Exercise 5

Question 1.

Let $Y_1, \cdots, Y_n \sim f_\theta(y)$ with an unknown parameter θ.

1. Show that the sample α-quantile $y_{(\alpha)}$ is the M-estimator corresponding to $\rho(y, \theta) = \alpha(y - \theta)_+ + (1 - \alpha)(\theta - y)_+$.
2. What is the asymptotic distribution of $y_{(\alpha)}$? Find an (asymptotic) $(1 - \delta)100\%$ confidence interval for the α-quantile $F_\theta^{-1}(\alpha)$ of the distribution f_θ.

Question 2.

The file Hornets.dat (Hornets.dat) provides the results of the research on hornets' cells building. The file contains the numbers of hornets in the *i*-th box and the numbers of cells per capita, *CPC*, (hornet) that were built in the *i*-th box.

1. Fit a linear model of *CPC* as a function of *log(Hornets)*. What can you say about the adequacy of the model? Try to find an appropriate transformation of the response and re-fit the model. Comment the results.
2. Fit robust regression using several M-estimators: Hampel, Huber, Tukey's bisquare, etc. Compare the results and compare them with the OLS fit from the previous paragraph.

Question 3.

The file Puromycin.dat (Puromycin.dat) contains the data on the substrate concentration of Puromycin, x (parts per million, ppm) and the initial rate, or "velocity", y, of the enzymatic reaction (counts/min^2) in the presence of Puromycin. The velocity is assumed to depend on the substrate concentration according to the Michaelis-Menten equation: $y=\theta_1 x/(\theta_2+x)$.

1. What is the physical/mathematical meaning of the parameters θ_1 and θ_2?
2. Using transformations of x and y transform the original nonlinear model to a linear one. Fit the corresponding linear model. Does it seem to be adequate?
3. Find the gradient matrix D for the Michaelis-Menten original nonlinear model. Fit the nonlinear model using the results of the previous paragraph for obtaining initial values for the parameters, and check the adequacy of the nonlinear model.
4. Test the hypothesis θ_1=200 applying F- and t-tests, comment the results.
5. Predict velocity of the enzyme reaction when the substrate concentration of Puromycin is at level 0.5. Give the corresponding confidence and prediction intervals.

Computational Notes for R users:

- To fit various robust regression models use **rlm** and **lqs** functions (see *help* for details).
- To fit nonlinear regression use the function **nls** (see *help* for details).