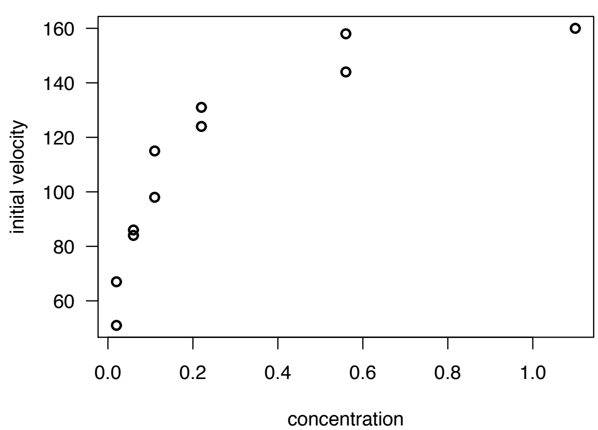
Non-linear Regression

A biochemical experiment



Michaelis-Menten equation

$$V = \frac{V_{\text{max}} \times C}{K + C}$$

V = initial velocity

C = concentration

V_{max} = maximum velocity

K = rate constant

Linearize

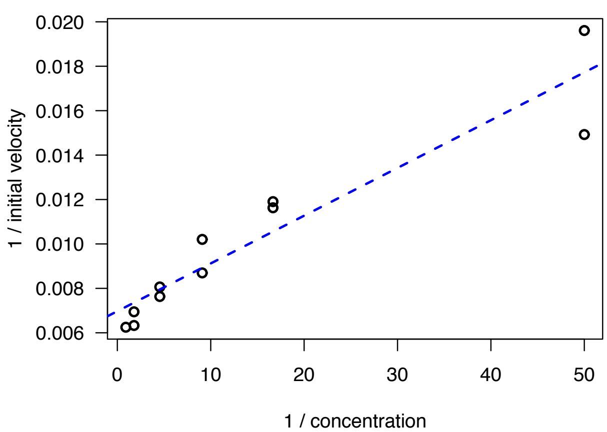
$$V = \frac{V_{\text{max}} \times C}{K + C}$$

$$\Rightarrow \quad \frac{1}{V} = \frac{K + C}{V_{\text{max}} \times C}$$

$$= \frac{K}{V_{\text{max}} \times C} + \frac{1}{V_{\text{max}}}$$

$$\Rightarrow \quad \frac{1}{V} = \left(\frac{1}{V_{\text{max}}}\right) + \left(\frac{K}{V_{\text{max}}}\right) \times \left(\frac{1}{C}\right)$$

Fit the line



Model:

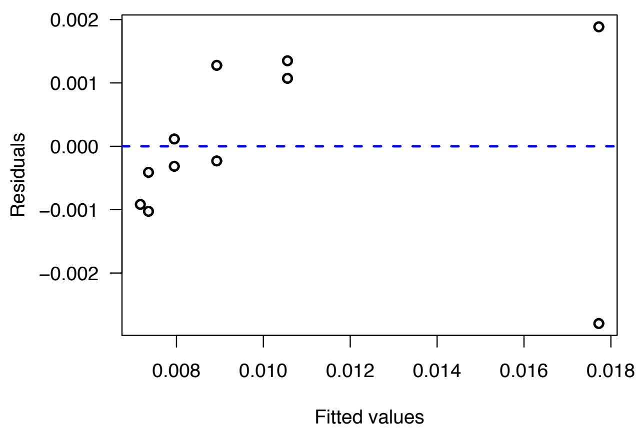
$$\frac{1}{V} = \beta_0 + \beta_1 \left(\frac{1}{C}\right) + \text{ error}$$

```
Intercept    0.00697
Slope        0.00022
```

$\hat{V}_{\text{max}}$ = 1/Intercept = 1/0.00697
= 143

$\hat{K}$ = Slope $\times$ $\hat{V}_{\text{max}}$ = 0.031

Residuals vs fitted values



Which is more reasonable?

