Non linear regression model

Nithya S

2024-01-12

Objective

To represent the substrate concentration (S) and the observed velocity (v) based on an enzymology experiment.

(a) Fit a non linear regression model that relates velocity to concentration using Michaelis-Menten equation .

(b) Analyse and examine whether you can fit a simple linear regression model that relates velocity and substrate concentration by using any suitable transformation.

URL for the dataset:

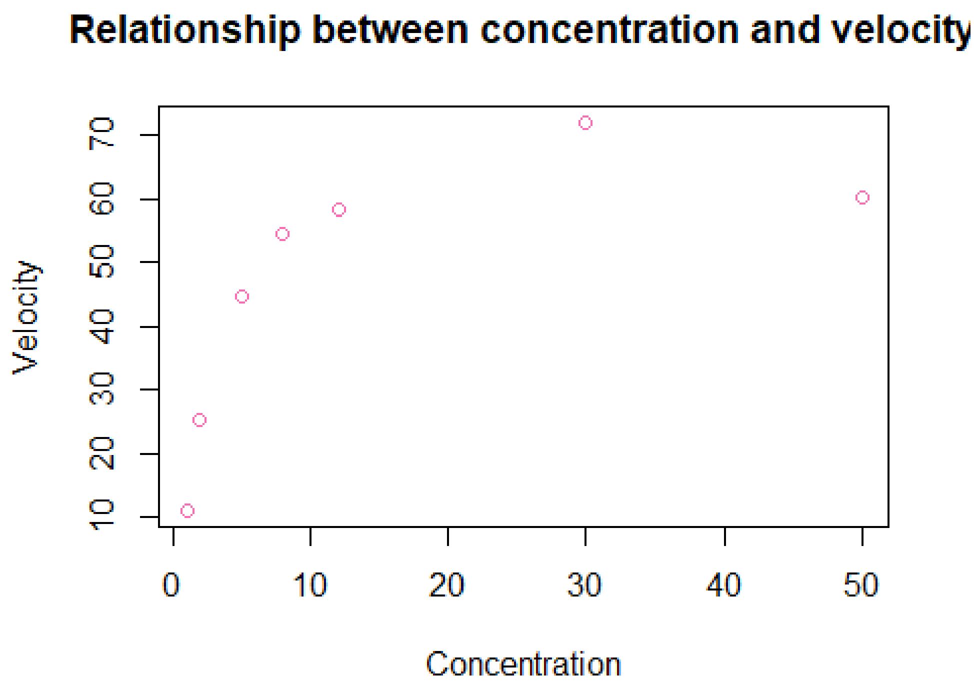
https://docs.google.com/spreadsheets/d/1X_rJrknx0vAdunPST5Dh0PQKffC5LeDE/edit?usp=sharing&ouid=110138716074493614410&rtpof=true&sd=true

We import the data to perform the analysis as follows:

```
library(readxl)
Dataset <- read_excel("G:/My Drive/Linear Regression/Datasets/Substrate
concentration.xlsx")
View(Dataset)
attach(Dataset)
```

To start off, we plot the data as follows:

```
plot(S,V,col="hotpink",main="Relationship between concentration and
velocity",xlab="Concentration",ylab="Velocity")
```

