MAST30027: Modern Applied Statistics

Week 5 Lab

1. Incubation temperature can affect the sex of turtles. An experiment was conducted with three independent replicates for each temperature and the number of male and female turtles born was recorded. The data can be found in the `turtle` dataset in the `faraway` package.

   Check for evidence of overdispersion in a binomial model for the sex of the turtle.

   What problems can arise if you ignore overdispersion?

2. Suppose that $Y_i \sim \text{Poisson}(\lambda_i)$, where $\lambda_i \propto t_i$. For example, if we record the number of burglaries reported in different cities, the observed number will depend on the number of households in these cities. In other cases, the size variable t may be time. For example, if we record the number of customers served by sales people, we must take account of the differing amounts of time worked.

   We can model the rate *per unit time* using a log link via

   $$\log(\lambda_i/t_i) = x_i^T\beta$$

   where x_i are known predictors and β unknown parameters. That is

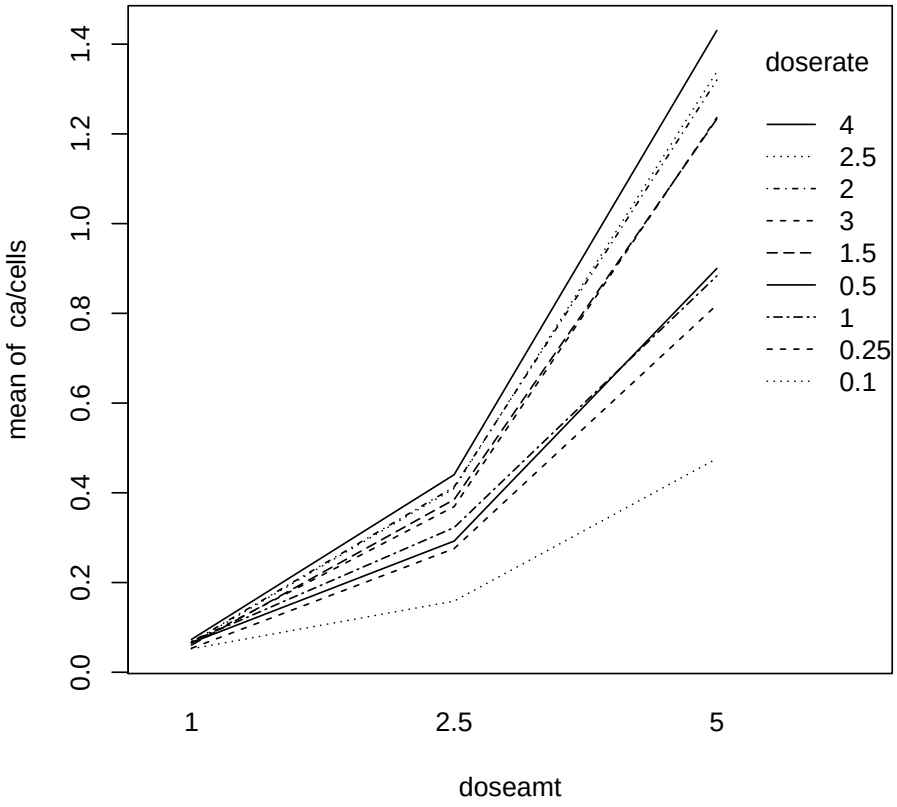
   $$\log(\lambda_i) = \log t_i + x_i^T\beta.$$

   This is of the form of a Poisson glm with log link, but where the coefficient of $\log t_i$ has been constrained to be 1. This is called a *rate model*.

   In an R model description we can fix the coefficient of a variable to 1 by enclosing it in the `offset` function, viz `y ~ offset(log(t)) + x1 + x2 + ⋯`.

   In Purott and Reeder (1976), some data is presented from an experiment conducted to determine the effect of gamma radiation on the numbers of chromosomal abnormalities (ca) observed. The number (cells), in hundreds of cells exposed in each run, differs. The dose amount (doseamt) and the rate (doserate) at which the dose is applied are the predictors of interest. We can plot the data as follows

   ```
   > library(faraway)
   > data(dicentric)
   > with(dicentric, interaction.plot(doseamt, doserate, ca/cells))
   ```



Fit a rate model to this data.

3. In Question 3, we fitted a rate model (a type of Poisson regression) to data on the effect of gamma radiation on chromosomal abnormalities.

   Show that these data are overdispersed compared to a Poisson distribution. Next test for an interaction between `doserate` and `doseamt`, firstly without allowing for overdispersion (fixing this dispersion $\phi = 1$), and secondly allowing for overdispersion. Do you get different answers?