Regression Methods: Problems MATH-408
Anthony Davison

Problem 1 (Smoking data) Consider the data on lung cancer deaths in British male physicians shown in the lectures. Suppose the number of deaths y in a cell of the table is a Poisson variable with mean $T\lambda(d,t)$, where T is man-years at risk, d is number of cigarettes smoked daily and t is time smoking (years), and that we use the standard epidemiological model

$$\lambda(d,t) = \beta_0 t^{\beta_1}\left(1+\beta_2 d^{\beta_3}\right),$$

where the background rate of lung cancer is $\beta_0 t^{\beta_1}$ for non-smokers and the additional risk due to smoking d cigarettes/day is $\beta_2 d^{\beta_3}$; note that the latter is zero if $d=0$. We anticipate that all the parameters β_r will be positive.

(a) With $x_j = (T_j, d_j, t_j)$, check that we can write

$$\begin{aligned} y_j &\stackrel{\text{ind}}{\sim} \text{Poiss}\{\eta_j(\beta)\}, \\ \eta_j(\beta) &= T_j\beta_0 t_j^{\beta_1}\left(1+\beta_2 d_j^{\beta_3}\right), \quad j=1,\ldots,n. \end{aligned}$$

In this case $n=63$, corresponding to the 9×7 cells of the data table.

(b) Give the log-likelihood function of this model, and hence compute the elements of the matrices used in the iterative weighted least squares algorithm.

Problem 2 The deviance and Pearson residuals are defined as

$$d_j = \text{sign}(\tilde\eta_j - \widehat\eta_j)[2\{\ell_j(\tilde\eta_j;\phi) - \ell_j(\widehat\eta_j;\phi)\}]^{1/2}, \quad j=1,\ldots,n,$$

with $\sum d_j^2 = D$ the deviance, and

$$P_j = u_j(\widehat\beta)/\sqrt{w_j(\widehat\beta)}, \quad j=1,\ldots,n.$$

These quantities are usually standardized to

$$r_{D_j} = \frac{d_j}{(1-h_{jj})^{1/2}}, \quad r_{Pj} = \frac{u_j(\widehat\beta)}{\{w_j(\widehat\beta)(1-h_{jj})\}^{1/2}}, \quad j=1,\ldots,n.$$

Prove that r_{D_j} and r_{Pj} both reduce to the usual standardized residual in the case of a normal linear model, and compute r_j^*.

Problem 3 Show that the gamma density

$$f(y;\mu,\nu) = \frac{1}{\Gamma(\nu)} y^{\nu-1}\left(\frac{\nu}{\mu}\right)^{\nu}\exp\left(-\frac{\nu y}{\mu}\right), \quad y>0, \quad \nu,\mu>0,$$

can be written as a GLM density and give its mean and variance and canonical link functions. Compare the latter with the link function $\eta = \log\mu$.

Problem 4 If X is Poisson with mean $\exp(x^{\mathrm{T}}\beta)$ and the binary variable Y indicates the event $X>0$, find the link function between $\mathrm{E}(Y)$ and $\eta = x^{\mathrm{T}}\beta$.