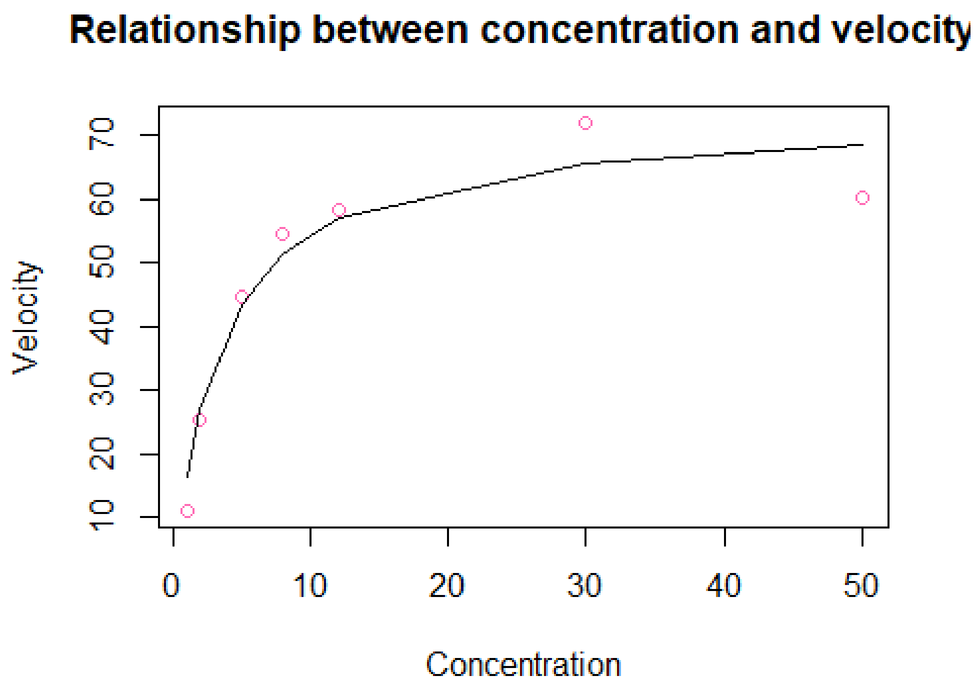


From the graph we infer as, there is an increasing[positive] relationship between concentration and velocity. We infer or decide the values of V_{max} and K_m values from the graph. The maximum velocity value (V_{max}) is 72. K_m is the concentration at which the velocity rate is half of V_{max}. From the data set, K_m is 5.

We fit the non linear regression model that relates velocity to concentration using Michaelis-Menten equation as follows.

```
MM_model=nls(V~V_max*S/(K_m+S),start=c(V_max=72,K_m=5),data=Dataset)
summary(MM_model)

##
## Formula: V ~ V_max * S/(K_m + S)
##
## Parameters:
##       Estimate Std. Error t value Pr(>|t|)
## V_max  73.2614     5.0198  14.594 2.73e-05 ***
## K_m     3.4372     0.8954   3.839   0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.594 on 5 degrees of freedom
##
## Number of iterations to convergence: 6
## Achieved convergence tolerance: 6.226e-06
```

```r
plot(S,V,col="hotpink",main="Relationship between concentration and velocity",xlab="Concentration",ylab="Velocity")
lines(S,predict(MM_model,lty=1,col="blue",lwd=3))
```



From the summary of the model, we see the estimated values of V_{max} and K_m is 73.2614 and 3.4372 respectively. The non-linear model built here is

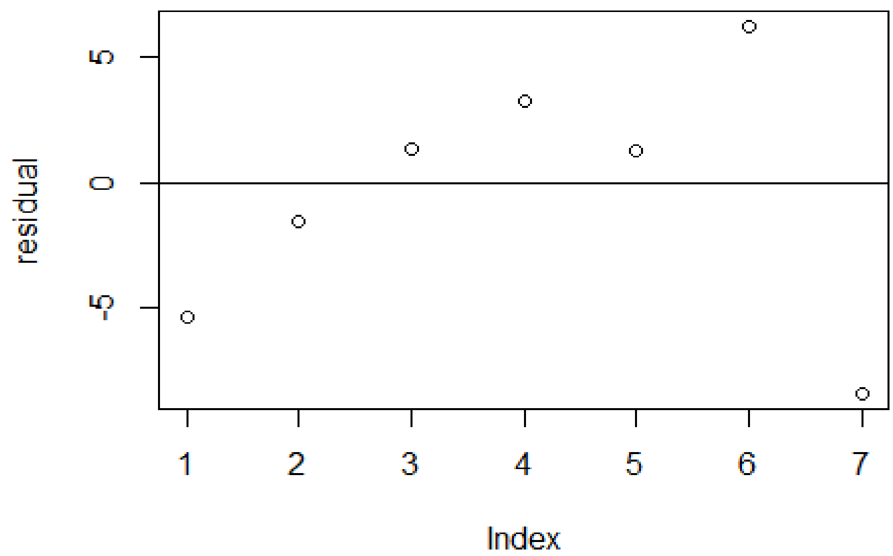
$$Velocity, V = \frac{(V_{max} * Concentration, S)}{(K_m + Concentration, S)}$$