$$\frac{1}{V} = \beta_0 + \beta_1 \left(\frac{1}{C}\right) + \text{ error}$$

$$V = \frac{V_{\text{max}} \times C}{K + C} + \text{ error}$$

Nonlinear regression

We imagine that

$$V_i = \frac{V_{\text{max}} \times C_i}{K + C_i} + \epsilon, \quad \epsilon \sim \text{ iid Normal}(0, \sigma^2)$$

We estimate V_{max} and K by the values for which

$$\text{RSS} = \sum_i \left(V_i - \frac{\hat{V}_{\text{max}} \times C_i}{\hat{K} + C_i}\right)^2$$

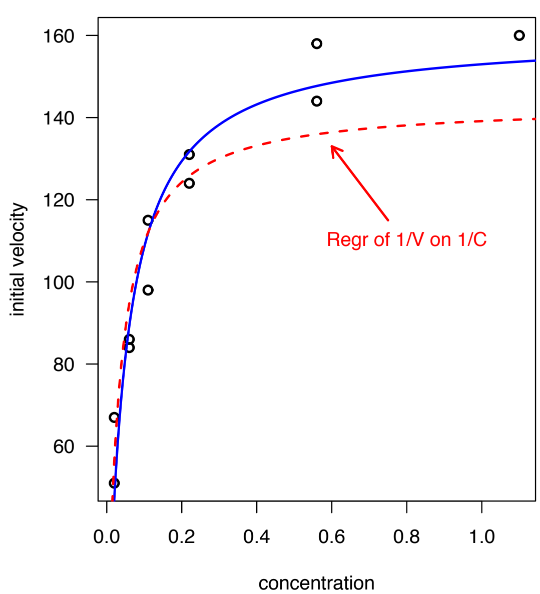
is minimized.

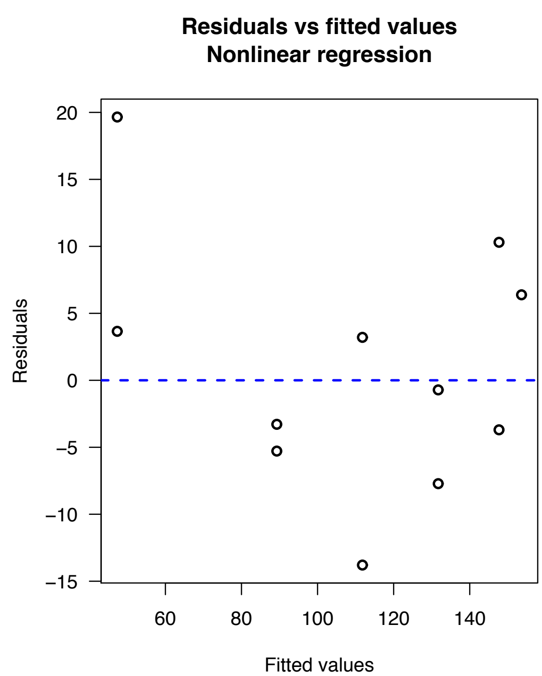
$\longrightarrow$ An iterative method; need "starting values".

Nonlinear regression in R

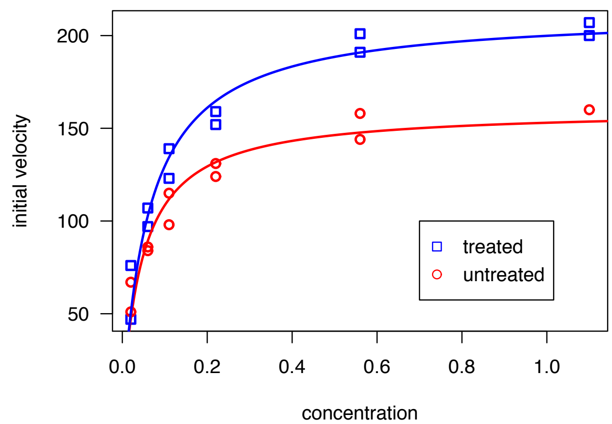
```
> nls.out <- nls(vel ~ (Vm * conc) / (K + conc),
                 data=mydata,
                 start = c(Vm=143, K=0.031))

> summary(nls.out)$param

          Est      SE   t-val     P-val
  Vm   160.28    6.48    24.7   1.4e-09
  K      0.048   0.008    6.1   1.7e-04
```

