AMS 247: Homework 2

Diana Gerardo

October 22, 2018

1. Let y_i be realizations of the independent random variables $Y_i \sim \text{Pois}(\mu_i)$, where $\text{E}(Y_i) = \mu_i$, for $i = 1, \ldots, n$.

   (a) Obtain the expression for the deviance for comparison of the full model, which assumes a different μ_i for each y_i, with a reduced model defiend by a Poisson GLM with link function $g(\cdot)$. That is, under the reduced model, $g(\cdot) = \eta_i = x_i^T \beta$, where $\beta = \left(\beta_1, \ldots, \beta_p\right)^T$ (with p < n) is the vector of regression coefficients corresponding to covariates $x_i = \left(x_{i1}, \ldots, x_{ip}\right)^T$.

   Solution

$$\begin{aligned} f(y_i|\mu_i) &= \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \\ &= \exp\left[-\mu_i + y_i \log(\mu_i) - \log(y_i!)\right] \\ &= \exp\left[y_i \log(\mu_i) - \mu_i + (-\log(y_i!))\right] \end{aligned}$$

   Notice that $\text{Pois}(\mu_i)$ is in the EDF.

$$f(y_i|\theta, \phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right]$$

   where $y_i = y_i$, $\theta_i = \log(\mu_i)$, $b(\theta_i) = \mu_i$, abd $c(y_i, \phi) = -\log(y_i!)$, $a_i(\phi) = 1$.
   Then the following are the MLEs under,

   Full Model: $\widetilde{\mu}_i = y_i \Longrightarrow \widetilde{\theta}_i = \log(\widetilde{\mu}_i)$ and $b\left(\widetilde{\theta}_i\right) = \widetilde{\mu}_i = y_i$
   Reduced Model: $\widehat{\mu}_i = g\left(x_i^T \widehat{\beta}\right)$. Where $\widehat{\beta}$ denotes the mle of β. Thus, $\widehat{\theta}_i = \log(\widehat{\mu}_i)$ and $b\left(\widehat{\theta}_i\right) = \widehat{\mu}_i$

   Next, the scale deviance by definition is,

$$D^* = \frac{1}{\phi} \times 2 \sum_{i=1}^{n} w_i \left[y_i\left(\widetilde{\theta}_i - \widehat{\theta}_i\right) - b\left(\widetilde{\theta}_i\right) + b\left(\widehat{\theta}_i\right)\right]$$

   where $\phi = w_i = 1$. Thus the Deviance (not scaled) is left,

$$\begin{aligned} D &= 2 \sum_{i=1}^{n} \left[y_i\left(\widetilde{\theta}_i - \widehat{\theta}_i\right) - b\left(\widetilde{\theta}_i\right) + b\left(\widehat{\theta}_i\right)\right] \\ &= 2 \sum_{i=1}^{n} \left[y_i\left(\log(\widetilde{\mu}_i) - \log(\widehat{\mu}_i)\right) - \widetilde{\mu}_i + \widehat{\mu}_i\right] \\ &= 2 \sum_{i=1}^{n} \left[y_i \log\left(\frac{\widetilde{\mu}_i}{\widehat{\mu}_i}\right) - \widetilde{\mu}_i + \widehat{\mu}_i\right] \\ &= 2 \sum_{i=1}^{n} \left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - y_i + \widehat{\mu}_i\right] \qquad \square \end{aligned}$$

(b) Show that the expression for the deviance simplifies to $2\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right)$, for the special case of the reduced model in part (a) with $g(\mu_i) = \log(\mu_i)$, and linear predictor that includes an intercept, that is, $\eta_i = \beta_1 + \sum_{j=2}^{p} x_{ij}\beta_j$, for $i = 1, \ldots, n$.

Solution

For the special case with $g(\mu_i) = \log(\mu_i)$, we have $\widehat{\mu_i} = g^{-1}\left(x_i^T \widehat{\beta}\right) = \exp\left(x_i^T \widehat{\beta}\right)$. Thus the deviance is now,

$$\begin{aligned} D &= 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - y_i + \widehat{\mu}_i \right] \\ &= 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right)\right] - 2\sum_{i=1}^{n}[y_i - \widehat{\mu}_i] \end{aligned}$$

Recall that our likelihood function is in EDF form. Then from lecture notes we know that the log likelihood is of the form:

$$f(y_i|\theta, \phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right]$$

Then the score function is:

$$U_j(\beta; y) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{a_i(\phi)\mathrm{Var}(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \tag{1}$$

For the log link, $\eta_i = \log(\mu_i)$, so $\mu_i = \exp(\eta_i)$, $a_i(\phi) = 1$, and $\frac{\partial \mu_i}{\partial \eta_i} = \exp(\eta_i) = \mu_i$. Since $\mathrm{Var}(y_i) = \mu_i$, equation (1) simplifies to (and set equal to 0):

$$\sum_{i=n}^{n} (y_i - \mu_i)\, x_{ij} = 0 \tag{2}$$

Since our log link contains an intercept term, equation (2) implied by that parameter is $\sum y_i = \sum \widehat{\mu}_i$. Thus,

$$\begin{aligned} D &= 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right)\right] - 2\sum_{i=1}^{n}[y_i - \widehat{\mu}_i] \\ &= 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right)\right] \qquad \square \end{aligned}$$

