Zurich University
of Applied Sciences

School of
Engineering

IDP Institute of
Data Analysis and
Process Design

# WBL Statistik 2024 — Nonlinear Regression

## A Powerful Tool With Considerable Complexity

## Half-Day 1: Estimation and Standard Inference

Andreas Ruckstuhl
Institut für Datenanalyse und Prozessdesign
Zürcher Hochschule für Angewandte Wissenschaften

# Outline:

**Half-Day 1** Estimation and Standard Inference
- The Nonlinear Regression Model
- Iterative Estimation - Model Fitting
- Inference Based on Linear Approximations

**Half-Day 2** Improved Inference and Visualisation
- Likelihood Based Inference
- Profile t Plot and Profile Traces
- Parameter Transformations

**Half-Day 3** Bootstrap, Prediction and Calibration
- Bootstrap
- Prediction
- Calibration

Outlook

# Your Lecturer

*Name:* **Andreas Ruckstuhl**
*Civil Status:* Married, 2 children
*Education:* Dr. sc. math. ETH

*Position:* Professor of Statistical Data Analysis, ZHAW Winterthur
Lecturer, WBL Applied Statistics, ETH Zürich

*Expierence:*

| | |
|---|---|
| 1987 – 1991 | Statistical Consulting and Teaching Assistant, Seminar für Statistik, ETHZ |
| 1991 – 1996 | PhD, Teaching Assistant, Lecturer in NDK/WBL, ETHZ |
| 1996 | Post-Doc, Texas A&M University, College Station, TX, USA |
| 1996 – 1999 | Lecturer, ANU, Canberra, Australia |
| Since 1999 | Institute of Data Analysis and Process Design, ZHAW |

# 1.1 The Nonlinear Regression Model

## The regression model

$$Y_i = h\left\langle x_i^{(1)}, \ldots, x_i^{(m)};\ \theta_1, \theta_2, \ldots, \theta_p \right\rangle + E_i \quad \text{with } E_i \text{ indep. } \mathcal{N}\left\langle 0, \sigma^2 \right\rangle$$

In case of the linear regression model

$$h\left\langle x_i^{(1)}, \ldots, x_i^{(m)};\ \theta_1, \theta_2, \ldots, \theta_p \right\rangle = \theta_1 \cdot 1 + \theta_2 x_i^{(2)} + \ldots + \theta_p x_i^{(p)} \quad (\text{i.e., } m = p)$$

**Examples of nonlinear regression function:**

$$h\langle x_i; \underline{\theta} \rangle = \frac{\theta_1 x_i^{\theta_3}}{\theta_2 + x_i^{\theta_3}}$$

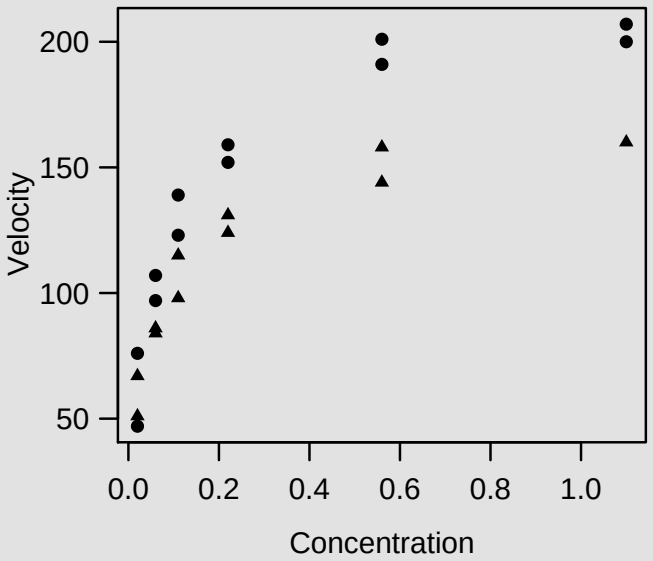
$$h\langle x; \underline{\theta} \rangle = \theta_1 \exp\left\langle \frac{\theta_2}{x_i} \right\rangle$$

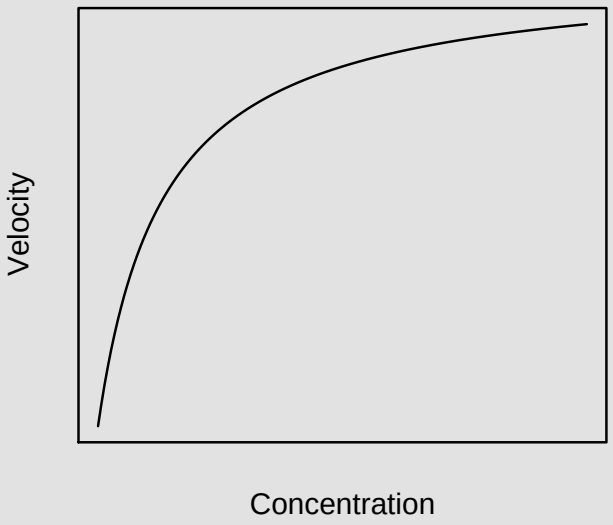
$$h\langle x; \underline{\theta} \rangle = \exp\left\langle \theta_1 \left(x_i^{(1)}\right)^{\theta_3} \exp\left\langle -\frac{\theta_2}{x_i^{(2)}} \right\rangle \right\rangle$$

# Example: Puromycin

The Michaelis-Menten model for enzyme kinetics relates the initial "velocity" of an enzymatic reaction to the substrate concentration

- ● teated with Puromycin
- △ not treated



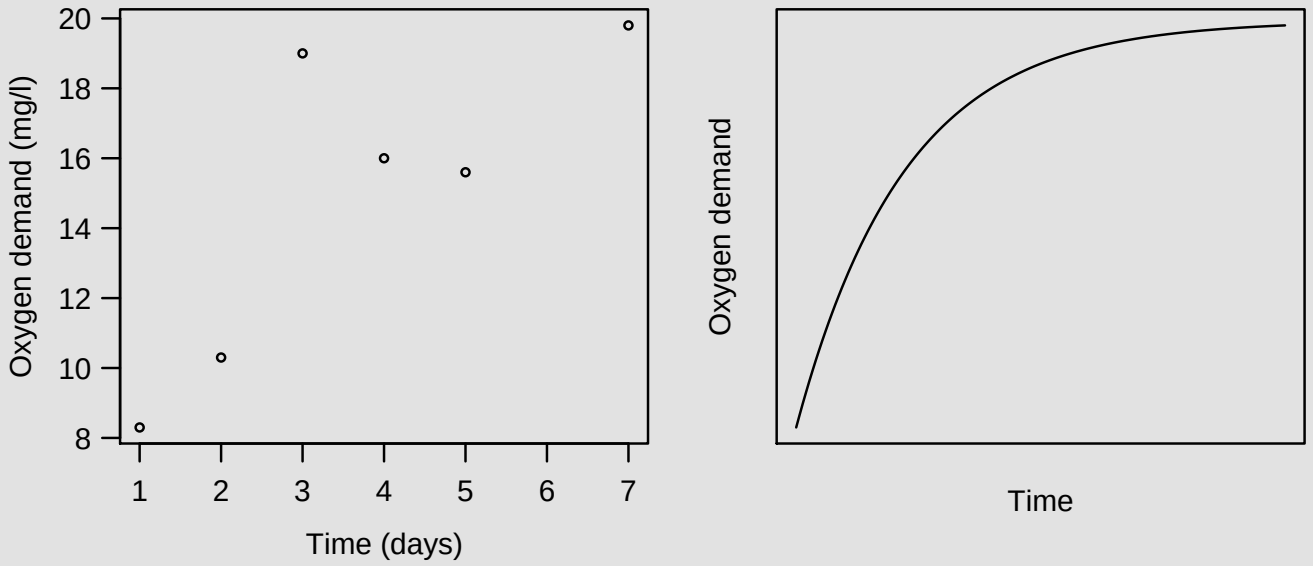




$$Y_i = \frac{\theta_1 \cdot x_i}{\theta_2 + x_i} + E_i \quad \text{with} \quad E_i \; i.i.d. \sim \mathcal{N}\left\langle 0, \sigma^2 \right\rangle \qquad \text{(Michaelis-Menten model)}$$

