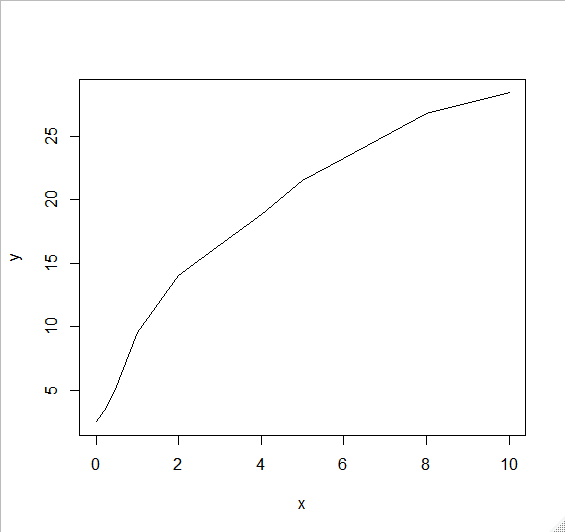
**Exercises Curve Fitting**

Fahmy et al measured the total amount of drug released from a nanoparticle over time. Some of their results are provided in the table below:

|  |  |
| --- | --- |
| Day | Micrograms drug released |
| 0 | 2.5 |
| 0.25 | 3.6 |
| 0.5 | 5.3 |
| 1 | 9.5 |
| 2 | 14.0 |
| 3 | 16.5 |
| 4 | 18.8 |
| 5 | 21.5 |
| 6 | 23.2 |
| 8 | 26.8 |
| 10 | 28.4 |



1. Answer the following questions
2. Create a linear model . Show the design matrix. Calculate the sum of squares and the determination coefficient (correlation coefficient R2)

**Answer:**

fahmy.lm=lm(y~x)

#verify parameters

summary(fahmy.lm)

Residuals:

Min 1Q Median 3Q Max

-4.0014 -2.4242 0.9682 2.3356 2.8160

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.8796 1.2593 4.669 0.00117 \*\*

x 2.6522 0.2614 10.147 3.17e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.762 on 9 degrees of freedom

Multiple R-squared: **0.9196**, Adjusted R-squared: 0.9107

F-statistic: 103 on 1 and 9 DF, p-value: 3.17e-06

cor(y,predict(fahmy.lm))^2

[1] 0.9196089

confint(fahmy.lm)

2.5 % 97.5 %

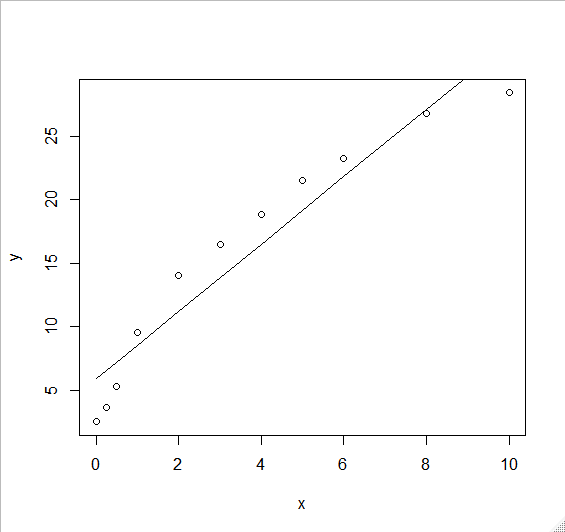
(Intercept) 3.030910 8.728314

x 2.060884 3.243482

#plot model vs prediction

plot(x,y)

points(x,predict(fahmy.lm),type="l")



1. Create a quadratic model . Show the design matrix. Calculate the sum of squares and the determination coefficient (correlation coefficient R2)

**Answer:**

#create cuadratic model

fahmy.cuad=lm(y~1+x+I(x^2))

summary(fahmy.lm)

Residuals:

Min 1Q Median 3Q Max

-0.9423 -0.6265 -0.4192 0.5105 1.7215

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.31044 0.58484 5.660 0.000476 \*\*\*

x 4.99091 0.32814 15.210 3.46e-07 \*\*\*

I(x^2) -0.25343 0.03393 -7.469 7.13e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.038 on 8 degrees of freedom

Multiple R-squared: **0.9899**, Adjusted R-squared: 0.9874

F-statistic: 392.8 on 2 and 8 DF, p-value: 1.033e-08

confint(fahmy.lm)

