Exercises Chapter 4

4.5 Exercises

4.1 The data in Table 4.5 show the numbers of cases of AIDS in Australia by date of diagnosis for successive 3-months periods from 1984 to 1988. (Data from National Centre for HIV Epidemiology and Clinical Research 1994.)

In this early phase of the epidemic, the numbers of cases seemed to be increasing exponentially.

(a) Plot the number of cases y_i against time period i ($i = 1, \ldots, 20$).

(b) A possible model is the Poisson distribution with parameter $\lambda_i = i^\theta$, or equivalently

$$\log \lambda_i = \theta \log i.$$

Plot $\log y_i$ against $\log i$ to examine this model.

(c) Fit a generalized linear model to these data using the Poisson distribution, the log-link function and the equation

$$g(\lambda_i) = \log \lambda_i = \beta_1 + \beta_2 x_i,$$

Table 4.5 *Numbers of cases of AIDS in Australia for successive quarter from 1984 to 1988.*

	Quarter			
Year	1	2	3	4
1984	1	6	16	23
1985	27	39	31	30
1986	43	51	63	70
1987	88	97	91	104
1988	110	113	149	159

where $x_i = \log i$. Firstly, do this from first principles, working out expressions for the weight matrix $\mathbf{W}$ and other terms needed for the iterative equation

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z}$$

and using software which can perform matrix operations to carry out the calculations.

(d) Fit the model described in (c) using statistical software which can perform Poisson regression. Compare the results with those obtained in (c).

4.2 The data in Table 4.6 are times to death, y_i, in weeks from diagnosis and $\log_{10}$(initial white blood cell count), x_i, for seventeen patients suffering from leukemia. (This is Example U from Cox and Snell 1981.)

Table 4.6 *Survival time, y_i, in weeks and $\log_{10}$(initial white blood cell count), x_i, for seventeen leukemia patients.*

y_i	65	156	100	134	16	108	121	4	39
x_i	3.36	2.88	3.63	3.41	3.78	4.02	4.00	4.23	3.73
y_i	143	56	26	22	1	1	5	65	
x_i	3.85	3.97	4.51	4.54	5.00	5.00	4.72	5.00	

(a) Plot y_i against x_i. Do the data show any trend?

(b) A possible specification for $\mathrm{E}(Y)$ is

$$\mathrm{E}(Y_i) = \exp(\beta_1 + \beta_2 x_i),$$

which will ensure that $\mathrm{E}(Y)$ is non-negative for all values of the parameters and all values of x. Which link function is appropriate in this case?

(c) The Exponential distribution is often used to describe survival times.

The probability distribution is $f(y;\theta) = \theta e^{-y\theta}$. This is a special case of the Gamma distribution with shape parameter $\phi = 1$ (see Exercise 3.12(a)). Show that $\mathrm{E}(Y) = 1/\theta$ and $\mathrm{var}(Y) = 1/\theta^2$.

(d) Fit a model with the equation for $\mathrm{E}(Y_i)$ given in (b) and the Exponential distribution using appropriate statistical software.

(e) For the model fitted in (d), compare the observed values y_i and fitted values $\widehat{y}_i = \exp(\widehat{\beta}_1 + \widehat{\beta}_2 x_i)$, and use the standardized residuals $r_i = (y_i - \widehat{y}_i)/\widehat{y}_i$ to investigate the adequacy of the model. (Note: $\widehat{y}_i$ is used as the denominator of r_i because it is an estimate of the standard deviation of Y_i—see (c) above.)

4.3 Let $Y_1, \ldots, Y_N$ be a random sample from the Normal distribution $Y_i \sim \mathrm{N}(\log\beta, \sigma^2)$ where σ^2 is known. Find the maximum likelihood estimator of β from first principles. Also verify equations (4.18) and (4.25) in this case.