$x$ substrate concentration [ppm]
$Y$ initial "velocity" [((number of counts)/min)/min]

# Example: Biochemical Oxygen Demand (BOD)

## Biochemical oxygen demand of stream water

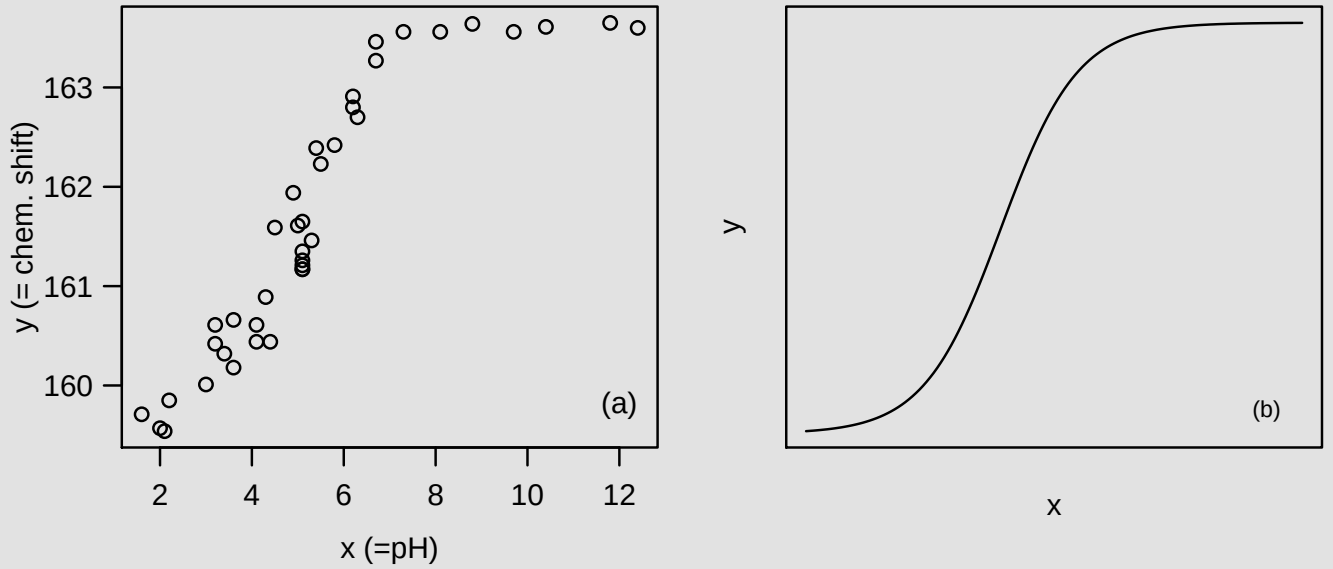



$$Y_i = \theta_1 \cdot \left(1 - e^{\theta_2 \cdot x_i}\right) + E_i \qquad \text{mit} \;\; E_i \; i.i.d. \sim \mathcal{N}\left\langle 0, \sigma^2 \right\rangle ,$$

where $Y$ is the biochemical oxygen demand (BOD) $[mg/\ell]$ and $x$ the incubation time [days]

# Example: Cellulose Membrane

Ratio of protonated to deprotonated carboxyl groups within the pore of celluose membrane versus pH value $x$ of the bulk solution




Theoretically, this relation is described by the Henderson-Hasselbalch equation,

$$Y_i = \frac{\theta_1 + \theta_2 \cdot 10^{\theta_3 + \theta_4 x_i}}{1 + 10^{\theta_3 + \theta_4 x_i}} + E_i \qquad i = 1, \ldots, n, \quad \text{with } E_i \; i.i.d. \sim \mathcal{N}\left\langle 0, \sigma^2 \right\rangle .$$

# Transformably Linear Models

Example: $h\langle x, \underline{\theta}\rangle = \theta_1 \cdot \exp\left\langle \frac{\theta_2}{x} \right\rangle$

Applying the log-transformation, we obtain

$$\begin{aligned}\log\langle h\langle x, \underline{\theta}\rangle\rangle &= \log\left\langle \theta_1 \cdot \exp\left\langle \frac{\theta_2}{x}\right\rangle\right\rangle \\ &= \log\langle\theta_1\rangle + \log\left\langle \exp\left\langle \frac{\theta_2}{x}\right\rangle\right\rangle \\ &= \log\langle\theta_1\rangle + \theta_2 \cdot \frac{1}{x}\end{aligned}$$

Hence

$$\log\left\langle h\left\langle x, \underline{\beta}\right\rangle\right\rangle = \beta_1 + \beta_2\widetilde{x}$$

The "complete" transformably linear regression model is

$$\log\langle Y_i\rangle = \beta_1 + \beta_2\widetilde{x}_i + E_i\,, \quad E_i \; i.i.d. \sim \mathcal{N}\left\langle 0, \sigma^2\right\rangle$$

The error term is additive

In the original representation, the regression model transforms to

$$\begin{aligned}Y_i &= \exp\left\langle \beta_1 + \beta_2\widetilde{x}_i + E_i\right\rangle \\ &= \theta_1 \cdot \exp\left\langle \frac{\theta_2}{x}\right\rangle \cdot \widetilde{E}_i\end{aligned}$$

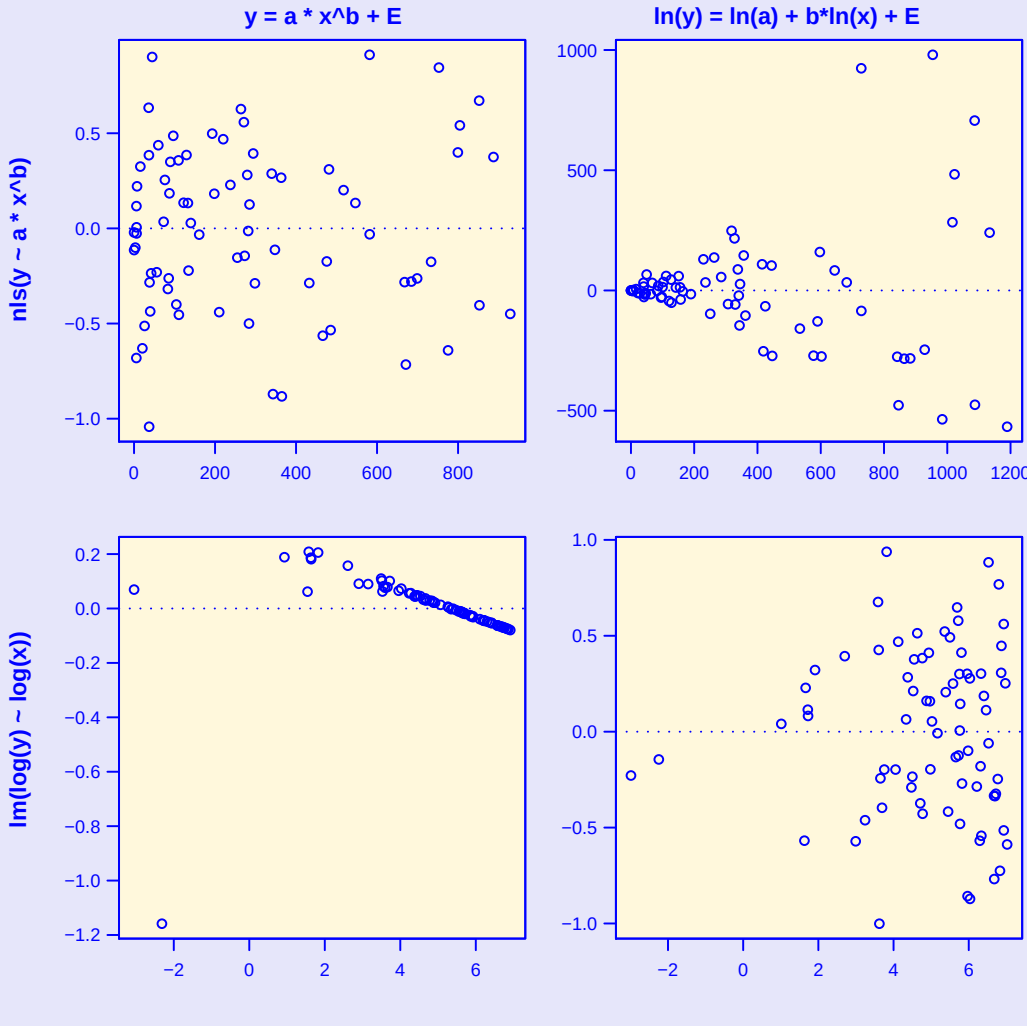
i.e., $\widetilde{E}_i$ is log-normally distributed and the error is multiplicative.