2. Let y_i, $i = 1, \ldots, n$, be realizations of independent random variables $Y_i \sim \mathrm{Ga}(\mu_i, \nu)$ distributions, with densities given by

$$f(y_i|\mu_i, \nu) = \frac{\left(\frac{\nu}{\mu_i}\right)^{\nu} y_i^{\nu-1} \exp\left(\frac{-\nu y_i}{\mu_i}\right)}{\Gamma(\nu)}, \quad y_i > 0, \nu > 0, \mu_i > 0$$

where $\Gamma(\nu) = \int_0^{\infty} t^{\nu-1} \exp(-t)\, dt$ is the gamma funciton.

(a) Express the gamma distribution as a member of the exponential dispersion family.

Solution

$$
\begin{aligned}
f(y_i|\mu_i,\nu) &= \exp\left[\nu\log\left(\frac{\nu}{\mu_i}\right)+(\nu-1)\log(y_i)-\frac{\nu y_i}{\mu_i}-\log(\Gamma(\nu))\right] \\
&= \exp\left[\frac{\log\left(\frac{\nu}{\mu_i}\right)}{\nu^{-1}}+(\nu-1)\log(y_i)-\frac{y_i\mu_i^{-1}}{\nu^{-1}}-\log(\Gamma(\nu))\right] \\
&= \exp\left[\frac{-\left(y_i\mu_i^{-1}-\log\left(\frac{\nu}{\mu_i}\right)\right)}{\nu^{-1}}+(\nu-1)\log(y_i)-\log(\Gamma(\nu))\right]
\end{aligned}
$$

which is part of the EDF,

$$f(y_i|\theta,\phi)=\exp\left[\frac{y_i\theta_i-b(\theta_i)}{a_i(\phi)}+c(y_i,\phi)\right]$$

where $y_i=y_i$, $\theta_i=\mu_i^{-1}$, $b(\theta_i)=\log\left(\frac{\nu}{\mu_i}\right)$, $a_i(\phi)=-\nu^{-1}$, $c(y_i,\phi)=(\nu-1)\log(y_i)-\log(\Gamma(\nu))$ □

(b) Obtain the scaled deviance and the deviance for the comparison of the full model, which incudes a different μ_i for each y_i, with a gamma GLM based on link function $g(\mu_i)=x_i^T\beta$, where $\beta=\left(\beta_1,\ldots,\beta_p\right)$ (with $p>n$) is the vector of regression coefficients corresponding to a set of p covariates.

Solution

The following are the MLEs under,

Full Model:$\widetilde{\mu}_i=y_i\Longrightarrow\widetilde{\theta}_i=\widetilde{\mu}_i^{-1}$ and $b\left(\widetilde{\theta}_i\right)=\log(\nu/\widetilde{\mu}_i)=\log(\nu/y_i)$
Reduced Model: $\widehat{\mu}_i=g\left(x_i^T\widehat{\beta}\right)$. Where $\widehat{\beta}$ denotes the mle of β. Thus $\widehat{\theta}_i=\widehat{\mu}_i^{-1}$ and $b\left(\widehat{\theta}_i\right)=\log(\nu/\widehat{\mu}_i)$

Then the scaled deviance is, with $w_i=1$,

$$
\begin{aligned}
D^* &= \frac{1}{\phi}\times 2\sum_{i=1}^{n} w_i\left[y_i\left(\widetilde{\theta}_i-\widehat{\theta}_i\right)-b\left(\widetilde{\theta}_i\right)+b\left(\widehat{\theta}_i\right)\right] \\
&= \frac{1}{\nu^{-1}}\times 2\sum_{i=1}^{n}\left[y_i\left(\widetilde{\mu}_i^{-1}-\widehat{\mu}_i^{-1}\right)-\log\left(\frac{\nu}{\widetilde{\mu}_i}\right)+\log\left(\frac{\nu}{\widehat{\mu}_i}\right)\right] \\
&= \nu\times 2\sum_{i=1}^{n}\left[y_i\left(\widetilde{\mu}_i^{-1}-\widehat{\mu}_i^{-1}\right)-\log\left(\frac{\widetilde{\mu}_i^{-1}}{\nu^{-1}}\right)+\log\left(\frac{\widehat{\mu}_i^{-1}}{\nu^{-1}}\right)\right] \\
&= \nu\times 2\sum_{i=1}^{n}\left[y_i\left(\widetilde{\mu}_i^{-1}-\widehat{\mu}_i^{-1}\right)-\log\left(\widetilde{\mu}_i^{-1}\right)+\log\left(\nu^{-1}\right)+\log\left(\widehat{\mu}_i^{-1}\right)-\log\left(\nu^{-1}\right)\right] \\
&= \nu\times 2\sum_{i=1}^{n}\left[y_i\left(\widetilde{\mu}_i^{-1}-\widehat{\mu}_i^{-1}\right)-\log\left(\widetilde{\mu}_i^{-1}\right)+\log\left(\widehat{\mu}_i^{-1}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
D^* &= \nu \times 2 \sum_{i=1}^{n} \left[y_i \left(\widetilde{\mu}_i^{-1} - \widehat{\mu}_i^{-1} \right) + \log \left(\frac{\widehat{\mu}_i^{-1}}{\widetilde{\mu}_i^{-1}} \right) \right] \\
&= \nu \times 2 \sum_{i=1}^{n} \left[y_i \left(\widetilde{\mu}_i^{-1} - \widehat{\mu}_i^{-1} \right) + \log \left(\frac{\widetilde{\mu}_i}{\widehat{\mu}_i} \right) \right] \\
&= \nu \times 2 \sum_{i=1}^{n} \left[y_i \left(y_i^{-1} - \widehat{\mu}_i^{-1} \right) + \log \left(\frac{y_i}{\widehat{\mu}_i} \right) \right] \\
&= \nu \times 2 \sum_{i=1}^{n} \left[1 - \frac{y_i}{\widehat{\mu}_i} + \log \left(\frac{y_i}{\widehat{\mu}_i} \right) \right]
\end{aligned}
$$