2.5 % 97.5 %

(Intercept) 1.9618035 4.6590773

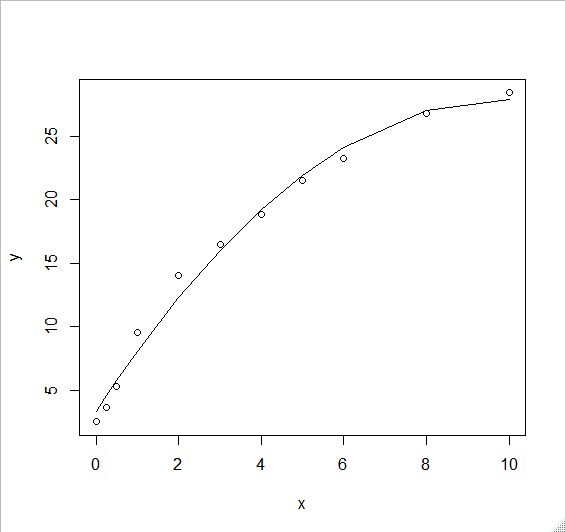
x 4.2342203 5.7475919

I(x^2) -0.3316676 -0.1751886

#plot model vs prediction

plot(x,y)

points(x,predict(fahmy.cuad),type="l")



1. Create a cubic model . Show the design matrix. Calculate the sum of squares and the determination coefficient (correlation coefficient R2)

**Answer:**

#create cubic model

fahmy.cub=lm(y~1+x+I(x^2)+I(x^3))

summary(fahmy.cub)

Residuals:

Min 1Q Median 3Q Max

-0.7741 -0.3458 -0.2206 0.3810 1.0540

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.58218 0.49204 5.248 0.00119 \*\*

x 6.52070 0.58053 11.232 9.9e-06 \*\*\*

I(x^2) -0.68649 0.15201 -4.516 0.00274 \*\*

I(x^3) 0.02958 0.01025 2.887 0.02343 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7494 on 7 degrees of freedom

Multiple R-squared: **0.9954**, Adjusted R-squared: 0.9934

F-statistic: 504.6 on 3 and 7 DF, p-value: 1.537e-08

confint(fahmy.cub)

2.5 % 97.5 %

(Intercept) 1.418690536 3.74565992

x 5.147958340 7.89343775

I(x^2) -1.045923711 -0.32704856

I(x^3) 0.005349176 0.05380357

1. What model would be the best fit? Why?

**Answer:**

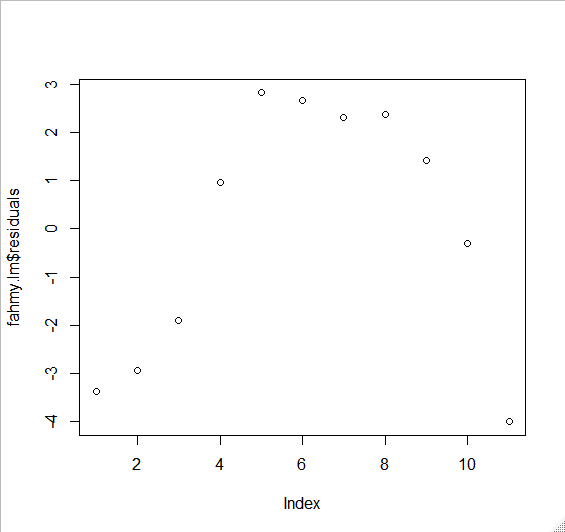
Cubic fit is best because it has the highest R2

1. Definition of Residual. For each data point, its residual is the difference between the actual data point and the value of the equation used to model the data (in this case a polynomial). That is, the residual is . The residuals should be normally distributed. Make a plot of x (Day) vs e. If the residuals are “s” shaped or “u” the fit is not a good one. Are any of the models created a good fit?

**Answer:**

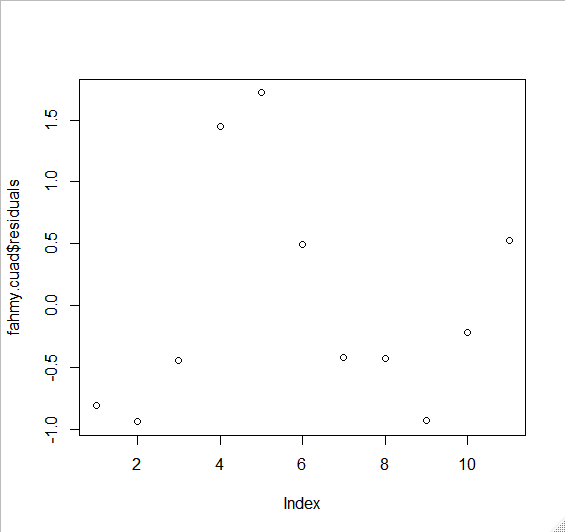
For linear

plot(fahmy.lm$residuals)



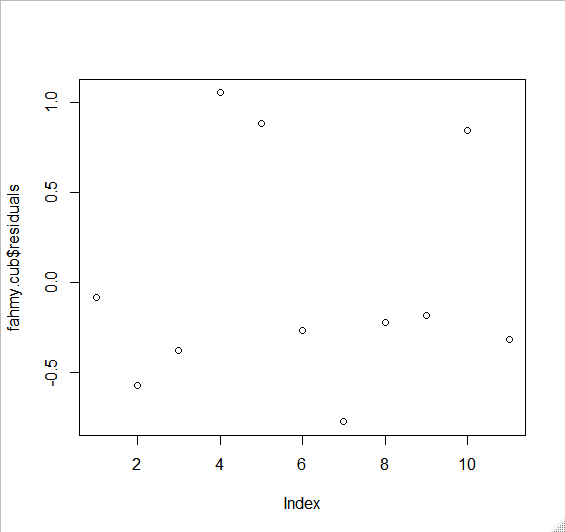
For quadratic

plot(fahmy.cuad$residuals)



For cubic

plot(fahmy.cub$residuals)



Cubic has the most even distribution about the mean error, which is zero.