## Conclusion:

Transform to a linear regression model only if required by the error structure.

☞ Check assumptions on error term by residual analysis.

If there is a deterministic model $y = \theta_1 \cdot x^{\theta_2}$, the random component may be either additiv or multiplicativ. – The Tukey-Anscombe plot of the fitted model will show clearly which model is more adequate for the data.



# A selection of transformably linear models

$$h\langle x, \underline{\theta}\rangle = 1/(\theta_1 + \theta_2 \exp\langle -x\rangle) \quad \longleftrightarrow \quad 1/h\langle x, \underline{\theta}\rangle = \theta_1 + \theta_2 \exp\langle -x\rangle$$

$$h\langle x, \underline{\theta}\rangle = \theta_1 x/(\theta_2 + x) \quad \longleftrightarrow \quad 1/h\langle x, \underline{\theta}\rangle = 1/\theta_1 + \theta_2/\theta_1 \tfrac{1}{x}$$

$$h\langle x, \underline{\theta}\rangle = \theta_1 x^{\theta_2} \quad \longleftrightarrow \quad \ln\langle h\langle x, \underline{\theta}\rangle\rangle = \ln\langle\theta_1\rangle + \theta_2 \ln\langle x\rangle$$

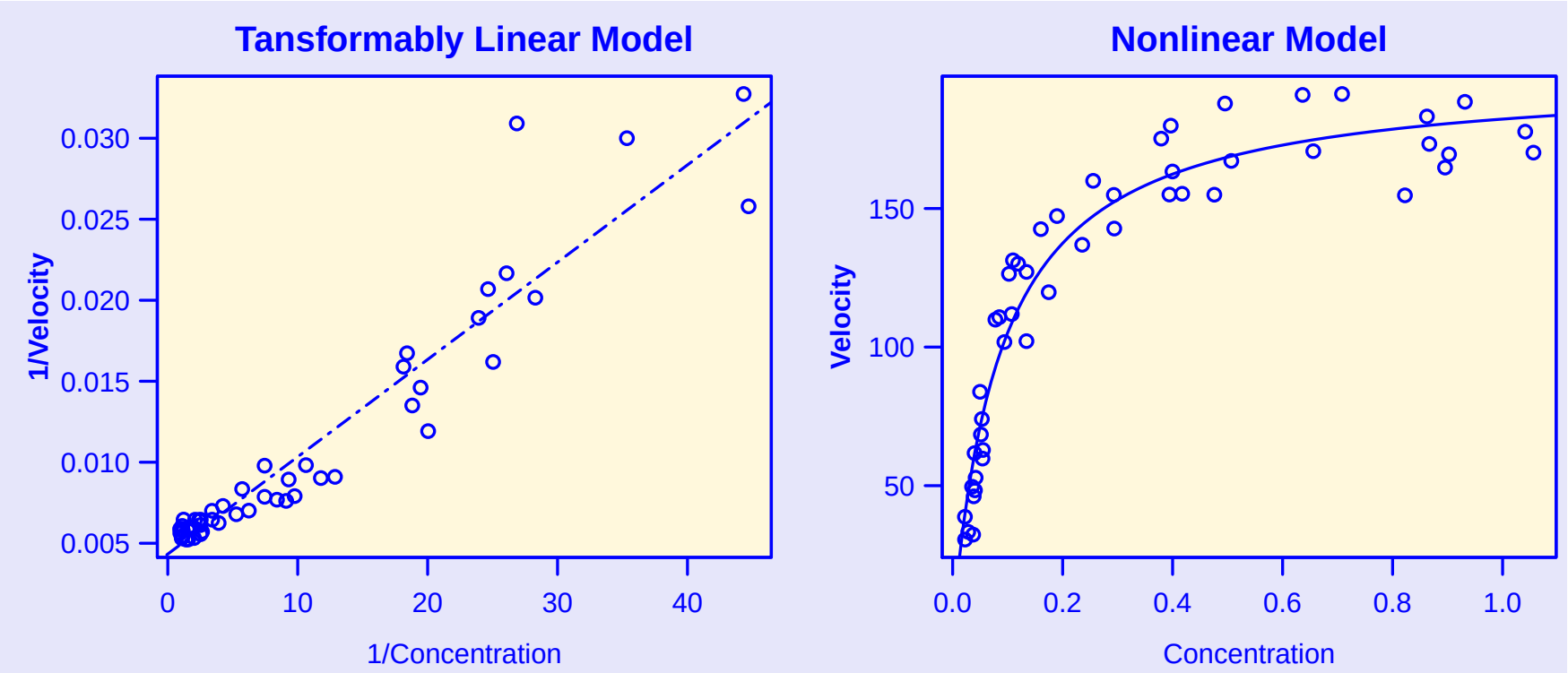
$$h\langle x, \underline{\theta}\rangle = \theta_1 \exp\langle\theta_2 g\langle x\rangle\rangle \quad \longleftrightarrow \quad \ln\langle h\langle x, \underline{\theta}\rangle\rangle = \ln\langle\theta_1\rangle + \theta_2 g\langle x\rangle$$

$$h\langle x, \underline{\theta}\rangle = \exp\left\langle -\theta_1 x^{(1)} \exp\left\langle -\theta_2/x^{(2)}\right\rangle\right\rangle \quad \longleftrightarrow \quad \ln\langle\ln\langle h\langle x, \underline{\theta}\rangle\rangle\rangle = \ln\langle -\theta_1\rangle + \ln\left\langle x^{(1)}\right\rangle - \theta_2/x^{(2)}$$

$$h\langle x, \underline{\theta}\rangle = \theta_1 \left(x^{(1)}\right)^{\theta_2} \left(x^{(2)}\right)^{\theta_3} \quad \longleftrightarrow \quad \ln\langle h\langle x, \underline{\theta}\rangle\rangle = \ln\langle\theta_1\rangle + \theta_2 \ln\left\langle x^{(1)}\right\rangle + \theta_3 \ln\left\langle x^{(2)}\right\rangle$$

Note that
it is not sufficient to examine the fit in a scatter plot of response versus exploratory variable,