where the Deviance is, $D = 2 \sum_{i=1}^{n} \left[1 - \left(y_i / \widehat{\mu}_i^{-1} \right) + \log \left(y_i / \widehat{\mu}_i^{-1} \right) \right]$ □

3. Consider the data set from: http://www.stat.columbia.edu/~gelman/book/data/fabric.asc , on the incidence of faults in the manufacturing of rolls of fabric. The first column contains the length of each roll (the covariates with values x_i), and the second contains the number of faults (the response with means μ_i).

   (a) Use R to fit a Poisson GLM, with logarithmic link,

   $$\log(\mu_i) = \beta_1 + \beta_2 x_i \tag{3}$$

   to explain the number of faults in terms of length of roll.

   Solution

```
# Call:
# glm(formula = faults ~ length, family = "poisson", data = fabric)
#
# Deviance Residuals:
#      Min        1Q    Median        3Q       Max
# -2.74127  -1.13312  -0.03904   0.66179   3.07446
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept) 0.9717506  0.2124693   4.574 4.79e-06 ***
# length      0.0019297  0.0003063   6.300 2.97e-10 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for poisson family taken to be 1)
#
#     Null deviance: 103.714  on 31  degrees of freedom
# Residual deviance:  61.758  on 30  degrees of freedom
# AIC: 189.06
#
# Number of Fisher Scoring iterations: 4
```

   From the summary above, increasing the fabric length by 1 unit multiplies the mean number of faults by $\exp\left(\widehat{\beta}_2\right) = \exp(0.0019297) = 1.001932$. Note that the dispersion parameter was taken to be 1 and that the residual deviance is greater than the degrees of freedom. We will discuss further in part (b). □

(b) Use the quasipoisson "family" in R to fit the regression model for the response mean in (3) using the quasi-likelihood estimation method, which allows for a dispersion parameter in the response variance function. Discuss the results.

Solution

```
# Call:
# glm(formula = faults ~ length, family = "quasipoisson", data = fabric)
#
# Deviance Residuals:
#      Min        1Q    Median        3Q       Max
# -2.74127  -1.13312  -0.03904   0.66179   3.07446
#
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) 0.9717506  0.3095033   3.140 0.003781 **
# length      0.0019297  0.0004462   4.325 0.000155 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for quasipoisson family taken to be 2.121965)
#
#     Null deviance: 103.714  on 31  degrees of freedom
# Residual deviance:  61.758  on 30  degrees of freedom
# AIC: NA
#
# Number of Fisher Scoring iterations: 4
```

From the summary of the quasi-poisson glm above, the estiamtes of $\beta = (\beta_1, \beta_2)$ and the residual deviance are the same as those shown in the Poisson GLM.

Recall in the Poisson GLM, we saw that the residual deviance was greater than the degrees of freedom, so the Poisson GLM may have overdispersion. Running a quasi-poisson model fits an extra dispersion parameter but results in the same residual deviance. This implies that there is extra variance not accounted for by the model or by the error structure.

