Myrto Limnios, Marco Piccininni

Problem 1 (Smoking data) Consider the data on lung cancer deaths in British male physicians described below.

Table 1: Lung cancer deaths in British male physicians (Doll and Hill, 1952). The table gives man-years at risk T/number of cases y of lung cancer, cross-classified by years of smoking t, taken to be age minus 20 years, and number of cigarettes smoked per day, d.

Years of smoking t	Daily cigarette consumption d						
	Nonsmokers	1–9	10–14	15–19	20–24	25–34	35+
15–19	10366/1	3121	3577	4317	5683	3042	670
20–24	8162	2937	3286/1	4214	6385/1	4050/1	1166
25–29	5969	2288	2546/1	3185	5483/1	4290/4	1482
30–34	4496	2015	2219/2	2560/4	4687/6	4268/9	1580/4
35–39	3512	1648/1	1826	1893	3646/5	3529/9	1336/6
40–44	2201	1310/2	1386/1	1334/2	2411/12	2424/11	924/10
45–49	1421	927	988/2	849/2	1567/9	1409/10	556/7
50–54	1121	710/3	684/4	470/2	857/7	663/5	255/4
55–59	826/2	606	449/3	280/5	416/7	284/3	104/1

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

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

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 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 $\eta = x^{\mathrm{T}}\beta$.

Problem 3 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^*.