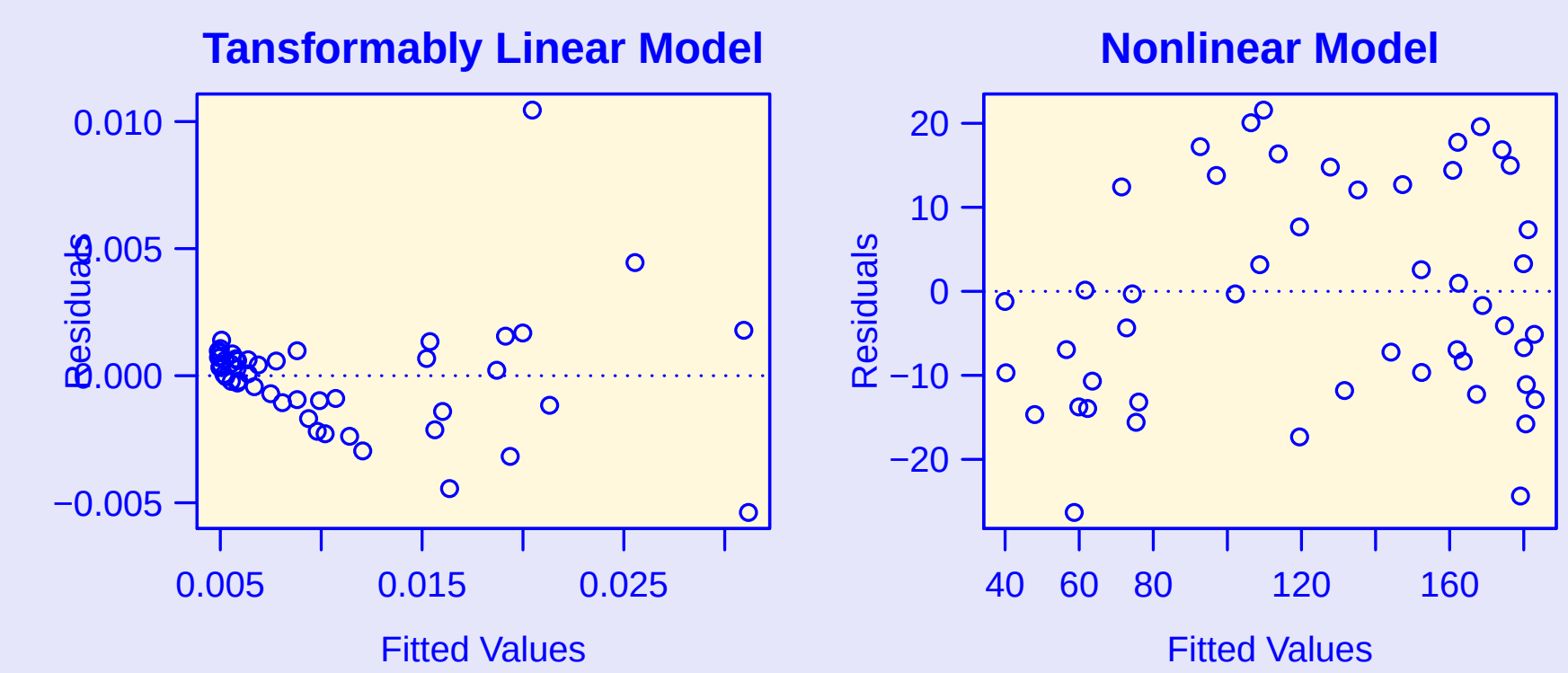
since the nonlinearity of the fitted curve prevents an assessment of the specified error structure.




**Which model fits the data better?**

Use Tukey-Anscombe plot to check whether

- the expectation of the error is zero
- the variability of the error is constant



# 1.2 Model Fitting Using an Iterative Algorithm

## The method of least squares:

Find the minimum of

$$S\langle\underline{\theta}\rangle = \sum_{i=1}^{n} (y_i - \eta_i\langle\underline{\theta}\rangle)^2 \qquad \text{mit } \eta_i\langle\underline{\theta}\rangle = h\langle\underline{\theta}, \underline{x}_i\rangle \ .$$

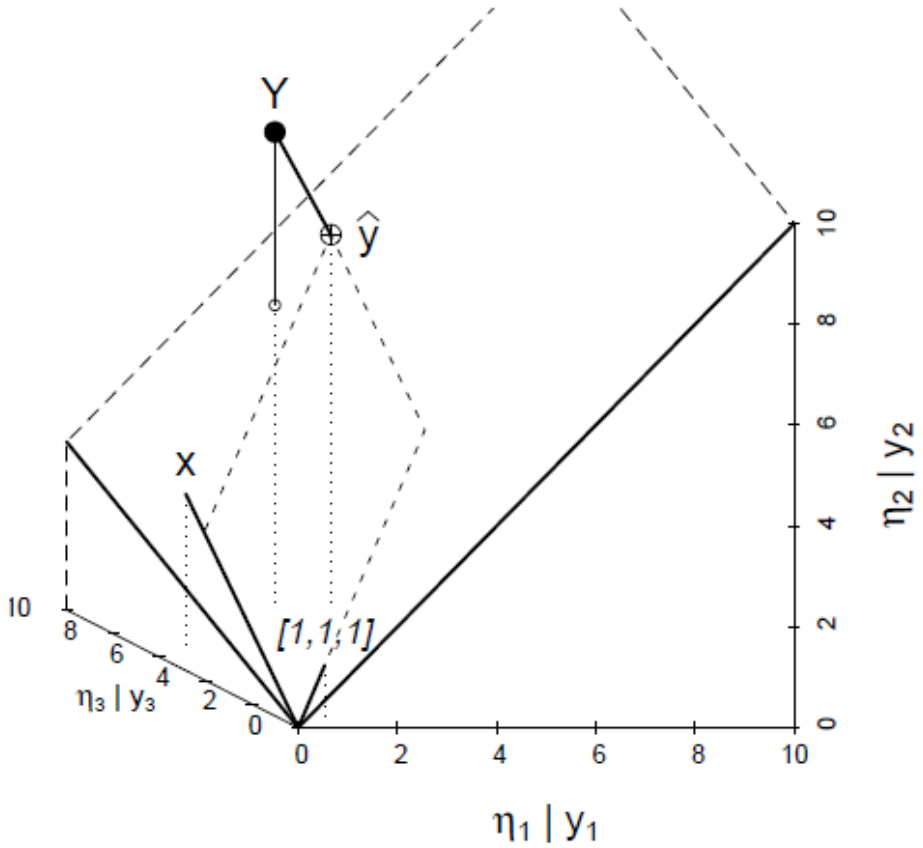
Key steps for minimising:

- **approximate** the surface $\underline{\eta}\langle\underline{\theta}\rangle$ at a temporarily best value $\underline{\theta}^{(\ell)}$ **by a tangent plane** where $\underline{\eta}\left\langle\underline{\theta}^{(\ell)}\right\rangle$ is the point of contact.
- **search the point on the plane**, which is closest to $\underline{Y}$ (that is a linear regression fitting problem).
- The new point lies on the plain but not on the surface. However, it defines a parameter vector $\underline{\theta}^{(\ell+1)}$ which will be used in the next iteration step.

# Geometrical Illustration

Geometric illustration of the least squares fit in the case of *linear* regression with three observations $(y_i, x_i)$, $i = 1, 2, 3$.

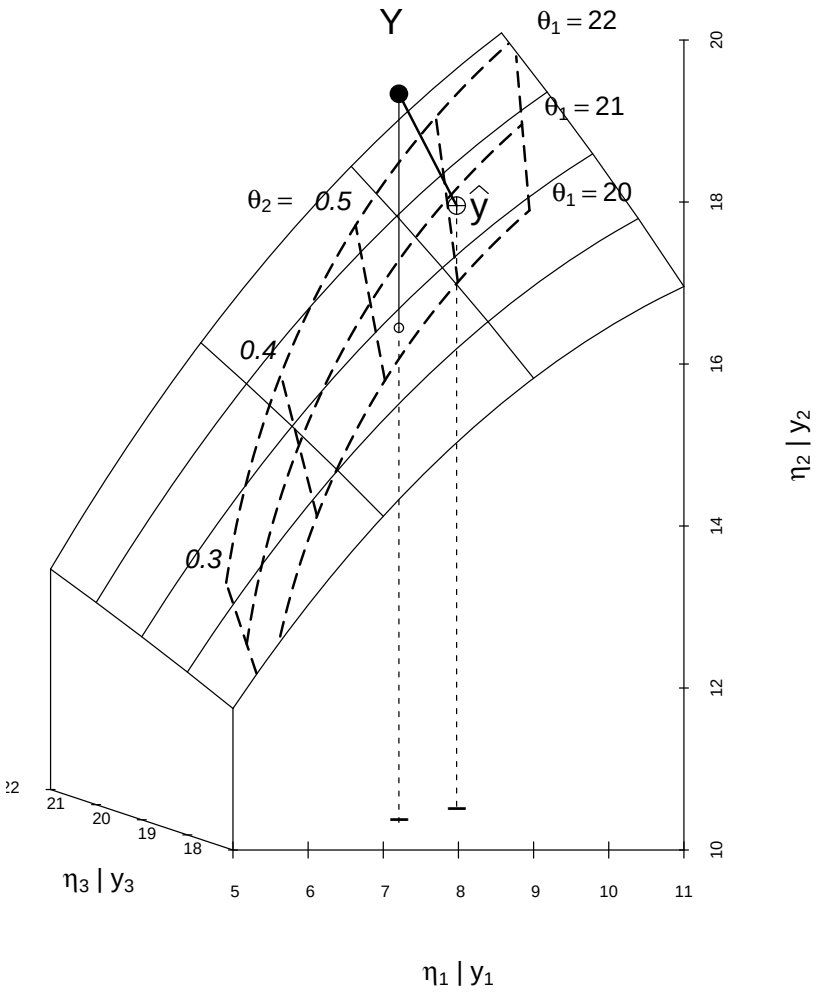
We can calculate the three model values $\eta_i\langle\underline{\beta}\rangle = \beta_0 + \beta_1 x_i$ and represent the corresponding vector $\underline{\eta}\langle\underline{\beta}\rangle = \beta_0\underline{1} + \beta_1\underline{x}$ as a point depending on given parameters $\beta_0$ and $\beta_1$.