In the Poisson GLM the dispersion parameter was forced to be $\widehat{\phi} = 1$ whereas in the Quasi-Possion GLM the dispersion parameter was estimated to be $\widetilde{\phi} = 2.121965$. Since $\widetilde{\phi} > 1$, it turns out the conditional variance is larger than the conditional mean. Thus, we have over dispersion. □

(c) Derive point estimates and asymptotic interval estimates for the linear predictor, $\eta_0 = \beta_1 + \beta_2 x_i$, at a new value x_0 for length of roll, under the standard (likelihood) estimation method from part (a), and the quas-likelihood estimation method from part (b). Evaluate the point and interval estimates at $x_0 = 500$ and $x_0 = 995$. (Under both cases, the asymptotic bivariate normality of $\left(\widehat{\beta}_1, \widehat{\beta}_2\right)$ to obtain the asymptotic distribution of $\widehat{\eta}_0 = \widehat{\beta}_1 + \widehat{\beta}_2 x_i$)

Solution

Poisson Likelihood:
$\widehat{\eta}_0 = x_0^T \widehat{\beta}, \quad E(\widehat{\eta}_0) = E\left(x_0^T \widehat{\beta}\right) = x_0^T \widehat{\beta}, \quad Var(\widehat{\eta}_0) = Var\left(x_0^T \widehat{\beta}\right) = x_0^T Var\left(\widehat{\beta}\right) x_0 = x_0^T J^{-}\left(\widehat{\beta}\right) x_0$

Therefore, a point estimate for η_0 is $\widehat{\eta}_0$, and the interval estimate is:

$$x_0^T \widehat{\beta} \pm z_{0.025} \sqrt{x_0^T J^- \left(\widehat{\beta}\right) x_0}$$

We calculate the inverse fisher matrix, $J^- \left(\widehat{\beta}, \widehat{\phi} = 1\right) = J^- \left(\widehat{\beta}\right)$, using the $vcov(\cdot)$ function in R which in general returns the variance-covariance matrix of the main parameters of a fitted model object.

For $x_0 = 500$:

- Point estimate for η_0 is $\widehat{\eta} = x_0^T \widehat{\beta} = 1.936624$
- Interval estimate is $(1.783435, 2.089812)$

For $x_0 = 995$:

- Point estimate for η_0 is $\widehat{\eta} = x_0^T \widehat{\beta} = 2.891849$
- Interval estimate is $(2.662676, 3.121021)$

Quasi-Poisson Likelihood:

$$\widetilde{\eta}_0 = x_0^T \widetilde{\beta}, \quad E\left(\widetilde{\eta}_0\right) = E\left(x_0^T \widetilde{\beta}\right) = x_0^T \widetilde{\beta}, \quad Var\left(\widetilde{\eta}_0\right) = Var\left(x_0^T \widetilde{\beta}\right) = x_0^T Var\left(\widetilde{\beta}\right) x_0 = x_0^T J^- \left(\widetilde{\beta}\right) x_0$$

Therefore a point estimate for η_0 is $\widetilde{\eta}$, and an interval estimate is:

$$x_0^T \widetilde{\beta} \pm z_{0.025} \sqrt{x_0^T J^- \left(\widetilde{\beta}\right) x_0}$$

We calculate the inverse fisher matrix, $J^- \left(\widetilde{\beta}, \widetilde{\phi}\right)$, using the $vcoc(\cdot)$ function in R which in general returns the variance-covariance matrix of the main parameters of a fitted model object.

For $x_0 = 500$:

- Point estimate for η_0 is $\widetilde{\eta} = x_0^T \widetilde{\beta} = 1.936624$
- Interval estimate is $(1.713475, 2.159773)$

For $x_0 = 995$:

- Point estimate for η_0 is $\widetilde{\eta} = x_0^T \widetilde{\beta} = 2.891849$
- Interval estimate is (2.5580143.225683) □

4. This problem deals with data collected as a the number of Ceriodaphnia organisms counted in a controlled environment in which reproduction is occuring among the organisms. The experimenter places into containers a varying concentration of particular component of jet fuel that impairs reproduction. It is anticipated that as the concentration of jet fuels grows, the number of organisms should decrease. The problem includes also a categorical covariate introduced through use of two different strains of the organisms.

   The data set is available from the course website: https://ams274-fall18-01.courses.soe.ucsc. edu/homework-assignments , where the first column includes the number of organisms, the second the concentration of jet fuel (in grams per liter), and the third strain of the organism (with covariate values 0 and 1).

   Build a Poisson GLM to study the effect of the covariates (jet fuel concentration and organism strain) on the number of Ceriodaphnia organisms. Use graphical explanatory data analysis to motivate possible choices for the link function and the linear predictor. Use classical measures of goodness-of-fit and mode comparison (deviancec,

AIC, and BIC), as well as Pearson and deviance residuals, to assess model fit and to compare different model formulations. Provide a plot of the estimated regression functions under your proposed model.

Solution

Graphical Explanatory Data Analysis
The only transformations for the link function and linear predictor are log and square root. The following scatterplots (Figures 1, 2) show the linear relationship between the number of ceriodaphnia and jet fuel with and/or without a transformation on the variables. On the scatterplots you can also see its corresponing correlation value, r. (Note: we used the `scatter.smooth(·)` function to plot the graphs. And we use the jitter function on each variable before plotting; this has the effect of adding a small amount of noise so that more points are visible on the graph).

