Exercises 6

1. Consider the one-way ANOVA model given in §2.5 of the lecture notes:

$$Y_{ij} = \mu_i + e_{ij} \qquad (i = 1, \ldots, I;\ j = 1, \ldots, n_i)\ .$$

By writing this in the general linear model form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $\mathbf{X}$ and $\boldsymbol{\beta}$ as given in the lecture notes, and using the formula $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{Y}$, show that the least-squares estimator of $\boldsymbol{\beta}$ is the vector of group means.

2. The Round Loch of Glenhead is a lake in southwest Scotland. In an investigation to examine the effect of various factors upon the alkalinity of the loch, alkalinity measurements were made every three months for a period of almost 15 years. The mean air temperature, in degrees Celcius, was also recorded for both the month (`Mtemp`) and the three-month period (`Qtemp`) preceding each alkalinity measurement; and the total precipitation (`Mrain`) was recorded for the month preceding each alkalinity measurement. The following R output is extracted from a linear model fit in which the alkalinity was regressed upon these various quantities (note that the notation `Mtemp:Mrain` denotes an interaction between `Mtemp` and `Mrain`).

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -12.3722     0.5401 -22.906  < 2e-16 ***
Qtemp         4.1581     0.9413   4.417 4.96e-05 ***
Mtemp        -1.9794     1.0250  -1.931  0.05882 .
Mrain         0.2127     0.5483   0.388  0.69967
Mtemp:Mrain  -1.7977     0.5852  -3.072  0.00335 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.525 on 53 degrees of freedom
Multiple R-Squared: 0.4255,Adjusted R-squared: 0.3821
F-statistic: 9.812 on 4 and 53 DF,  p-value: 5.142e-06
```

   (a) Write down the mathematical representation of the model summarised by the output above, taking care to define any notation that you use. Give the estimated values of any model parameters.

   (b) The interaction term `Mtemp:Mrain` is highly significant. What is the interpretation of this interaction term?

   (c) Comment on the significance of the main effect terms `Mtemp` and `Mrain`.

   (d) Suppose now that `Mtemp` and `Qtemp` were measured in degrees Fahrenheit rather than degrees Celcius (recall that if x denotes temperature in degrees Celsius, then the corresponding temperature in degrees Fahrenheit is $32 + 9x/5$). What, if anything, would be the effect of this upon (i) the structure (ii) the parameters of the model you wrote down in part (a)? Would the change in measurement units affect the significance of the interaction term? What about the main effects?

3. Verify that the following distributions belong to the exponential family, and deduce the mean and variance from the general results in Section 3.1.1 of the notes:

   (a) $\mathrm{Bin}(n, \pi)$

   (b) $\Gamma(\alpha, \lambda)$ where both α and λ are unknown (pdf as in Rice Section 2.2)

In case (b), suggest a canonical link function and write down the variance function.

4. In each of the situations below, state whether the indicated model can be regarded as a GLM and give reasons for your answer. Greek letters indicate unknown parameters, Y_i denotes the i^{th} observation of a response variable corresponding to the value x_i of an explanatory variable. Assume independent observations for $i = 1, ..., N$.

   (a) $Y_i = 0$ or 1 where $P(Y_i = 1) = \mathrm{e}^{-\beta x_i}$.

   (b) $Y_i \sim$ exponential with mean μ_i where $\mu_i = \exp(\beta_0 + \beta_1 x_i)$.

   (c) $Y_i \sim$ Poisson with mean μ_i where $\mu_i = n_i \mathrm{e}^{\beta x_i}$ and n_i is known.

   (d) $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \log(\beta_0 + \beta_1 x_i)$.