The fit





A second set of data



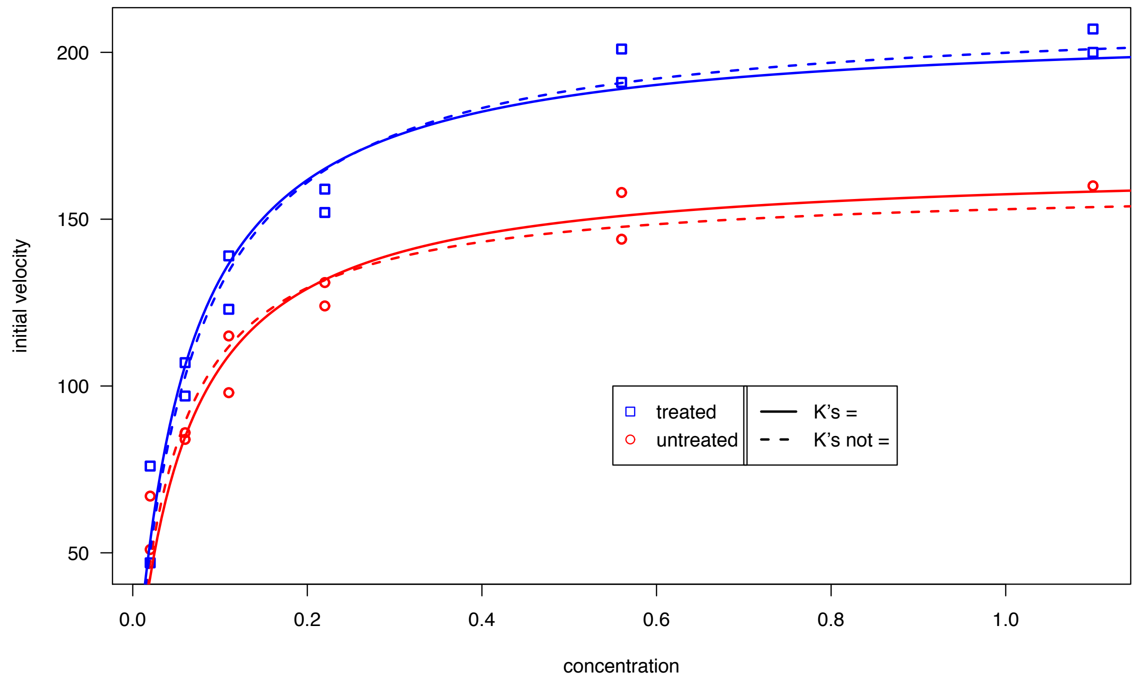
$$V = \frac{(V_{\text{max}} + \Delta V_{\text{max}}\,\mathsf{x}) \times C}{(K + \Delta K\,\mathsf{x}) + C} + \text{ error}$$

x = 0/1 if cells were untreated/treated.