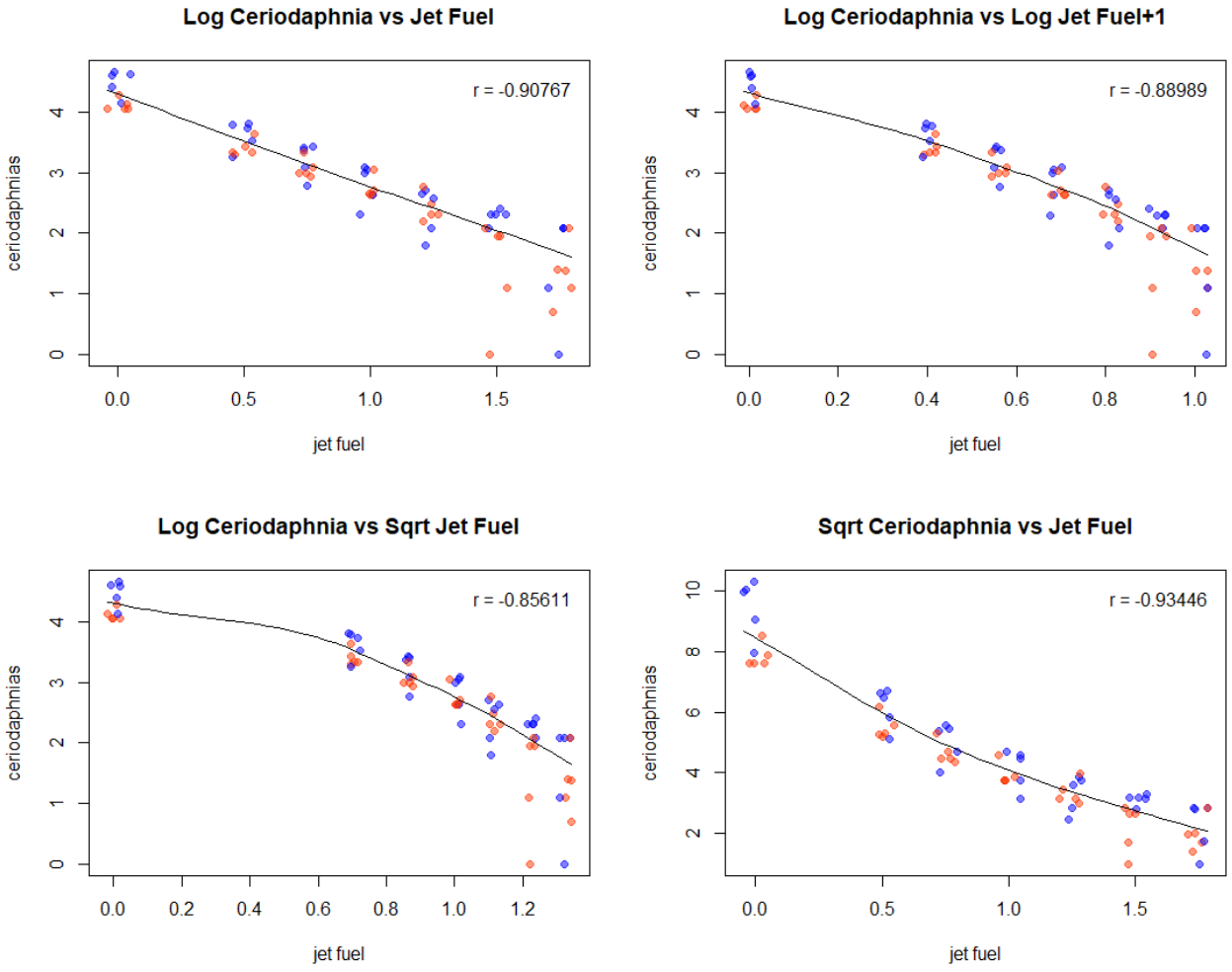

Figure 1:

