ST 437/537 Homework 6, Spring 2018

ST 537: Q1-2. ST 437: Q2

1 Let the observed data be: $[Y_i, X_{i1}, \ldots, X_{iK}]$ for $i = 1, \ldots, n$ where Y_i is the scalar response and $X_i = (X_{i1}, \ldots, X_{iK})^T$ is a K-dimensional vector of covariates. It is assumed that $Y_i \sim Poisson(\mu_i)$ where μ_i is the mean of the response Y_i and $\log \mu_i = X_i^T \beta$ for some unknown K-dimensional parameter β and Y_i's are independent over i.

i Write down the log-likelihood function of the model, $\ell(\beta)$

ii Derive the estimating equations that are used to estimate the maximum likelihood estimate, $\widehat{\beta}$. Hint: Calculate the derivative function with respect to the parameter $\mathcal{D}(\beta) = \partial \ell(\beta)/\partial \beta$.

2 Consider the militiamen example (Hand et al. , 1994) where the numbers of Prussian militiamen killed by being kicked by a horse in each of 10 separate corps of militiamen are measured between 1875 – 1894. Here is a snapshot of the data; the full data is found under the name "militiamen.txt" in the data folder.

Table 1:

Obs	Year	Corps	Number of men killed
1	1875	1	0
2	1875	2	0
3	1875	3	0
4	1875	4	0
5	1875	5	1
6	1875	6	1
7	1875	7	0
8	1875	8	0
9	1875	9	1
10	1875	10	0
11	1876	1	0
⋮	⋮	⋮	⋮

i Write down a mathematical model that accounts only for the main effects of year and corps.

ii Use the R function `glm()` to fit this model in R. Make sure to specify the 'family=poisson(link = "log")' and use the 'log' link function.

iii Are the differences in the number of men killed attributed to systematic effects of year or corps?

iv Assess formally that the years 1875 and 1880 have the same effect on the number of horse kick deaths?. Use anova() to carry out this study.