Stat 201 - Generalized Linear Models

Optional Assignment

1. Review the R codes presented during this week (look in the directory `R material`).
2. As we have seen, the Normal density $N(\mu, \sigma^2)$ with σ^2 known is a member of the canonical one-parameter exponential family. Assuming $Y \sim N(\mu, \sigma^2)$, use this fact for computing
   (a) the mean $\mathbb{E}Y$
   (b) the variance $\mathrm{Var}Y$
   (c) the score function $U(\mu)$
   (d) the expectation and the variance of the score function
   (e) verify that
   $$\mathbb{E}U'(\mu) = -I(\mu)$$
3. Prove that under the (non necessarily canonical) one-parameter exponential family
   $$\mathbb{E}U'(\theta) = -I(\theta)$$
4. As we have seen, the Binomial distribution $\mathrm{Bin}(n, p)$ is a member of the canonical one-parameter exponential family. Assuming that $Y \sim \mathrm{Bin}(n, p)$, use this fact for computing
   (a) the mean $\mathbb{E}Y$
   (b) the variance $\mathrm{Var}Y$
   (c) the score function $U(p)$
   (d) the information number $I(p)$
5. Import the file `deaths.txt`: it includes the number of deaths by cancer occurred in Italy in 1999 and the population at risk (exposure), clustered by age and gender.
   (a) Plot death rates by age and intepret the pattern you see
   (b) Obtain the least square estimates of the parameters of the model
   $$\texttt{model1} \quad \log \mathrm{rate} = \beta_0 + \beta_1 \mathrm{age}$$
   and show the fit with an appropriate picture
   (c) Do you think that the `model1` is acceptable for these data? Why?
   (d) Are you able to suggest an alternative model that is better than `model1`? Motivate your answer with an analysis similar to that done in `examples chapter 2.R`.