# Geometrical Illustration

Geometric illustration of the least squares fit in the case of *nonlinear* regression with three observations $(y_i, x_i)$, $i = 1, 2, 3$.

We can calculate the three model values $\eta_i\langle\underline{\beta}\rangle = \theta_1 \exp\langle 1 - \theta_2 x_i\rangle$ and represent the corresponding vector $\underline{\eta}\langle\underline{\beta}\rangle = [\eta_1\langle\underline{\beta}\rangle, \eta_2\langle\underline{\beta}\rangle, \eta_3\langle\underline{\beta}\rangle]^T$ as a point depending on given parameters $\theta_1$ and $\theta_2$.



# Algebraically formulated

☞ Gauss-Newton Algorithm

1. Linear approximation of $\underline{\eta}_i\left\langle\underline{\theta}\right\rangle$ at $\underline{\theta}^{(\ell)}$:

$$\underline{\eta}_i\left\langle\underline{\theta}\right\rangle \approx \underline{\eta}_i\left\langle\underline{\theta}^{(\ell)}\right\rangle + \boldsymbol{A}^{(\ell)}\left(\underline{\theta}-\underline{\theta}^{(\ell)}\right),$$

   where $\boldsymbol{A}^{(\ell)}$ defines the tangent plane in the $\ell$-th iteration step, i.e., $\boldsymbol{A}^{(\ell)} = \boldsymbol{A}\left\langle\underline{\theta}^{(\ell)}\right\rangle$ is the derivative matrix of $\underline{\eta}\left\langle\underline{\theta}\right\rangle$ at $\underline{\theta}^{(\ell)}$.

2. (Locally) linear regression model

$$\underline{\widetilde{Y}}^{(\ell)} \approx \boldsymbol{A}^{(\ell)}\,\underline{\beta}^{(\ell)} + \underline{E}$$

   where $\underline{\widetilde{Y}}^{(\ell)} = \underline{Y} - \underline{\eta}\left\langle\underline{\theta}^{(\ell)}\right\rangle$ and $\underline{\beta}^{(\ell)} = \underline{\theta} - \underline{\theta}^{(\ell)}$

3. Least-squares estimation for $\beta^{(\ell)} \to \widehat{\beta}^{(\ell)}$.

   Set $\underline{\theta}^{(\ell+1)} = \underline{\theta}^{(\ell)} + \underline{\widehat{\beta}}^{(\ell)}$.

4. Repeat steps 1 to 3 until the procedure converges.

   result $\underline{\widehat{\theta}} = \underline{\theta}^{(\ell+1)}$

# Starting Values

- interpret the behaviour of the regression function in terms of the parameter analytically or graphically
- transform the regression function to obtain simpler, preferably linear, behaviour
- use your knowledge from previous or similar experiments

Example Puromycin (2) - using transformation

$$y \approx h\langle x, \underline{\theta}\rangle = \frac{\theta_1 \cdot x_i}{\theta_2 + x_i}$$

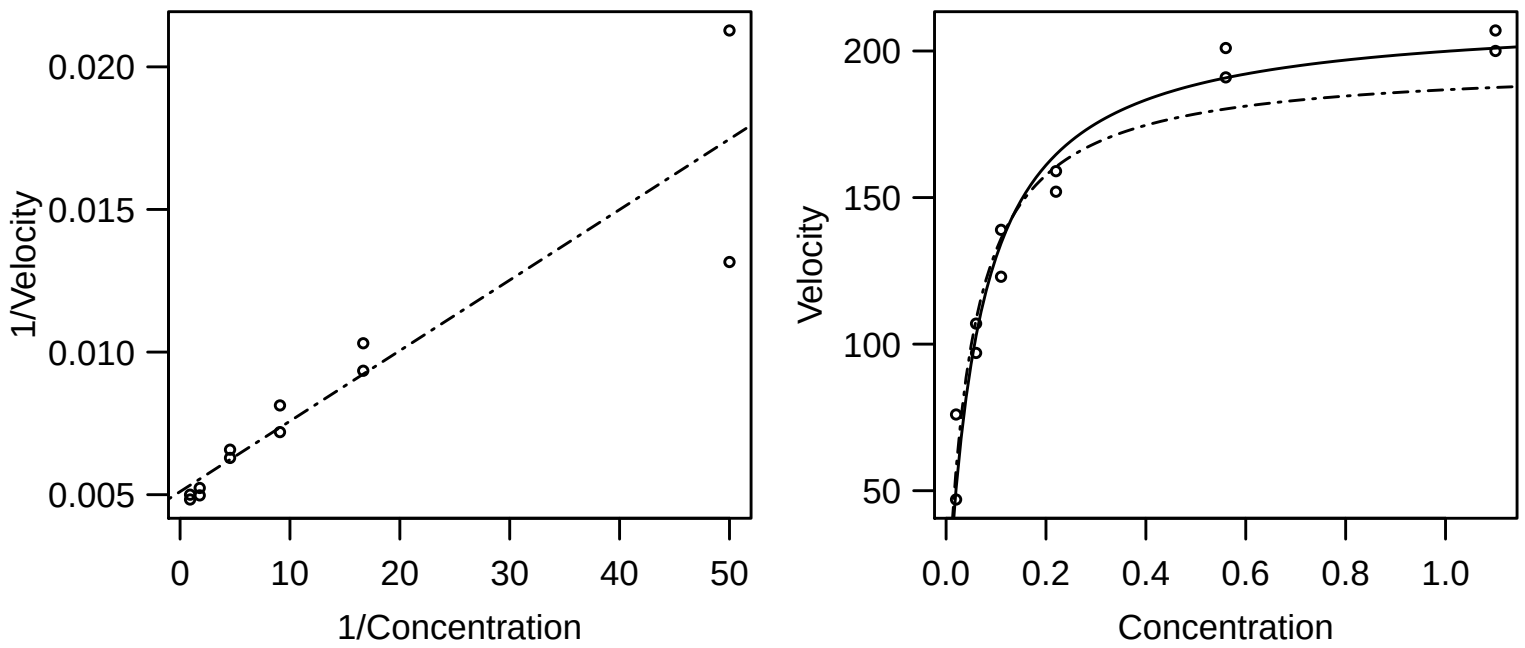
transform to linearity $$\widetilde{y} = \frac{1}{y} \approx \frac{1}{h\langle x, \underline{\theta}\rangle} = \frac{\theta_2}{\theta_1} \cdot \frac{1}{x} + \frac{1}{\theta_1}$$

that is $$\widetilde{y} \approx \beta_1 \widetilde{x} + \beta_0$$

linear regression (robustly estimated) ☞ $$\widehat{\underline{\beta}} = (0.005,\ 0.00025)^T$$

starting values: $$\widehat{\theta}_1^0 = \frac{1}{\widehat{\beta}_0} \approx 196 \qquad \widehat{\theta}_2^0 = \frac{\widehat{\beta}_1}{\widehat{\beta}_0} \approx 0.048$$

# Example Puromycin (3)

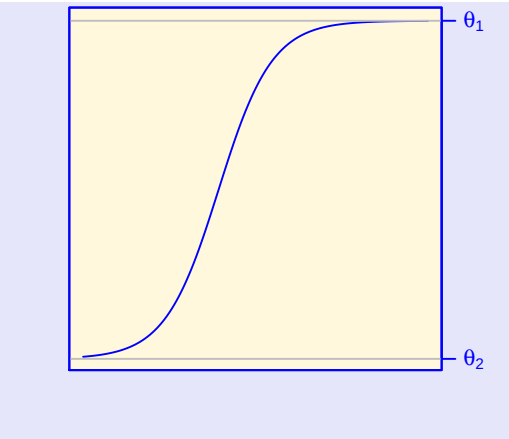



Left: Regression line used for determining the starting values $\theta_1$ and $\theta_2$.