The estimated nonlinear model is

$$Velocity, V = \frac{73.2614 * S}{3.4372 + S}$$

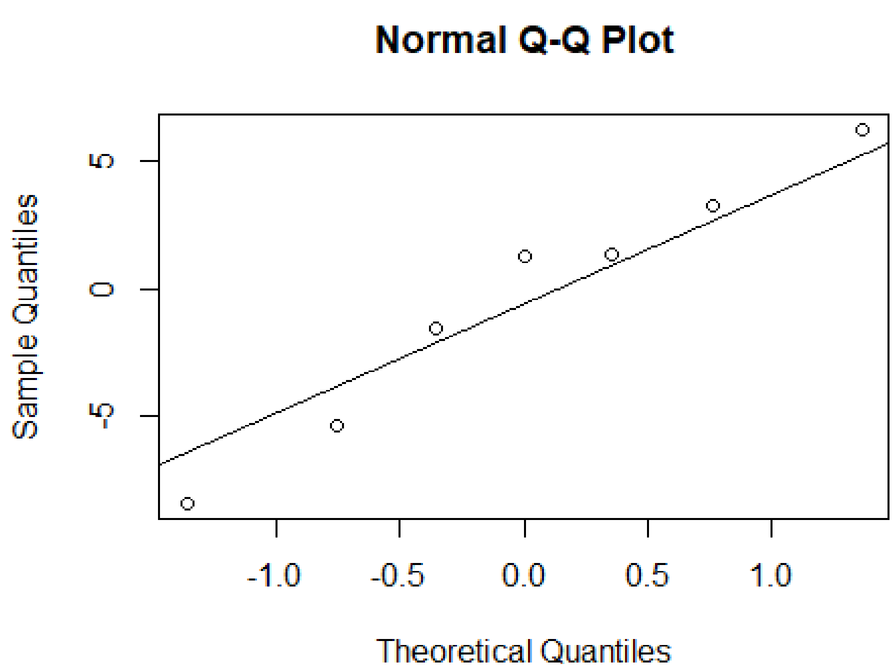
We also see from the summary that, the regression coefficients are significant at 5% significance level. Since the values are significant, we proceed for residual analysis.

```r
residual=resid(MM_model)
residual

## [1] -5.410856 -1.548379  1.384119  3.255559  1.250639  6.269496 -8.449099
## attr(,"label")
## [1] "Residuals"

plot(residual)
abline(0,0)
```



Now to check for Normality condition

```
qqnorm(resid(MM_model)) # QQ plot-of residual
qqline(resid(MM_model)) # plots the points
```



If not all, majority of the points fall on the line thus the quartile of normal and residual are almost same, hence it indicates that the residuals follow a normal distribution. However the assumption of normality has to be verified by using a statistical test.(but to know if the deviation of the points lying away from the line, we use the test to further confirm the normality.)

Hypothesis to testing for normality:

H_0: Errors follow normal distribution.

v/s

H_1: Errors do not follow normal distribution.

Normality using shapiro test

```
shapiro.test(resid(MM_model))

##
##  Shapiro-Wilk normality test
##
## data:  resid(MM_model)
## W = 0.9624, p-value = 0.839
```

At 0.05 level of significance the p value is 0.8504 which is greater than 0.05, thus we fail to reject null and and say that the residuals follow normal distribution. Hence the assumption of errors.

Since the assumptions of the residuals are satisfied we conclude that the non-linear model is possible for the given data set or is the better fit .

To analyse and examine whether we can fit a simple linear regression model that relates velocity and substrate concentration by using any suitable transformation.

Inverse linear transformation can be applied to our model. We should have a function V to be linearly related to a function of S. This can be done by taking the reciprocal on both sides of our initial model. The new model will become:

$$\frac{1}{V} = \frac{S + K_m}{(V_{max} * S)}$$

$$\frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m}{V_{max} * S)}$$

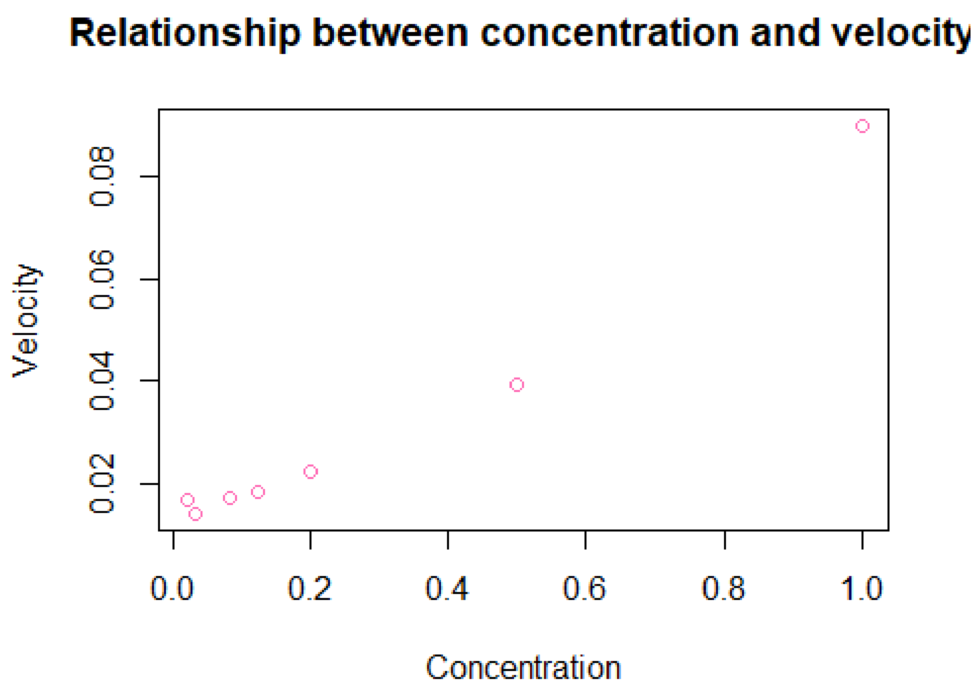
The above model will reduce to a linear model when simplified.

```
library(readxl)
New_data <- read_excel("G:/My Drive/Linear Regression/Datasets/Substrate
concentration.xlsx")
View(New_data)
attach(New_data)
```

```
## The following objects are masked from Dataset:
## 
##     S, V
```

We plot the above in a graph to understand the relationship now that we have taken the inverse

```
S_new=1/S
V_new=1/V

plot(S_new,V_new,col="hotpink",main="Relationship between concentration and velocity",xlab="Concentration",ylab="Velocity")
```



The non-linear model has now become a linear model due to inverse transformation which is visible in the graph. Now we build a linear regression model for the transformed variables and validate the same.

```
Reg_model=lm(V_new~S_new)
Reg_model

## 
## Call:
## lm(formula = V_new ~ S_new)
## 
## Coefficients:
## (Intercept)        S_new  
##     0.00996      0.07551
```

```r
summary(Reg_model)
```

```
##
## Call:
## lm(formula = V_new ~ S_new)
##
## Residuals:
##          1          2          3          4          5          6
7
##  0.0046231 -0.0083435 -0.0027401 -0.0010499  0.0009298  0.0014119
0.0051687
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.009960   0.002519   3.954   0.0108 *
## S_new       0.075507   0.005814  12.988 4.82e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005083 on 5 degrees of freedom
## Multiple R-squared:  0.9712, Adjusted R-squared:  0.9655
## F-statistic: 168.7 on 1 and 5 DF,  p-value: 4.824e-05
```

From the summary of the model, we see the estimated values of concentration[X] is 0.07551 a The linear model built here is

$$Velocity[Y] = B_0 + B_1 Concentration[C] + E$$

where velocity[Y] is the dependent variable, B_0 is the intercept, B_1 is the slope parameter and concentration[X] is the independent variable while E is the error term. The estimated nonlinear model is