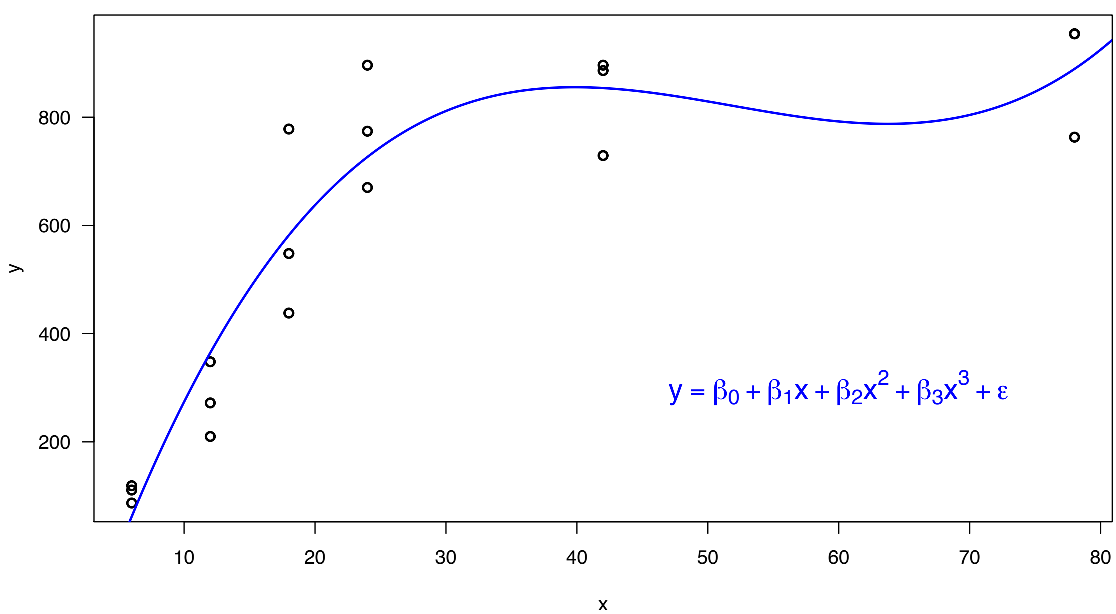
Estimation in R

```
> nls.outC <- nls(vel ~ ((Vm +dV * x) * conc)/
                         (K + dK*x + conc),
                  data=puro,
                  start=c(Vm=160, K=0.048, dV=0, dK=0))

> summary(nls.outC)$param

           Est      SE   t-val     P-val
  Vm    160.28    6.90    23.2   2.0e-15
  K       0.048   0.008    5.8   1.5e-05
  dV     52.40    9.55     5.5   2.7e-05
  dK      0.016   0.011    1.4   1.7e-01
```

K's equal vs not



From last time...



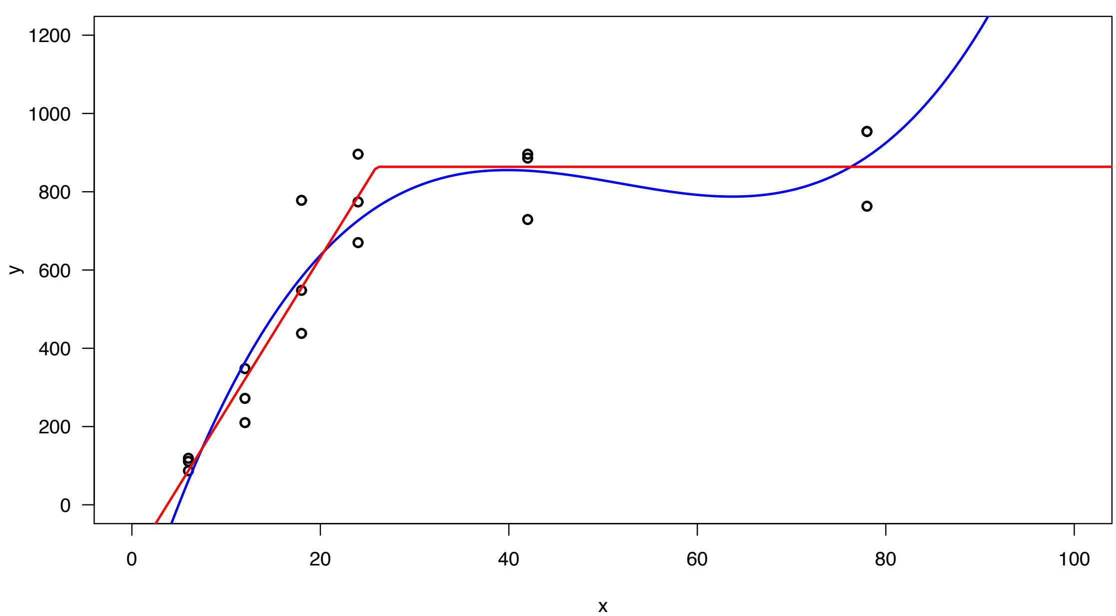
An alternative model

Linear "spline"

