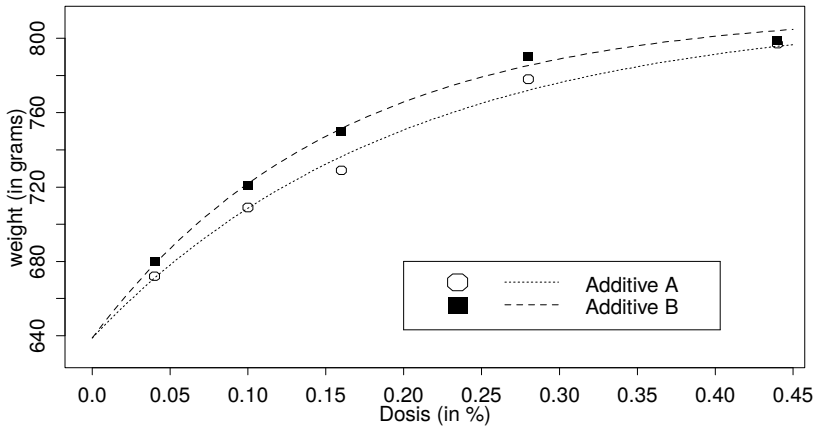
Series 5

1. Two different methionine food additives in different doses have been given to newborn turkeys in order to investigate the growth of turkeys (for details see Noll et al., 1984). The doses ranged from 0.04% to 0.44% of the total feeding of the turkeys. 15 turkeys were assigned to each of these dose-additive combinations. The target variable is the average weight of the 15 turkeys in each dose-additive combination. The data `http://stat.ethz.ch/Teaching/Datasets/cas-das/body.dat` contains the result of this experiment.



The growth Y_i (variable `weight`) of the turkeys depending on the doses $x_{i,k}$ of the two different methionine food additives ($k = 1, 2$) is described by the nonlinear model

$$Y_i = h\langle x_{i,1}, x_{i,2}\rangle = \theta_1 + \theta_2 \left\{1 - \exp\langle\theta_3(\theta_4 x_{i,1} + x_{i,2})\rangle\right\} + E_i,$$

where $x_{i,1}$ is the dose of the additive A (variable `sourceA`) and $x_{i,2}$ is the dose of the additive B (variable `sourceB`).
Considering the form of the curve, we can assume that $\theta_3 < 0$ and $\theta_4 > 0$. The model thus assumes that the intercept θ_1 and the horizontal asymptote $\theta_1 + \theta_2$ are the same for both methionine food additives and that the curves only differ in θ_3 and $\theta_3 * \theta_4$.

Now we would like to find out whether there is a difference of the "growth curves" for the two food additives. In other words, we would like to know if θ_4 is significantly different from 1.

a) Determine the starting values of the parameters θ_1, θ_2, θ_3 and θ_4.
Hint:

- A starting value for θ_1 can be obtained by using the fact that for $x_{i,1} = x_{i,2} = 0$ we have $h\langle \boldsymbol{x}_i\rangle = \theta_1$.
- A starting value for $\theta_1 + \theta_2$ can be obtained by noting that $h\langle x_{i,1}, x_{i,2}\rangle \to \theta_1 + \theta_2$ for the dose $\boldsymbol{x}_i = (x_{i,1},\ x_{i,2})^{\mathrm{T}} \to \infty$.
- When θ_1 and θ_2 are known, it is possible to linearize the equation to $\log\left(1 - \frac{Y_i - \theta_1}{\theta_2}\right) = \ldots$. We can then get starting values for θ_3 and θ_4 by using least-squares or robust MM-estimation for the linearized model. Note that there is no intercept!

b) Apply nonlinear regression. What are the estimates for the unknown parameters?
R-Hint:

- Set the argument `control = nls.control(maxit = 1000)` in the `nls()` function (this only needs to be done to make sure subtask **d)** is working).
- If you have problems in solving **a)**, use $\theta_1^{(0)} = 638$, $\theta_2^{(0)} = 175$, $\theta_3^{(0)} = -5$ and $\theta_4^{(0)} = 1$ as starting values.

c) Is there a difference in the curves for the two food additives? In other words, is θ_4 significantly different from 1?
Hint: Calculate the approximate 95%-confidence-intervals based on the Wald-statistic (by hand) as in Section 1.3.e. in the script.

d) Look at the likelihood profile traces. Which problems have to be investigated more closely?
R-Hint: The R-function `p.profileTraces()` is in the package `sfsmisc`.

```
> library(sfsmisc)
> r.prof <- profile(r.nls); p.profileTraces(r.prof)
```

e) Can we trust the conclusion in **c)**?
Hint: Look at the profile t-plot for θ_4.

f) Determine the 95%-confidence-intervals for all parameters θ_i based on the profile t-functions.
R-Hint: Use the function `confint()` from the package `MASS`.

```
> confint(profile(...))  # or confint(nls-obj.)
```

2. Isomerization of n-Pentan: Description of the data by Bates and Watts (1988)

"Data on the reaction rate of the catalytic isomerization of n-pentane to isopentane versus the partial pressures of hydrogen, n-pentane, and isopentane were given in Carr (1960) and are reproduced in `isomer.dat`.
Isomers are molecules of the same atomic composition (identical molecular formula, i.e. same number of each element contained therein), but different spatial arrangement. Isomerization is the conversion of one isomer into another.Catalytic isomerization employs catalysts to speed the reaction. The reaction rate depends on various factors, such as partial pressures of the products and the concentration of the catalyst. The differential reaction rate was expressed as grams of isopentane produced per gram of catalyst per hour (hr^{-1}), and the instantaneous partial pressure of a component was calculated as the mole fraction of the component times the total pressure, in pounds per square inch absolute (psia)."

Source: Bates D. M. and Watts D. G. (1988), *Nonlinear Regression Analysis & Its Applications*, John Wiley & Sons. Carr N. L. (1960), *Kinetics of catalytic isomerization of n-pentane*, Industrial and Engineering Chemistry, 52 (391–396).

A common model for rates of reaction is the *Hougen-Watson-Model*, which has the following special case:

$$y \approx h(\mathbf{x}, \boldsymbol{\theta}) = \frac{\theta_1 \theta_3 (x_2 - x_3/1.632)}{1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}, \tag{1}$$

where y is the reaction rate, x_1 is the partial pressure of hydrogen, x_2 is the partial pressure of n-pentan and x_3 is the partial pressure of isopentan. Note that the components of the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$ must be **positive** to get a meaningful physical model.
The data is stored in `http://stat.ethz.ch/Teaching/Datasets/cas-das/isomer.dat`.

a) Starting Values: Model (1) is a transformable linear model. Linearize it and determine the starting values $\boldsymbol{\theta}^{(0)}$ both with least-squares and with the regression-MM-estimation (`lmrob()`). Which starting values would you choose and why?
Hint: Rewrite the equation above in the form $\widetilde{y} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3$ (get all θ_k's on the right hand side of the equation). By fitting a least-squares or robust-MM-regression model to the data, you will get the estimates $\widehat{\beta}_k$. What is the relation between the θ_k's and the β_k's?

b) Estimation: Estimate the parameters for the nonlinear model. The summary of the result (`summary(..., corr = TRUE)`) hints serious problems. Which ones? Note two of them.

c) Profile t-plots and likelihood profile traces: Look at the profiles t-plots and the likelihood profile traces. Can you confirm your suspicion built up in **b)**?
R-Hint: Since the regression function is strongly nonlinear, the argument `delta.t` in the function `profile()` must be set to 0.2 and we can only plot the t-profiles for the parameters θ_2, θ_3 and θ_4. Use the R-function `p.profileTraces()` to get plots for the profiles t-functions and the likelihood profile traces:

```
> library(sfsmisc)
> r.prof <- profile(r.nls, which = 2:4, delta.t = 0.2)
> p.profileTraces(r.prof)
```

3. Isomerisation of n-Pentan (continued).

In Exercise 2 we have seen that the linear approximation around the parameter estimates is very bad. We now try to improve this approximation.

a) In Section 2.3.f in the script, a reparametrization for the model

$$y \approx h(\mathbf{x}, \boldsymbol{\theta}) = \frac{\theta_1 \theta_3 (x_2 - x_3/1.632)}{1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}$$

is suggested. Apply this parametrization and estimate the new parameters.
Hint: Compare with the solution of Exercise 2.

b) How do the new parameters affect the profile t-plots and the likelihood profile traces?

c) Compute a 95%-confidence-interval for each transformed parameter - once based on the Wald-statistics (approximate - by hand), once based on the profile t-function. Compare the two methods.

d) Calculate an approximate 95%-confidence-interval for θ_1 using the relation between θ_1 and φ_3 defined in part **a)**.
Hint: See Sections 2.3.i. and 2.3.j. in the script.

e) How well does the model fit the data? (Residual analysis).

Exercise hour: Monday, June 24, morning.