$$Y^{hat} = 0.00996 + 0.07551X$$

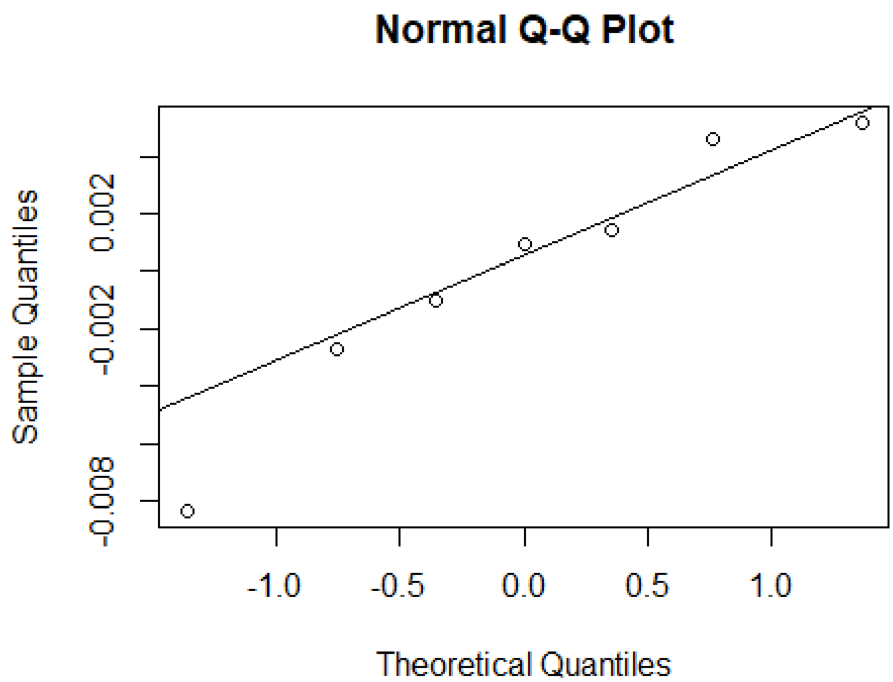
We also see from the summary that, the regression coefficients are significant at 5% significance level. Since the values are significant, we proceed for residual analysis.

We check for the normality by

1. QQ-plot

2. Shapiro test

```r
qqnorm(resid(Reg_model)) # QQ plot-of residual
qqline(resid(Reg_model)) # plots the points
```



If not all, majority of the points fall on the line thus the quartile of normal and residual are almost same, hence it indicates that the residuals follow a normal distribution. However the assumption of normality has to be verified by using a statistical test.(but to know if the deviation of the points lying away from the line, we use the test to further confirm the normality.)

Hypothesis to testing for normality:

H_0: Errors follow normal distribution.

v/s

H_1: Errors do not follow normal distribution.

```
shapiro.test(resid(Reg_model))

##
##  Shapiro-Wilk normality test
##
## data:  resid(Reg_model)
## W = 0.93643, p-value = 0.6068
```

At 0.05 level of significance the p value is 0.6068 which is greater than 0.05, thus we fail to reject null and and say that the residuals follow normal distribution. Hence the assumption of errors.

Hypothesis testing for constant variance:

To test if the errors have constant variance.

H_0: Errors have constant variance

v/s

H_1: Errors have no constant variance

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

bptest(Reg_model)

##
##  studentized Breusch-Pagan test
##
## data:  Reg_model
## BP = 1.2317, df = 1, p-value = 0.2671
```

Since p-value is 0.2671 which is greater than 0.05, thus we fail to reject H_0 and say that the errors have constant variance. Hence the assumption of constant variance is validated.

Thus the fitting of a simple linear regression model that relates velocity and substrate concentration has been done by using inverse transformation. By performing the residual analysis, it is evident that the model is a better fit after the inverse transformation.

Conclusion

A non-linear regression model that relates velocity to concentration using Michaelis-Menten equation was performed and the parameters were estimated. The estimated nonlinear model is,

$$Velocity, V = \frac{73.2614 * S}{3.4372 + S}$$

To check for normality, QQ plot and Shapiro test was performed. The non-linear model did satisfy the normality condition.

Suitable transformation for the model was “INVERSE Transformation”.

$$\frac{1}{V} = \frac{S + K_m}{(V_{max} * S)}$$

$$\frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m}{V_{max} * S)}$$

The linear model built here is

$$Velocity\left[\frac{1}{Y}\right] = B_0 + B_1 Concentration\left[\frac{1}{C}\right] + E$$

The estimated nonlinear model is

$$Y^{hat} = 0.00996 + 0.07551X$$

When the residual analysis was performed, the normality and the constant variance was satisfied. Hence the above linear model is a better fit.