1. The given data can also be modelled using the following exponential equation:

This equation models a situation where a system goes from a zero value to the maximum value C with a rate constant a. This time, the parameters C and a have a conceptual meaning (maximum drug released and rate constant for the release). Make a fit for the new model

1. Fit the model

**Answer:**

fahmy.nlm<-nls(y~a\*(1-exp(-b\*x)),start=list(a=10,b=1))

summary(fahmy.nlm)

Formula: y ~ a \* (1 - exp(-b \* x))

Parameters:

Estimate Std. Error t value Pr(>|t|)

a 28.68587 1.63849 17.507 2.93e-08 \*\*\*

b 0.30430 0.04233 7.188 5.15e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.504 on 9 degrees of freedom

Number of iterations to convergence: 8

Achieved convergence tolerance: 2.088e-06

confint(fahmy.nlm)

2.5% 97.5%

a 25.3172146 33.6357252

b 0.2122815 0.4248374

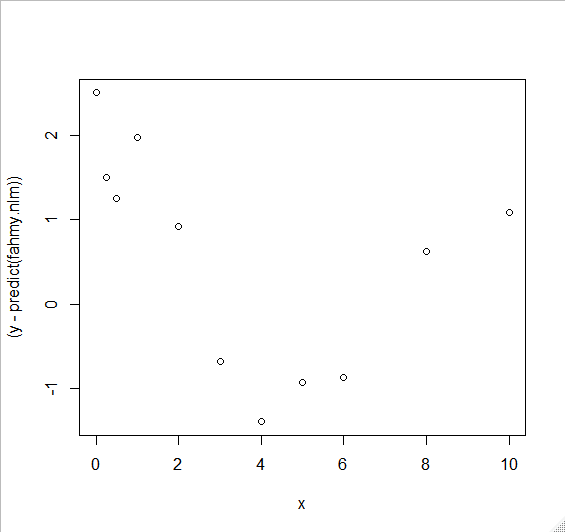
1. Show and plot the residuals. Is this model a good fit?

**Answer:**

cor(y,predict(fahmy.nlm))^2

[1] 0.9892419

plot(x,(y-predict(fahmy.nlm)))



Hard to tell if it is a good fit. The residuals start very badly and then improve.

1. One of the most valuable things a model can do is to predict the behaviour of the system in conditions not originally measured. We held out on you until now, giving you only data through 10 days. However, the published data goes through 21 days:

|  |  |
| --- | --- |
| Day | Micrograms |
| 12 | 28.4 |
| 16 | 28.5 |
| 21 | 29.5 |

Your models should be good at predicting what will happen between days 10 and 21. Without refitting the models, calculate the values of your 4 models at days 12 to 21.

1. Plot the residuals only for days 12 to 21

**Answer:**

new.days <- data.frame(x = c(12, 16, 21))

predict(fahmy.nlm, newdata = new.days)

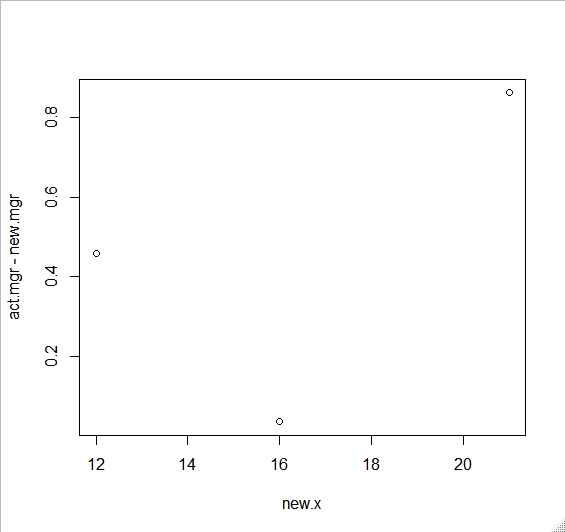
[1] 27.94147 28.46548 28.63774

new.x<-c(12, 16, 21)

new.mgr <- predict(fahmy.nlm, newdata = new.days)

act.mgr <- c(28.4,28.5,29.5)

plot(new.x,act.mgr-new.mgr)



1. Plot the fit for all data points from 0 to 21. Calculate . Plot residuals.

**Answer:**

1. Plot micrograms vs the prediction. A good fit would be a straight line

**Answer:**

1. Answer the questions. Which of the models is the most predictive through 21 days?

**Answer:**