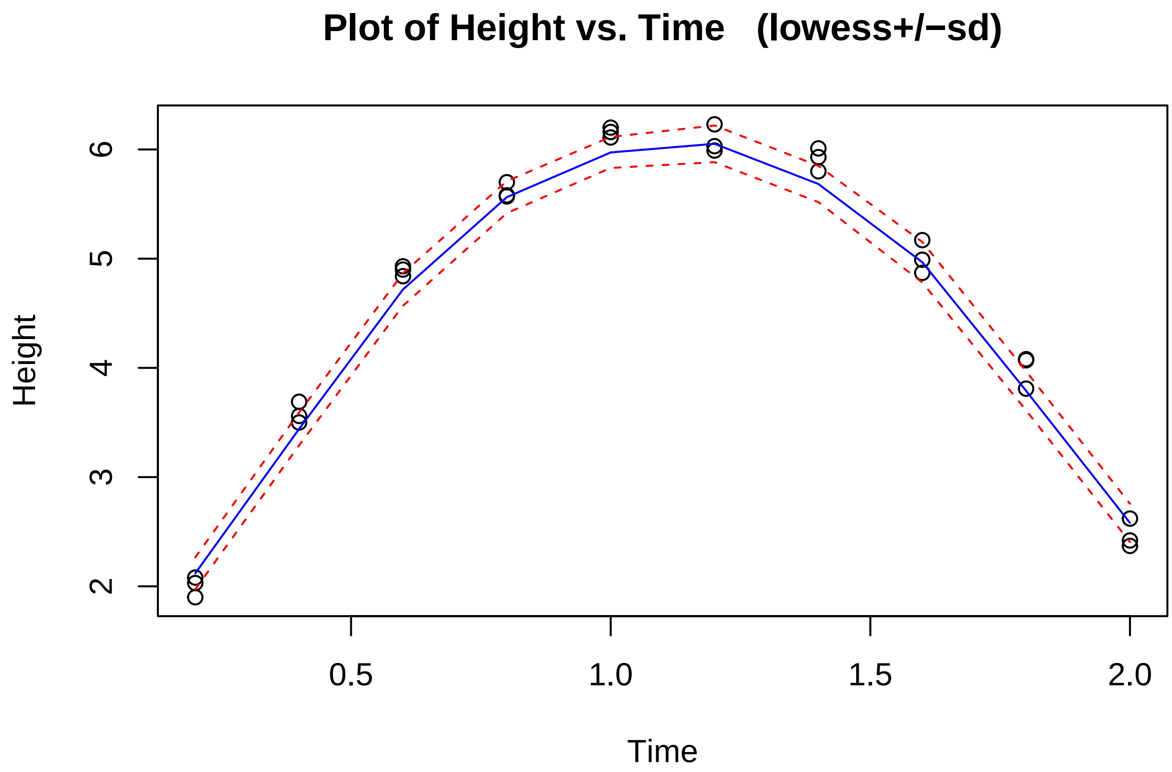

```
cooks20x(Gravity1.fit)
```




```
plot(Height ~ Time, data = Gravity.df)
x = seq(0, 2.0, 0.2)
lines(x, predict(Gravity1.fit, data.frame(Time = x)), col="red")
```




```
trendscatter(Height ~ Time, data = Gravity.df)
```




```
summary(Gravity1.fit)
```

```
##
## Call:
## lm(formula = Height ~ 0 + Time + I(Time^2), data = Gravity.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21923 -0.09375  0.01783  0.08031  0.19202
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## Time      10.98026    0.06406   171.4   <2e-16 ***
## I(Time^2) -4.87468    0.03949  -123.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1082 on 28 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 2.936e+04 on 2 and 28 DF,  p-value: < 2.2e-16
```

```
confint(Gravity1.fit)
```

```
##                2.5 %    97.5 %
## Time      10.849036 11.111475
## I(Time^2) -4.955563 -4.793798
```

Methods and Assumption Checks

We fitted a linear model with quadratic term as exploratory plots revealed some curvature. After fitting the quadratic, the residuals were fine, there were no problems with normality and no unduly influential points. We have independence from taking a random sample.

Our model is:

$$Height_i = \beta_0 + \beta_1 \times Time_i + \beta_2 \times Time_i^2 + \epsilon_i$$

where $\epsilon_i \sim iid\ N(0, \sigma^2)$ Our model explains 99% of the total variation in the response variable, so it will be reasonable for prediction.

Executive Summary

We are interested in addressing a few questions of interest in the Executive Summary.

First, we can find the estimated value of g in the summary, which is equal to $-4.874ms^{-2}/0.5 = -9.75ms^{-2}$

Second, we have 95% confidence that the $-0.5g$ is in the interval between -5.01 and -4.73. Such that we have 95% confidence that the g is in the interval between 9.46 and 10.02. We find that the confidence interval contains a theoretical value of 9.8. So we can say that our estimated value of g consistent with the theoretical value of 9.80.

Thirdly, we can conclude from the summary that the true value of the intercept term coefficient is 0, because its p-value is very close to 1, so we have to accept the null hypothesis, that is, the intercept is 0. This is consistent with our analysis.

Lastly, we can find that we have 95% confidence that the $-0.5g$ is in the interval between -4.96 and -4.79. Such that we have 95% confidence that the g is in the interval between 9.59 and 9.91. Compared to the CI for g with the CI we calculated earlier, the interval is narrowed and the estimate will be more accurate.

In conclusion, our model explains 99% of the total variation in the response variable, and so will be reasonable for prediction.