Right: Regression function $h\langle x; \underline{\theta}\rangle$ based on the starting values $\underline{\theta} = \underline{\theta}^{(0)}$ (– – – – –) and based on the least-squares estimation $\underline{\theta} = \widehat{\underline{\theta}}$ (———), respectively.

# Example: Cellulose membrane (2) - starting values

$$h\langle x; \underline{\theta}\rangle = \frac{\theta_1 + \theta_2 \cdot 10^{\theta_3+\theta_4 x}}{1 + 10^{\theta_3+\theta_4 x}} \qquad \text{mit } \theta_4 < 0$$




We know: $h\langle x; \underline{\theta}\rangle \longrightarrow \theta_1$ for $x \to \infty$

$h\langle x; \underline{\theta}\rangle \longrightarrow \theta_2$ for $x \to -\infty$

From data,
we obtain $\theta_1^{(0)} = 163.7$ und $\theta_2^{(0)} = 159.5$

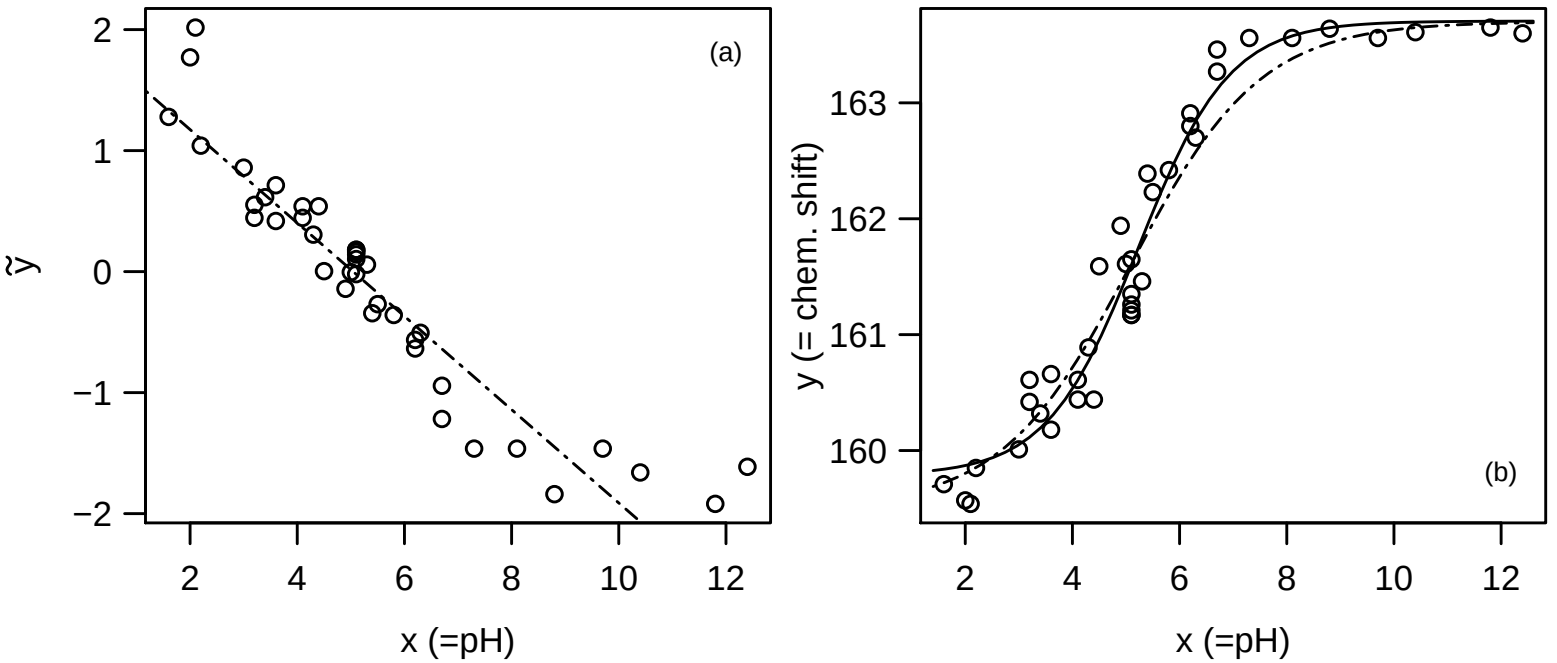
Let $\widetilde{y}_i = \log_{10}\left\langle \dfrac{\theta_1^{(0)} - y_i}{y_i - \theta_2^{(0)}} \right\rangle$,

hence $\widetilde{y}_i = \theta_3 + \theta_4\, x_i$.

Simple linear regression results in starting values for both $\theta_3$ and $\theta_4$

$$\theta_3^{(0)} = 1.83 \qquad \text{and} \qquad \theta_4^{(0)} = -0.36\,.$$

# Example: Cellulose membrane (3)




(a) Regression line used for determining the starting values $\theta_3$ and $\theta_4$ .

(b) Regression function $h\langle x; \underline{\theta}\rangle$ based on the starting values $\underline{\theta} = \underline{\theta}^{(0)}$ (– – – – –) and based on the least-squares estimation $\underline{\theta} = \widehat{\underline{\theta}}$ (———), respectively.

# Self-Starter Function

- For repeated use of the same nonlinear regression model
  ☞ use an automated way of providing starting values.
- Basically, collect all the manual steps which are necessary to obtain the initial values for a nonlinear regression model into a function.

- Self-starter functions are specific for a given mean function and calculates starting values for a given dataset.
- Let `SSmicmen()` (c.f. next slide) be the self-starter function for the Michaelis-Menton model. You can then run the fitting process as
  `nls(rate ~ SSmicmen(conc, Vm, K), data=D.minor)`
- How to write your own self-starter functions
  see help or, e.g., Ritz & Streibig (2008), Sec 3.2
- With the standard installation of R,
  the following self-starter functions are implemented:

# Self-Starter Functions in the Standard Installation

| Model | Mean Function | Name of Self-Starter Function |
|---|---|---|
| Biexponential | $A1 \cdot e^{-x \cdot e^{lrc1}} + A2 \cdot e^{-x \cdot e^{lrc2}}$ | `SSbiexp(x, A1, lrc1, A2, lrc2)` |
| Asymptotic regression | $Asym + (R0 - Asym) \cdot e^{-x \cdot e^{lrc}}$ | `SSasymp(x, Asym, R0, lrc)` |
| Asymptotic regression with offset c0 | $Asym \cdot (1 - e^{-(x-c0) \cdot e^{lrc}})$ | `SSasympOff(x, Asym, lrc, c0)` |
| Asymptotic regression (c0 = 0) | $Asym \cdot (1 - e^{-x \cdot e^{lrc}})$ | `SSasympOrig(x, Asym, lrc)` |
| First-order compartment | $x1 \cdot \frac{e^{lKe+lKa-lCl}}{e^{lKa}-e^{lKe}} \cdot (e^{-x2 \cdot e^{lKe}} - e^{-x2 \cdot e^{lKa}})$ | `SSfol(x1, x2, lKe, lKa, lCl)` |
| Gompertz | $Asym \cdot e^{-b2 \cdot b3^x}$ | `SSgompertz(x, Asym, b2, b3)` |
| Logistic | $A + \frac{B-A}{1+e^{(xmid-x)/scal}}$ | `SSfpl(x, A, B, xmid, scal)` |
| Logistic (A = 0) | $\frac{Asym}{1+e^{(xmid-x)/scal}}$ | `SSlogis(x, Asym, xmid, scal)` |
| Michaelis-Menten | $Vm \cdot \frac{x}{K+x}$ | `SSmicmen(x, Vm, K)` |
| Weibull | $Asym - Drop \cdot e^{-e^{lrc} \cdot xpwr}$ | `SSweibull(x, Asym, Drop, lrc, pwr)` |

