**Stat 401 - lab 13 - part b: additional information**

Goals:

 How to fit a Poisson regression to a general count

 How to fit a logistic regression to binomial data

 How to fit overdispersed versions of Poisson and Binomial models

We will use case study 22.1 (male elephant mating success). The response variable is the number of matings for each of 41 elephants in an 8 year study of a population in Amboseli Park, Kenya. The questions are whether there is an association between elephant age and the number of matings, and whether this relationship has a detectable maximum. The data are in matings.csv. Load the data.

**How to fit a Poisson regression to count data**

The concept is very similar to fitting a logistic regression and multiple regression. The major difference is getting started, since JMP doesn't automatically know the difference between a count and a continuous variable.

Analyze / Fit model, and put Matings (blue ramp) into the Y box and Age (blue ramp) into the X box.

 We will ignore the attempts variable when we fit a Poisson regression

Change the Personality to Generalized Regression.

 Generalized regression extends multiple regression to many different response distributions

Change the Distribution to Poisson.

Just before you click Run, the Fit Model dialog box should look like:



Click Run.

The results include a box at the top labelled Model Launch. This provides a second list of options for model fitting, many of which are especially for many X variables. What you specify in this box is not used unless you click Go in that box. The default estimation method is Maximum Likelihood, which is the standard method for Poisson regression.

The results box has a lot of information about the fit. The results I find most useful are:

The estimated regression coefficients are in the Parameter Estimates for Original Predictors box. You should have a regression coefficient for age of 0.0689. Standard errors and Wald tests of true coefficient = 0 are provided for each coefficient. Type III (partial) tests of each parameter are obtained by expanding the Effect Tests box. The type III tests provide new information when for red bar variables, i.e. those that indicate groups.

The negative log likelihood, AICc and BIC statistics are in the Model Summary box.

Further information about the model and its predictions is obtained by clicking the red triangle by Maximum Likelihood for XXX. Some useful options are:

 Diagnostic plots - includes option for residual vs predicted value plots

 Save Columns - allows you generate formulae for predicted values and their variance.

 These formulae are added to the Data Window, so you can add new X values and get predictions.

 Regression reports - most show up by default. Can request coefficients for centered and scaled X's

**How to fit binomial data**

We will assume that the data in matings.csv are the result of a known number of attempts, so the data are Binomial not Poisson. Since each elephant was observed for 8 years, I set the number of attempts to 24 for each one. Now the response is the probability that an attempt results in a successful mating. The data are derived from 0/1 responses for each elephant, but the binomial format assumes that each male has his own probability, which may depend on age.

Analyze / Fit Model, set Personality to Generalized Regression and Distribution to Binomial

Binomial data needs two pieces of information about the response: # events and # trials

These are provided as two Y variables. The first is events; the second is trials. Put Matings (first) and Attempts (second) into the Y box and Age into the X box. Just before clicking Run, the Fit Model dialog should look like:



Click Run.

The results and follow-on options are very similar to those for Poisson regression. The coefficient for age is not exactly the same as that in Poisson regression because the Binomial model coefficient describes changes in the log odds of a successful mating while the Poisson model coefficient describes changes in the log mean number of matings.

**How to fit overdispersed Poisson or Binomial data**

The Generalized Regression model can account for overdispersion by changing the distribution to Beta Binomial (for overdispersed Binomial) and Negative Binomial (for overdispersed Poisson). The parameter estimates include an additional coefficient that quantifies (and tests) the amount of overdispersion. Dispersion values of 0 are equivalent to the non-overdispersed response distribution (Binomial or Poisson).

The documentation implies it is possible to fit a second form of overdispersed data (what the book fits), but I can't find it in JMP 13 Pro.