

## HW #5 Stat 511 Spring 2008

Refer to 2004 HW#6 2004 and 2003 HW #7 (and the first part of 2007 HW#8) and the associated keys for patterns for the R code needed to do the following problem.

The book of Bates and Watts has several biochemical oxygen demand data sets taken from a University Wisconsin M.S. thesis of Marske similar to the data set used on HW#7 of 2003. Here is one of them.

Time, $t$ (days)	1	2	3	4	5	7	9	11
BOD, $y$ (mg/l)	.047	.74	1.17	1.42	1.60	1.84	2.19	2.17

Two standard possible nonlinear models here are

$$y_i = \theta_1 (1 - \exp(-\theta_2 t_i)) + \varepsilon_i \quad (\text{exponential decay})$$

and

$$y_i = \frac{\theta_1 t_i}{\theta_2 + t_i} + \varepsilon_i \quad (\text{Michaelis-Menten})$$

(Note that both mean functions are 0 at  $t = 0$  and have asymptote  $\theta_1$  for large  $t$ . In the first,  $\theta_2$  is the reciprocal of the time at which the mean reaches 63% of its limiting value and in the second,  $\theta_2$  is the time at which the mean reaches half of its limiting value.)

a) Plot  $y$  vs  $t$  and make "eye estimates" of the model parameters for both models. Use these as starting values and employ the `nls` package in R to fit both models. Then replot the data along with 2 fitted mean curves, one for each model. Does one or the other of the mean functions "look" better than the other? Which fitted model has the smaller "deviance"? Compute and plot residuals versus  $t$  for both models. Normal plot the two sets of residuals. On the basis of these, do you see obvious problems with an "iid normal" model for the  $\varepsilon_i$  in either of the models?

Regardless of how your analysis in a) comes out, for the rest of this problem use the iid normal errors version of the exponential decay model.

b) Verify that `nls` outputs  $MSE(\hat{\mathbf{D}}\hat{\mathbf{D}})^{-1}$  as an estimated covariance matrix for  $(\hat{\theta}_{1OLS}, \hat{\theta}_{2OLS})'$  and that the standard errors it produces are the square roots of diagonal elements of this matrix. Use this matrix and make approximate 95% confidence limits for the mean BOD at 6 days and then approximate 95% prediction limits for an observed BOD at 6 days.

c) The time at which mean BOD reaches 50% of its limiting value is a function of  $\boldsymbol{\theta}$ . Find a sensible point estimate of this quantity and an estimated standard deviation for this estimate.

d) Make a contour plot of  $\sum_{i=1}^8 (y_i - \theta_1 (1 - \exp(-\theta_2 t_i)))^2$ , placing on the plot contours from which you can identify both an approximately 95% joint confidence region for  $\boldsymbol{\theta}$ , and approximate 95% confidence limits for the two parameters  $\theta_1$  and  $\theta_2$  individually.

- e) Compare your individual confidence limits from d) to 95%  $t$  intervals for the two parameters based on the estimated covariance matrix from b). Add the MASS package to your R environment in order to get access to the function `confint()` and use it to make approximate 95% intervals for  $\theta_1$  and  $\theta_2$ . How do these intervals compare to the others?
- f) Make two different approximately 95% confidence intervals for  $\sigma$ , one based on carrying over the linear model result that  $SSE / (n - k) \sim \chi^2_{n-k}$  to the nonlinear model, and the other based on the "profile likelihood" material. Does the function `confint()` provide limits for  $\sigma$ ?
- g) (See part l) of the 2004 HW#6) Plot approximately 95% *simultaneous* confidence limits for all values of mean BOD for  $t \leq 15$  based on inversion of the approximately 95% joint confidence region in d).