# 3 Inference Based on Linear Approximations

As a look on the summary output of the Example "Cellulose Membrane" shows it look very similar to the summary output of a fitted linear regression model:

```
> Mem.fit <- nls(delta ~ (T1 + T2*10^(T3+T4*pH))/(10^(T3+T4*pH)+1),
                 D.membran, start=list(T1=163.7, T2=159.5, T3=1.83, T4=-0.36))
> summary(Mem.fit)
Formula: delta ~ (T1 + T2 * 10^(T3 + T4 * pH))/(10^(T3 + T4 * pH) + 1)
 Parameters:
      Value   Std. Error   t value   Pr(>|t|)
θ1  163.706       0.1262   1297.26    < 2e-16   ***
θ2  159.785       0.1594   1002.19    < 2e-16   ***
θ3    2.675       0.3813      7.02   3.65e-08   ***
θ4   -0.512       0.0703     -7.28   1.66e-08   ***

 Residual standard error: 0.293137 on 35 degrees of freedom
 Number of iterations to convergence: 7
 Achieved convergence tolerance: 3.652e-06
```

# The Asymptotic Properties

The asymptotic properties are based on the local linearization of the model (cf. iterative estimation procedure, SLIDE 16, step 2)

$$\underline{Y} = \underline{\eta}\left\langle \underline{\theta} \right\rangle + \boldsymbol{A}\,\underline{\widehat{\beta}} + \underline{E}$$

where $\boldsymbol{A}\left\langle \underline{\theta} \right\rangle$ is the $n \times p$ matrix of partial derivatives.

If the estimation procedure has converged, then $\underline{\widehat{\beta}} = \underline{0}$.

Asymptotic Distribution of the Least Squares Estimator

$$\underline{\widehat{\theta}} \overset{a}{\sim} \mathcal{N}\left\langle \underline{\theta}, \boldsymbol{V}\left\langle \underline{\theta} \right\rangle \right\rangle$$

with asymptotic covariance matrix

$$\boldsymbol{V}\left\langle \underline{\theta} \right\rangle = \sigma^2 \left( \boldsymbol{A}\left\langle \underline{\theta} \right\rangle^T \boldsymbol{A}\left\langle \underline{\theta} \right\rangle \right)^{-1}$$

# Application in Practise

To explicitly determine the covariance matrix $V\langle\underline{\theta}\rangle$, we plug-in estimates instead of true parameters:

- $\boldsymbol{A}\langle\underline{\theta}\rangle$ is calculated using $\widehat{\underline{\theta}}$ ☞ $\widehat{\boldsymbol{A}}$.
- For the error variance $\sigma^2$ we plug-in the usual estimation.

Hence,

$$\widehat{\boldsymbol{V}} = \widehat{\sigma}^2 \left(\widehat{\boldsymbol{A}}^T \widehat{\boldsymbol{A}}\right)^{-1}$$

where

$$\widehat{\sigma}^2 = \frac{S\left\langle\widehat{\underline{\theta}}\right\rangle}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n}\left(y_i - \eta_i\left\langle\widehat{\underline{\theta}}\right\rangle\right)^2 \qquad \text{and} \qquad \widehat{\boldsymbol{A}} = \boldsymbol{A}\left\langle\widehat{\underline{\theta}}\right\rangle .$$

# Approximate 95%-confidence interval

Hence, an approximate 95%-confidence interval for $\beta_k$ is

$$\widehat{\theta}_k \pm \widehat{\mathrm{se}}\left\langle \widehat{\beta}_k \right\rangle \cdot q_{0.975}^{t_{n-p}},$$

where $\widehat{\mathrm{se}}\left\langle \widehat{\beta}_k \right\rangle$ is the square root of the $k$th diagonal element of $\widehat{\boldsymbol{V}}$.

Note: Theoretically correct is to use the quantiles of the Gaussian distribution. In practise, however, $t$ quantiles are preferred.

Example "Cellulose Membrane"
From the summary output

Parameters:

| | Value | Std. Error | t value | Pr($>\|t\|$) | |
|---|---|---|---|---|---|
| $\theta_1$ | 163.706 | 0.1262 | 1297.26 | $<$ 2e-16 | *** |
| $\theta_2$ | 159.785 | 0.1594 | 1002.19 | $<$ 2e-16 | *** |
| $\theta_3$ | 2.675 | 0.3813 | 7.02 | 3.65e-08 | *** |
| $\theta_4$ | -0.512 | 0.0703 | -7.28 | 1.66e-08 | *** |

Residual standard error: 0.293137 on 35 degrees of freedom

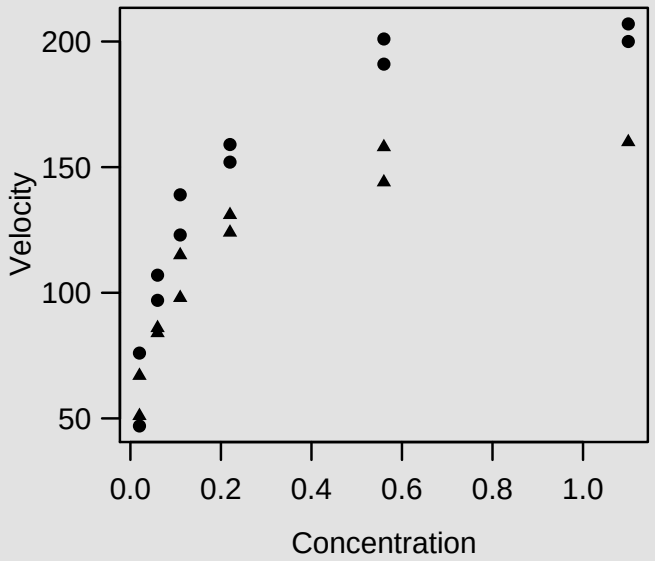
we can calculate the 95% confidence interval for $\theta_1$:

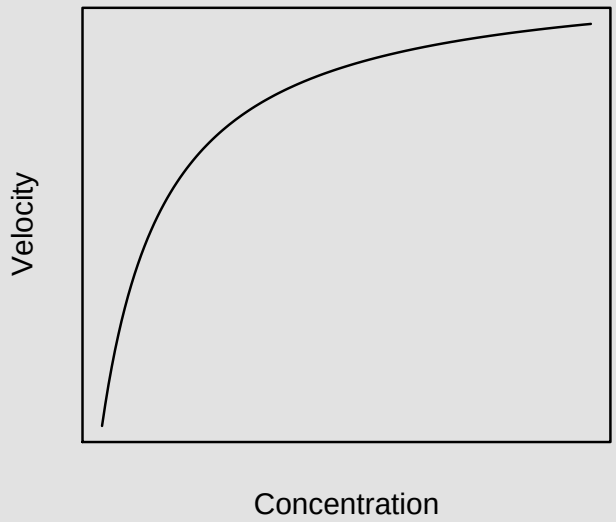
$$163.71 \pm 0.13 \cdot q_{0.975}^{t_{35}} = 163.71 \pm 0.26$$

# Example: Puromycin - back to the initial data set

The Michaelis-Menten model for enzyme kinetics relates the initial "velocity" of an enzymatic reaction to the substrate concentration

- ● teated with Puromycin
- △ not treated




$$Y_i = \frac{\theta_1 \cdot x_i}{\theta_2 + x_i} + E_i \quad \text{with} \quad E_i \; i.i.d. \sim \mathcal{N}\left\langle 0, \sigma^2 \right\rangle \qquad \text{(Michaelis-Menten model)}$$

$x$ substrate concentration [ppm]
$Y$ initial "velocity" [(number/min)/min]

# Example: Puromycin (4)

Modell: $$Y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + E_i \ .$$

Model with and without treatment (all data):

$$Y_i = \frac{(\theta_1 + \theta_3 z_i) x_i}{\theta_2 + \theta_4 z_i + x_i} + E_i \ .$$

$$\text{where } z_i = \begin{cases} 1 & \text{for „with“} \\ 0 & \text{for „without“} \end{cases}$$

Working hypothesis: **Only** the asymptotic velocity $\theta_1$ is influenced by adding Puromycin. Hence

Null hypothesis: $\theta_4 = 0$

R output for the example Puromycin

Parameters:

| | Value | Std. Error | t value | Pr($>\|t\|$) |
|---|---|---|---|---|
| $\theta_1$ | 160.286 | 6.8964 | 23.24 | 2.04e-15 |
| $\theta_2$ | 0.048 | 0.0083 | 5.76 | 1.50e-05 |
| $\theta_3$ | 52.398 | 9.5513 | 5.49 | 2.71e-05 |
| $\theta_4$ | 0.016 | 0.0114 | 1.44 | 0.167 |

Residual standard error: 10.4 on 19 df

Since the P-value of 0.167 is larger than the level of 5%

the null hypothesis is not rejected on the 5% level.