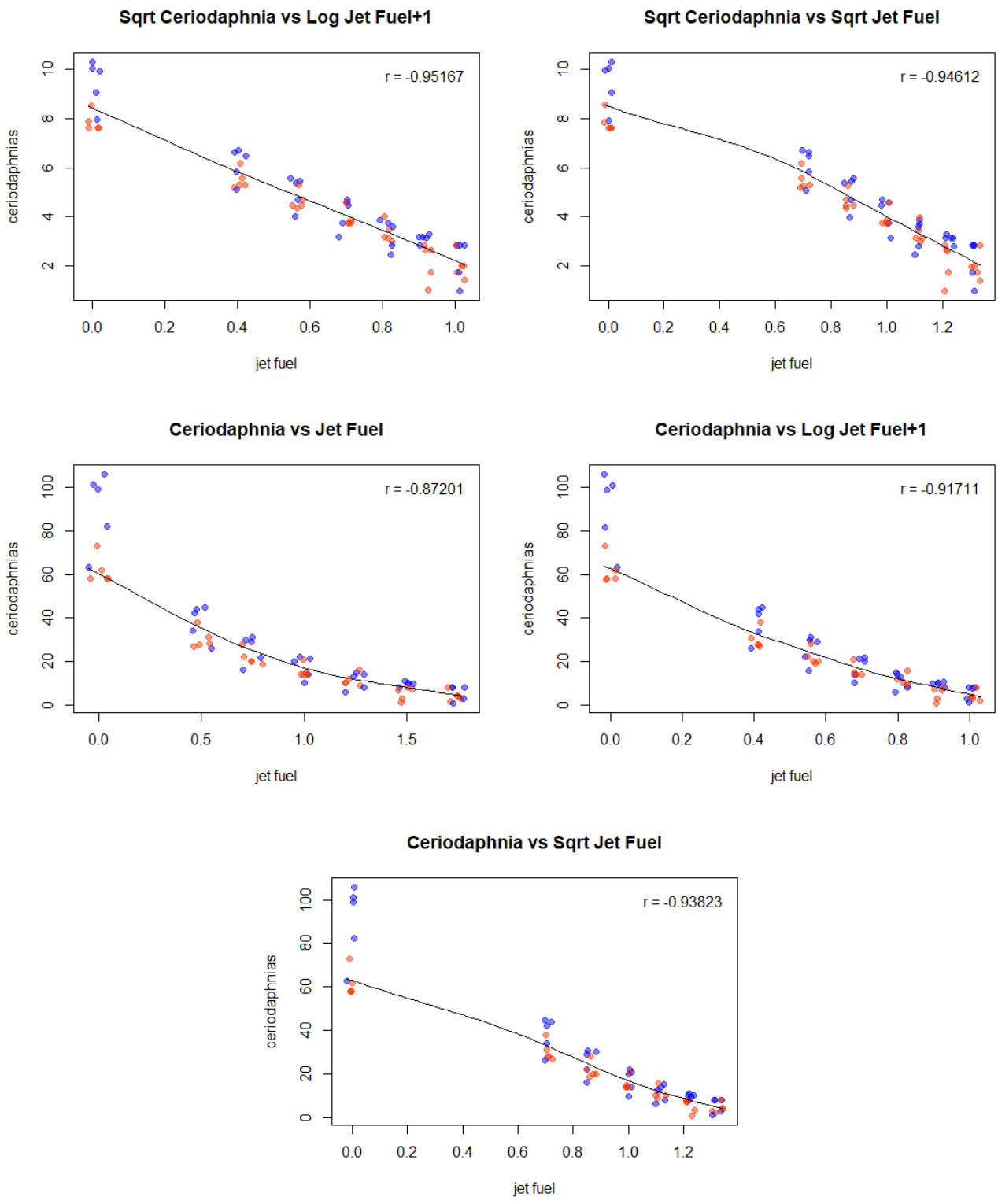
Sqrt Ceriodaphnia vs Log Jet Fuel+1
r = -0.95167
ceriodaphnias
jet fuel
Sqrt Ceriodaphnia vs Sqrt Jet Fuel
r = -0.94612
ceriodaphnias
jet fuel
Ceriodaphnia vs Jet Fuel
r = -0.87201
ceriodaphnias
jet fuel
Ceriodaphnia vs Log Jet Fuel+1
r = -0.91711
ceriodaphnias
jet fuel
Ceriodaphnia vs Sqrt Jet Fuel
r = -0.93823
ceriodaphnias
jet fuel

Figure 2:

In Figures 1 and 2, the blue points are the strained organisms (strain = 1) whereas the orange points are the unstrained organisms (strain = 0). The black curve is the fitted line regarding all strained and unstrained organisms, and we see that each graphs shows a negative correlation between the variables. Looking at each scatterplot carefuly, we can also fit two different slopes for each strain group but we will discuss this later.

Looking at the correlatin value for each scatterplot can help narrow our choice of transformation. The strongest correlations in Figure 1 are "Log Ceriodaphnia vs Jet Fuel" and "Sqrt Ceriodaphnia vs Jet Fuel". In Figure 2, "Sqrt Ceriodaphnia vs Log Jet Fuel + 1", "Sqrt Ceriodaphnia vs Sqrt Jet Fuel", "Ceriodaphnia vs log Jet Fuel + 1 ", and "Ceriodaphnia vs Sqrt Jet Fuel" all have a very strong negative correlation.

To narrow our choices further we will look at the contour density graph for each scatterplot with the strongest correlation values as previously mentioned.