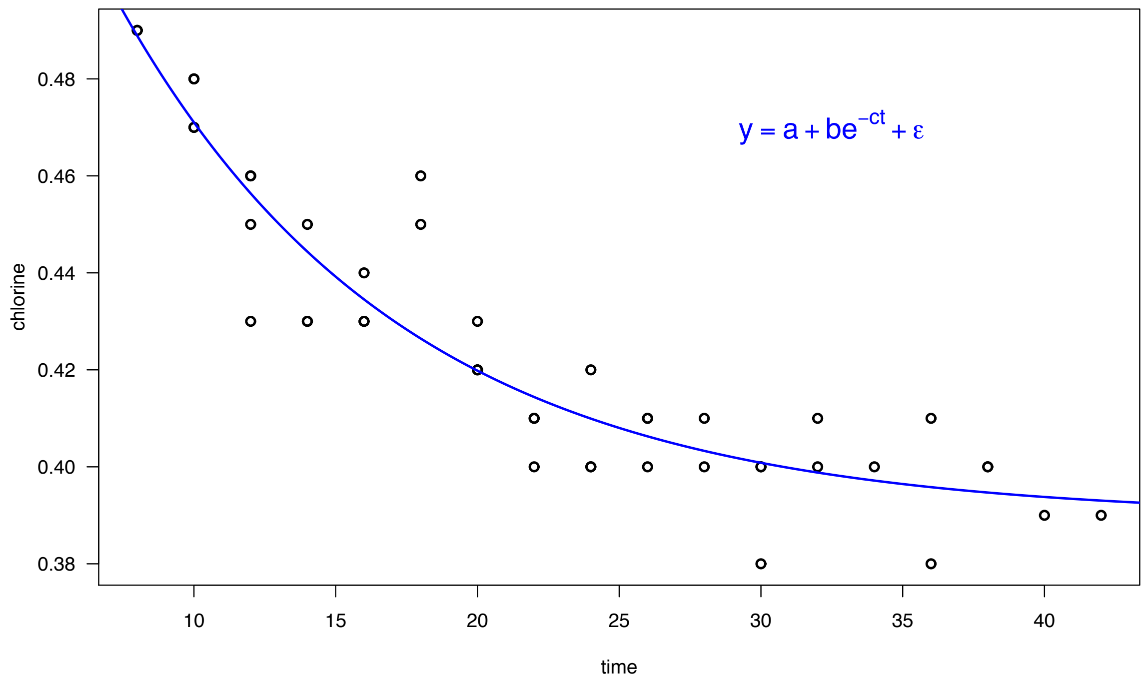
$$y = \begin{cases} \beta_0 + \beta_1 x + \epsilon & \text{if } x \leq x_0 \\ \beta_0 + \beta_1 x_0 + \epsilon & \text{if } x \geq x_0 \end{cases}$$

```
> f <- function(x,b0,b1,x0)
          ifelse(x<x0, b0+b1*x, b0+b1*x0)

> nls.out <- nls(y ~ f(time,b0,b1,x0), data=mydata,
                 start=c(b0=200, b1=100, x0=30))

> summary(nls.out)$param
         Est      SE   t-val      P-val
  b0    -146    71.3    -2.0    5.8e-02
  b1      39     4.3     9.0    2.1e-07
  x0      26     1.8    14.6    2.9e-10
```

Results



One last example...



R code

```
> nls.out <- nls(chlorine ~ a + b*exp(-c*time), data=cl,
                start=c(a=0.05, b=0.49, c=0.1))

> summary(nls.out)$param

         Est       SE    t-val    P-val
  a    0.390    0.006     66.7  1.9e-43
  b    0.219    0.031      7.0  1.7e-08
  c    0.099    0.018      5.5  2.4e-06
```