Exercises 7

1. Consider the example with Poisson counts discussed in the lectures for which the log-linear model is
$$\log \mu_i = \beta_0 + \beta_1 x_i \ (i = 1, \ldots, N).$$
   (a) Show that log-likelihood function as a function of β_0 and β_1 (given the data) is
$$\ell = -\sum_{i=1}^{N} \mathrm{e}^{\beta_0+\beta_1 x_i} + \sum_{i=1}^{N} y_i(\beta_0 + \beta_1 x_i) + \text{ constant}$$
   (b) Do the following from first principles, ie without using the general results in the lecture notes Section 3.2.1.
      (i) Obtain the likelihood equations (by differentiating ℓ with respect to β_0 and β_1).
      (ii) Find the matrix of second derivatives of the log-likelihood, and use this to obtain the information matrix.

   (c) Now obtain the matrices $\mathbf{X}$ and $\mathbf{W}$ for this problem and hence obtain the information matrix using the result given in Section 3.2.1 of the lecture notes. Check that your result for information matrix agrees with that obtained in part (b)(ii).

2. Refer to R output 10 (Poisson example — speed and mistakes) and use a hand calculator for the numerical calculations in the following. The notation is that used in the lecture for this example.
   (a) Verify the fitted value for Speed = 1 (see end of output).
   (b) Verify that the estimates of β_0 and β_1 in the R output satisfy the likelihood equations obtained in question 1 above (you may use the fitted values given at the end of the output).
   (c) From the algebraic result for the information matrix obtained above, verify the estimated covariance matrix of the maximum likelihood estimators of β_0 and β_1 given in the R output, and hence verify the standard errors associated with the two estimates.
   (d) Obtain algebraically the form of the deviance for Poisson responses given in Section 3.2.3 (ii) of the notes, and hence verify the value of the (residual) deviance given in the R output.
   (e) One test of the null hypothesis $\beta_1 = 0$ is obtained by the 'z-value' in the R output. Identify another test of the same null hypothesis in the output (hint: look at Section 3.2.3 (iii)).
   (f) Calculate an approximate 95% confidence interval for β_1 (hint: refer to section 3.2.2 (i)).
   (g) Calculate an approximate standard error of the estimate $\hat{\eta} = \hat{\beta}_0 + 3\hat{\beta}_1$ of $\eta = \beta_0 + 3\beta_1$. Hence calculate an approximate 95% confidence interval for η and deduce an approximate 95% confidence interval for the expected response when Speed = 3.

3. (a) For a GLM with a canonical link function, show that the likelihood equations are
$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{a_i\,(\phi)} \right) x_{ij} = 0 \ (j = 1, \ldots, p) \ .$$
   (**Hint:** for the canonical link we have $\eta_i = \theta_i$ — see p50 of the notes).

   (b) Consider a GLM with a canonical link, and such that $a_i(\phi) = \phi$ is the same for all observations. Show that if the model contains an intercept term, the *residuals* $\{y_i - \mu_i\}$ from a maximum likelihood fit sum to zero, and that their sample correlation with each of the covariates is also zero.

4. (a) Let $\mathbf{M}$ be a $N \times p$ matrix with (i,j)th entry m_{ij}. Also, let $\mathbf{A}$ be a diagonal $N \times N$ matrix with ith diagonal entry a_{ii} and all other entries zero; and let $\mathbf{V}$ be a $N \times 1$ vector with ith entry v_i. Show that $\mathbf{M}'\mathbf{A}\mathbf{V}$ is a $p \times 1$ matrix with jth entry $\sum_{i=1}^{N} m_{ij} a_{ii} v_i$.

   (**Hint:** you will need to use the fact that if $\mathbf{F}$ and $\mathbf{G}$ are two matrices such that the product $\mathbf{FG}$ is defined, then the (i,j)th element of $\mathbf{FG}$ is $\sum_k \mathbf{F}_{ik}\mathbf{G}_{kj}$ in an obvious notation — this is just the usual definition of matrix multiplication).

   (b) For any GLM, let $\mathbf{U}(\boldsymbol{\beta}) = (\partial\ell/\partial\beta_1 \ldots \partial\ell/\partial\beta_p)'$ be the vector of log-likelihood derivatives. Starting from equation (3.8) of the lecture notes and with all quantities as defined subsequently in the notes, show that the likelihood equation $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ can be written in matrix-vector form as

$$\mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{W}\mathbf{z}\ .$$

   (c) According to the result in part (b), the maximum likelihood estimate of $\boldsymbol{\beta}$ clearly satisfies $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{z}$. Why, then, is an iterative procedure usually necessary to find the estimate as described in the notes?