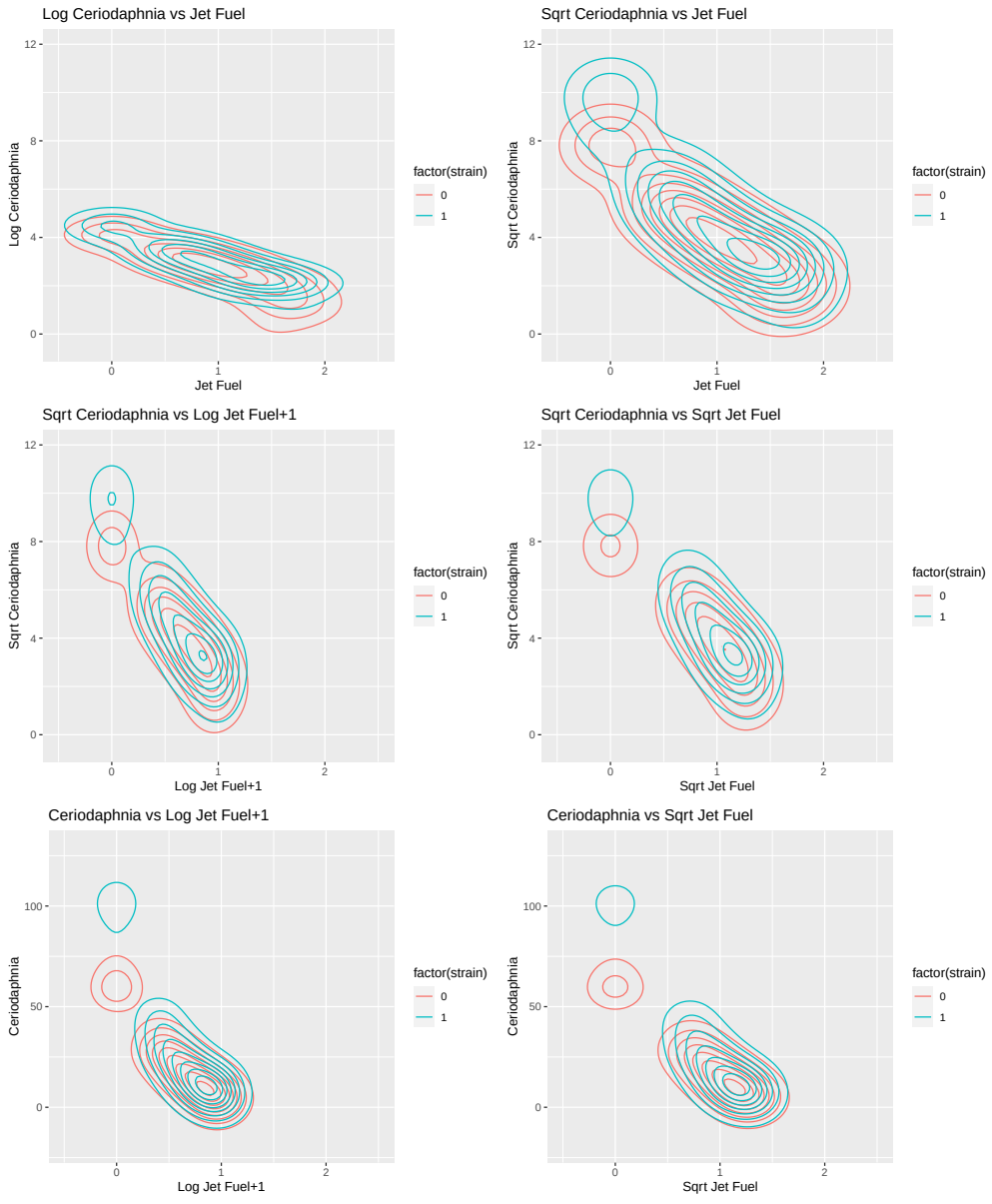

Figure 3:

In Figure 3, the bottom 4 density contour graphs have formed data clusters whereas "Sqrt Ceriodaphnia vs Jet Fuel" is tending to form a data cluster. Thus, "Log Ceriodaphnia vs Jet Fuel" is the best graph since it has not split into a cluster. Based on these Figures we assume models based on a log link function and untransformed predictors will be a good model fit. For model comparison purposes, we narrow our choices for a poisson glm model with a log link function and a poisson glm with a square root link function. We will also have a 3rd model, a poisson glm with log link function that includes an interaction term with jet fuel and strain type.
Poisson GLMs:

$$\text{Model 1:} \quad Y_i \sim \text{Pois}(\mu_i), \quad g(\mu_i) = \log(\mu_i) = \beta_1 + \beta_2 x_i, \quad \mu_i = \exp(\beta_1 + \beta_2 x_i)$$

$$\text{Model 2:} \quad Y_i \sim \text{Pois}(\widetilde{\mu}_i), \quad g(\widetilde{\mu}_i) = \sqrt{\widetilde{\mu}_i} = \widetilde{\beta}_1 + \widetilde{\beta}_2 x_i, \quad \widetilde{\mu}_i = (\beta_1 + \beta_2 x_i)^2$$

$$\text{Model 3:} \quad Y_i \sim \text{Pois}(\widehat{\mu}_i), \quad g(\widehat{\mu}_i) = \log(\widehat{\mu}_i) = \widehat{\beta}_1 + \widehat{\beta}_2 x_i + \widehat{\beta}_3 x_{1i} x_{2i}, \quad \widehat{\mu}_i = \exp\left(\widehat{\beta}_1 + \widehat{\beta}_2 x_i + \widehat{\beta}_3 x_{1i} x_{2i}\right)$$

Model Comparison: Now we use classical measures and criterions to assess goodness of fit.

