Exercise 1: Biochemical Oxygen Consumption (BOC)

The biochemical oxygen consumption is used as an indicator of the degree of pollution in waste water. After sealing of a bottle of waste water and keeping it a certain number of days at constant temperature, the seal is broken and the oxygen loss is measured in mg/l (`boc`). This is done for a total of 24 bottles, 4 in each group, indicating the 'number of days sealed' (`days`).

The results were:

days				
1	105	97	104	106
2	136	161	151	153
3	173	179	174	174
5	195	182	201	172
7	207	194	206	213
10	218	193	235	229

We want to give a description of the oxygen loss (`boc`) over time (`days`)

1. Make a scatter plot of `boc` vs. `days` and comment on the adequacy for making a linear regression on this scale.

   The biologists claim that the relation between `boc` and `days` can be described by a relation of the form

   `boc`=γ exp(-β/`days`)

2. Give an interpretation of the parameter γ
   (If you like mathematics, you may also give an interpretation of β. Try e.g. to look at the time until half the maximum `boc` is obtained)
3. Transform the above *mathematical relation* (i.e. the theoretical equation) with the natural logarithm and judge from a suitable scatter plot, whether this transformed relation may be used as a basis for a linear regression (do the assumptions seem to be fulfilled?)
4. Perform the above mentioned regression analysis and use the results for construction of a 95% confidence interval for the asymptotic oxygen loss, when the number of days of sealing gets very large.
5. Perform a numerical test of linearity in the regression above by comparing it to the anova-model, with 6 unrelated `days`-groups. Is linearity reasonable?
   Can we test linearity with more power?

turn the page!

Exercise 2: Height vs. age in the `Juul`-data

We want to model height as a function of age, with the aim of comparing boys and girls (and eventually also for constructing reference charts, i.e. normal regions for each age).

We once again use the data in `juul2.txt` and restrict ourselves to look at the ages 5 to 20.

1. Fit a straight line for male height as a function of age, and look at the fit graphically.
   What is the matter (not surprisingly) with this model?
2. Look at fits of polynomia of second and third order, perhaps after transformation of height. How can we compare these models to similar models for females?
3. You may also want to campare the above fits with the automatic *smoothing*-possibilities in SAS, using '`i=smxx`' in the `symbol`-statement (`xx` denoting a number between 1 and 99, indicating the degree of smoothness).

   Instead of polynomias, we will now fit a piecewise straight line, i.e. a **linear spline**.

4. Introduce cutpoints for a linear spline, e.g. at the ages 10, 12, 13 and 15. Define variables denoting number of years above these respective thresholds and fit a linear spline for boys.
5. At which age do we first see a deviation from a simple straight line?
6. Compare with similar results for the girls.
7. Fit a model for both genders simultaneously and find out at which age we first see a difference between girls and boys.
   Quantify this difference, in numbers as well as in words.