95% confidence interval for $\theta_4$:
$0.016 \pm 0.0114 \cdot q_{0.975}^{t_{19}} =$
[-0.0079, 0.0399]

# Inference for the expected value $E\langle Y|x_o\rangle = h\langle x_o; \underline{\theta}\rangle$ at $x_o$

## Linear Regression

$h\langle \underline{x}_o, \underline{\beta}\rangle = \underline{x}_o^T \underline{\beta}$ is estimated by

$$\widehat{\eta}_o = \underline{x}_o^T \widehat{\underline{\beta}}.$$

$(1-\alpha)\cdot 100\%$ confidence interval for $h\langle \underline{x}_o, \underline{\beta}\rangle$ is

$$\widehat{\eta}_o \pm q_{1-\alpha/2}^{t_{n-p}} \cdot \mathrm{se}\langle \widehat{\eta}_o\rangle$$

with $\mathrm{se}\langle \widehat{\eta}_o\rangle = \widehat{\sigma}\sqrt{\underline{x}_o^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\underline{x}_o}$

## Nonlinear Regression

$h\langle \underline{x}_o, \underline{\theta}\rangle$ is estimated by

$$\widehat{\eta}_o = h\left\langle \underline{x}_o, \widehat{\underline{\theta}}\right\rangle .$$

$(1-\alpha)\cdot 100\%$ confidence interval for $h\langle \underline{x}_o, \underline{\theta}\rangle$ is

$$h\left\langle x_o, \widehat{\underline{\theta}}\right\rangle \pm q_{1-\alpha/2}^{t_{n-p}} \cdot \mathrm{se}\langle \widehat{\eta}_o\rangle$$

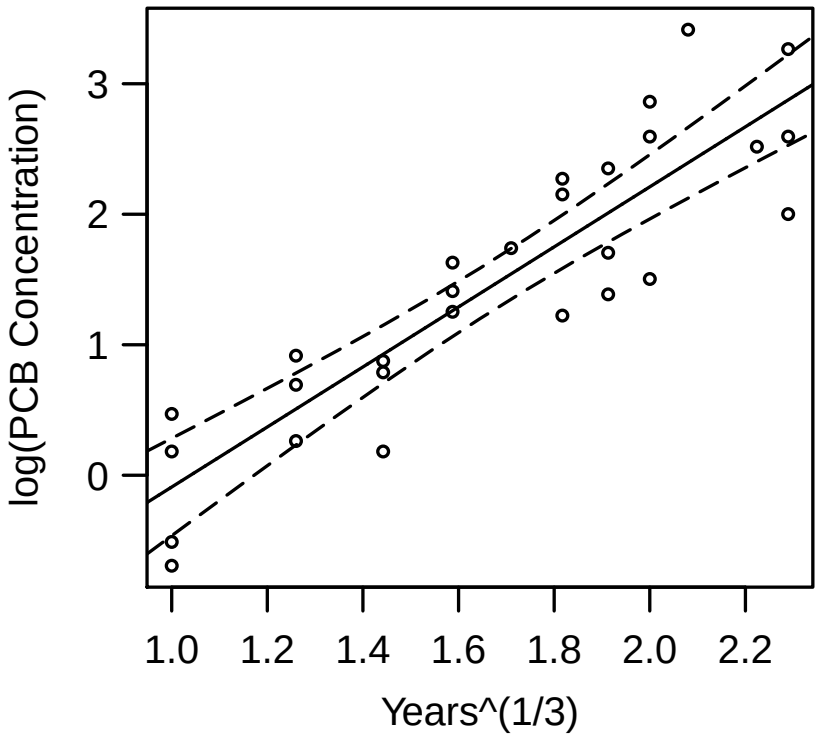
with $\mathrm{se}\langle \widehat{\eta}_o\rangle = \widehat{\sigma}\sqrt{\widehat{\underline{a}}_o^T\left(\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}}\right)^{-1}\widehat{\underline{a}}_o}$

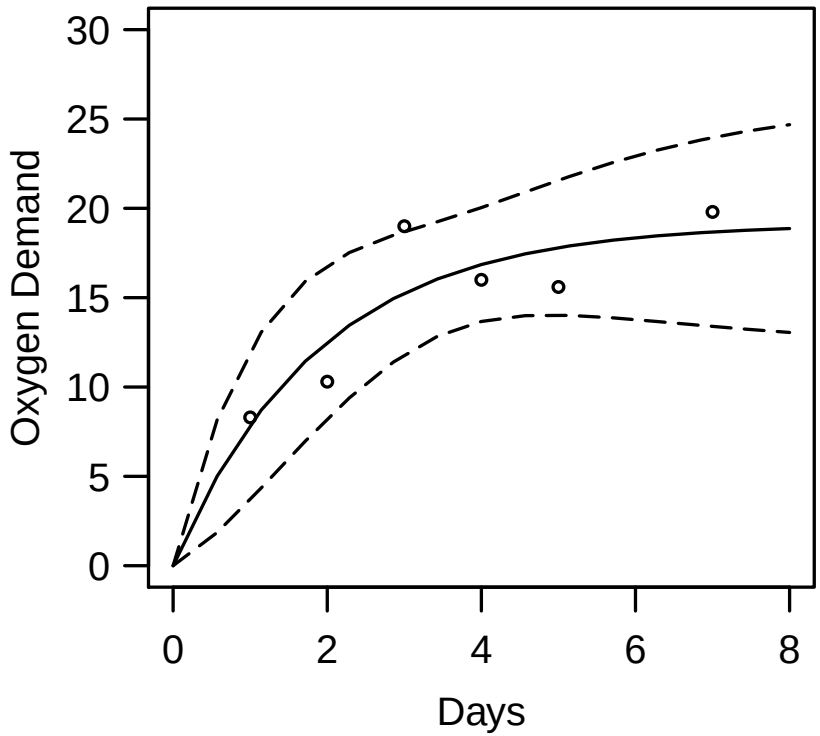
and $\widehat{\underline{a}}_o = \left.\frac{\partial h\langle x_o, \underline{\theta}\rangle}{\partial \underline{\theta}}\right|_{\underline{\theta}=\widehat{\underline{\theta}}}.$

# Confidence Band

Left: Confidence band (i.e., pointwise confidence intervals) for a fitted straight line (linear regression model).

Right: Confidence band for the fitted curve $h\langle x, \underline{\theta}\rangle$ of the example 'Biochemical Oxygen Demand'.



# Variable Selection

How about variable selection in nonlinear regression?

- There is no one-to-one correspondence between predictor variables and parameter as in linear regression!
  ☞ the number of variables may differ from the number of parameters.
- There are hardly ever problems, where some of the variables are in question
  because model is derived from subject matter theory!
- However, there are problems where a nested submodel may be adequat to describe the data;
  cf. Example Puromycin, SLIDE 28.
- Model Selection: If we have a collection of candidate models
  - which need *not* to be nested submodels of each other and
  - the subject matter is somehow indifferent to this models,
  - but we want to find the most appropriate model for the data

  one can use Akaike's information criterion (AIC) to select the best model and run a residual analysis to confirm the selection.

# Take Home Message Half-Day 1

- In **nonlinear regression**,

$$Y_i = h\langle \mathbf{x}_i, \underline{\theta} \rangle + E_i \ ,$$

functions $h$ are analysed which are not linear functions of the unknown parameters $\underline{\theta}$.

Such models are often derived from the **subject matter theory**.

- The flexibility of this model class is bought by a **more complex estimation and inference theory**.
  - Parameter estimation is done by an iterative procedure which needs **appropriate starting values**.
  - Inference is based on an asymptotic theory. For finite sample size the results just hold **approximately**
- **Model assumptions are assessed** like in linear regression modelling.