Table 1: Model Comparison

	AIC	BIC	Chi.sq	Res.Dev
Model 1	415.95082	422.69631	79.83014	86.37646
Model 2	473.75208	480.49757	149.9367	144.1777
Model 3	416.16993	422.91542	77.61464	84.59557

From the Table 1, Model 1 is a good model fit according to the AIC and BIC. Model 3 is also a good model fit according to the Pearson Chi Square test and Residual Deviance. As assumed previously, a model based on a log link function and untransformed predictors will be a good model fit. From here, we will only focus on Models 1 and 3. Lets look at the Pearson and Deviance esidual plots of M1 and M3.

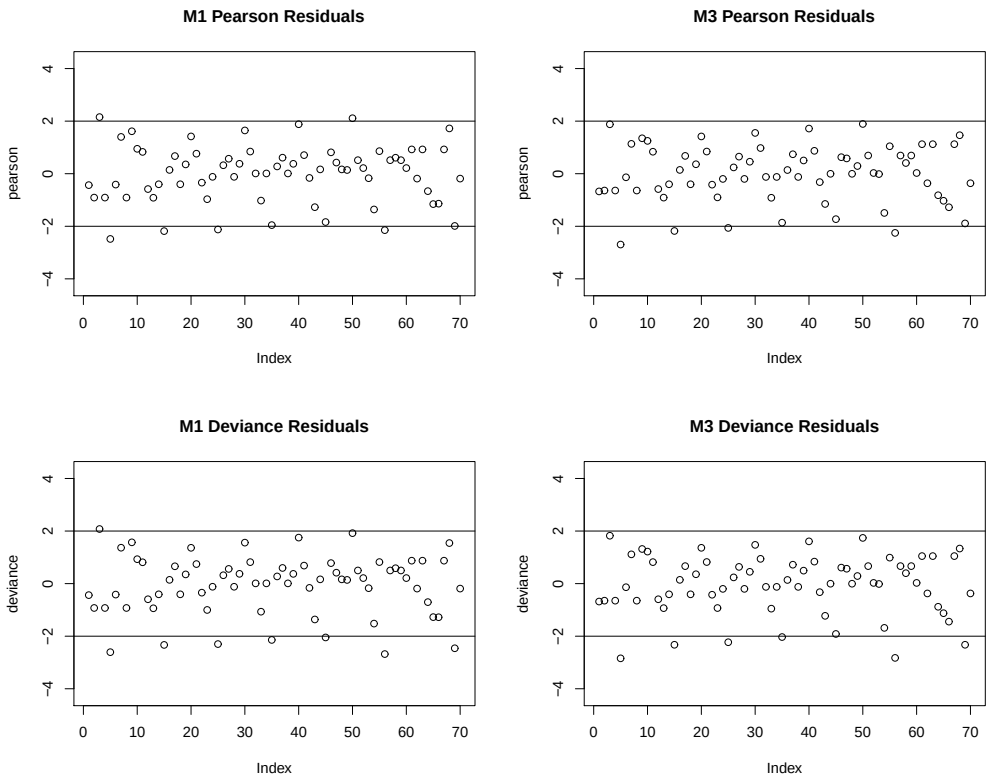

Figure 4:

Comparing the Pearson and Deviance Residuals between Model 1 and Model 3, we can see that Model 3 has just couple less outliers than Model 1 thus explaining why the Pearson Chi- Square test and the Residual Deviance test favors Model 3. Next, we will plot the estimated regression functions under M1 and M3. As shown below, we see no large difference between the two models. In conclusion, M1 and M3 seem to perform similarly under each criteria. But one can perhaps say that M1 is the better model since it is "simpler" by not including an interaction term.

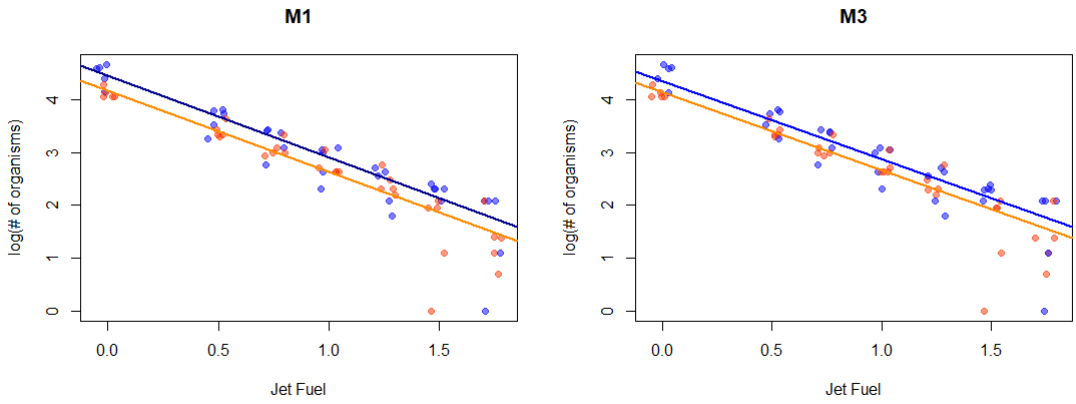

Figure 5: