Assignment 6

Stat 623

Due on Thursday Nov 19, 2020

1. Thall and Vail (1990) give a data set (seizure.data) on two-week seizure counts for 59 epileptics. The number of seizures was recorded for a baseline period of 8 weeks, and then patients were randomly assigned to a treatment group or a control group. Counts y_1, y_2, y_3 and y_4 were then recorded for four successive two-week periods. The subject's age is the only covariate. Treatments were either a 'placebo' (0) or 'progabide' (1). The counts in the baseline 8-week period are given by the variable base. For simplicity, we shall only consider y_4 as our response variable and ignore counts y_1, y_2 and y_3.

   (a) Fit a Poisson regression model to the response variable y_4 that makes use of subject's age, treatment group assigned and baseline counts. Summarize your results using Tables and graphs. What conclusions do you draw from the analysis?

   (b) Perform residual analysis to the above model. What conclusions do you draw?

   (c) Fit a negative binomial regression to y_4 with subject's age, treatment group assigned and baseline counts. Summarize your results using Tables and graphs.

   (d) Does a negative binomial regression provide a better fit to the data? If yes, how would you justify this.

2. Suppose that y follows an exponential distribution with mean μ. That is

$$f(y) = \mu^{-1} e^{-y/\mu}, \quad y > 0, \mu > 0.$$

   In homework 4, you wrote down the pdf of y as

$$\exp\left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}.$$

   and identified θ, $b(\theta)$, ϕ, $a(\phi)$ and $c(y, \phi)$.

   (a) Suppose that $y_1, y_2, \ldots, y_n$ are n independent random variables and y_i follows an exponential distribution with mean μ_i. Furthermore, assume that

$$g(\mu_i) = \eta_i = x_i^T \beta,$$

for some suitable link function g. Let $r_{d,i}, r_{p,i}, r_{a,i}$ and $r_{w,i}$ respectively denote the deviance, Pearson, Anscombe and working residuals for the i observation y_i after fitting a generalized regression model. Write down formulas for $r_{d,i}, r_{p,i}, r_{a,i}$ and $r_{w,i}$ in terms of y, X, θ, μ, b and g.

(b) Simplify the expressions in 1(b) when g is the canonical link function.