Stat 201 - Generalized Linear Models

Optional Assignment

1. Review the R codes presented during this week (look in the directory `R material`).
2. Load the `leukemia` dataset from the library `Dobson`. It includes times to death, y_i, in weeks from diagnosis and $\log_{10}$ (initial white blood cell count), x_i, for leukemia patients. Plot the data to have a rough idea of the relationship between the two varables. Assume that

$$\begin{aligned} y_i \sim & f(y_i;\theta_i) = \theta_i e^{-\theta_i y_i} \\ \mathbb{E}Y_i = & \exp(\beta_0 + \beta_1 x_i) \end{aligned}$$

   (a) Is this a GLM? Motivate your answer.
   (b) Compute the score vector $\boldsymbol{U}(\boldsymbol{\beta})$ and the information matrix $\mathbf{I}(\boldsymbol{\beta})$.
   (c) Provide the Fisher scoring iteration for obtaining the MLE of $\boldsymbol{\beta}$
   (d) Write an R code that uses such iteration and obtain the MLE of $\boldsymbol{\beta}$
   (e) Obtain an approximate 95% confidence interval for β_1
   (f) Provide the deviance statistic of the model
   (g) By comparing the deviances for two appropriate models, test the null hypothesis $\beta_1 = 0$ against the alternative hypothesis $\beta_1 \neq 0$. What can you conclude about the use of the initial white blood cell count as a predictor of survival time?

3. Dobson-Barnett book: Exercise 5.1