Homework 3

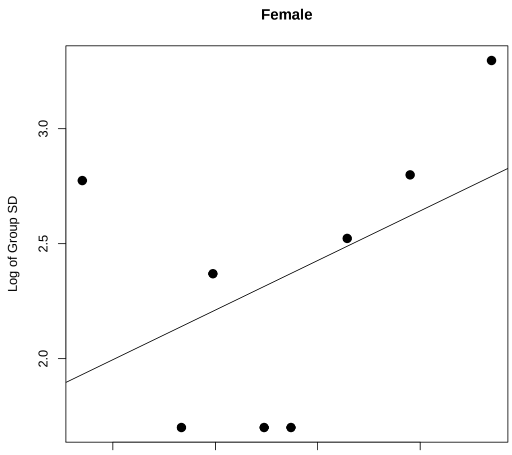
Steve Harms
STAT 520

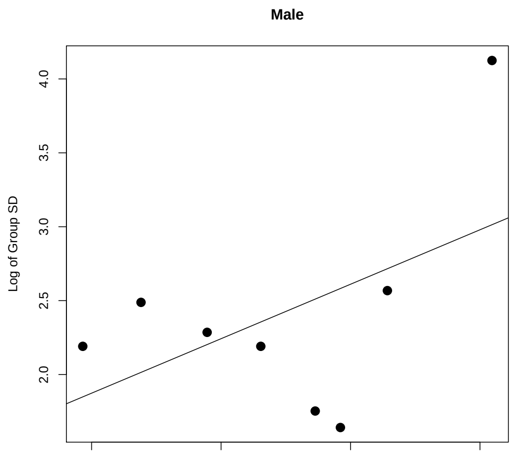
September 29, 2018

1 Part 1

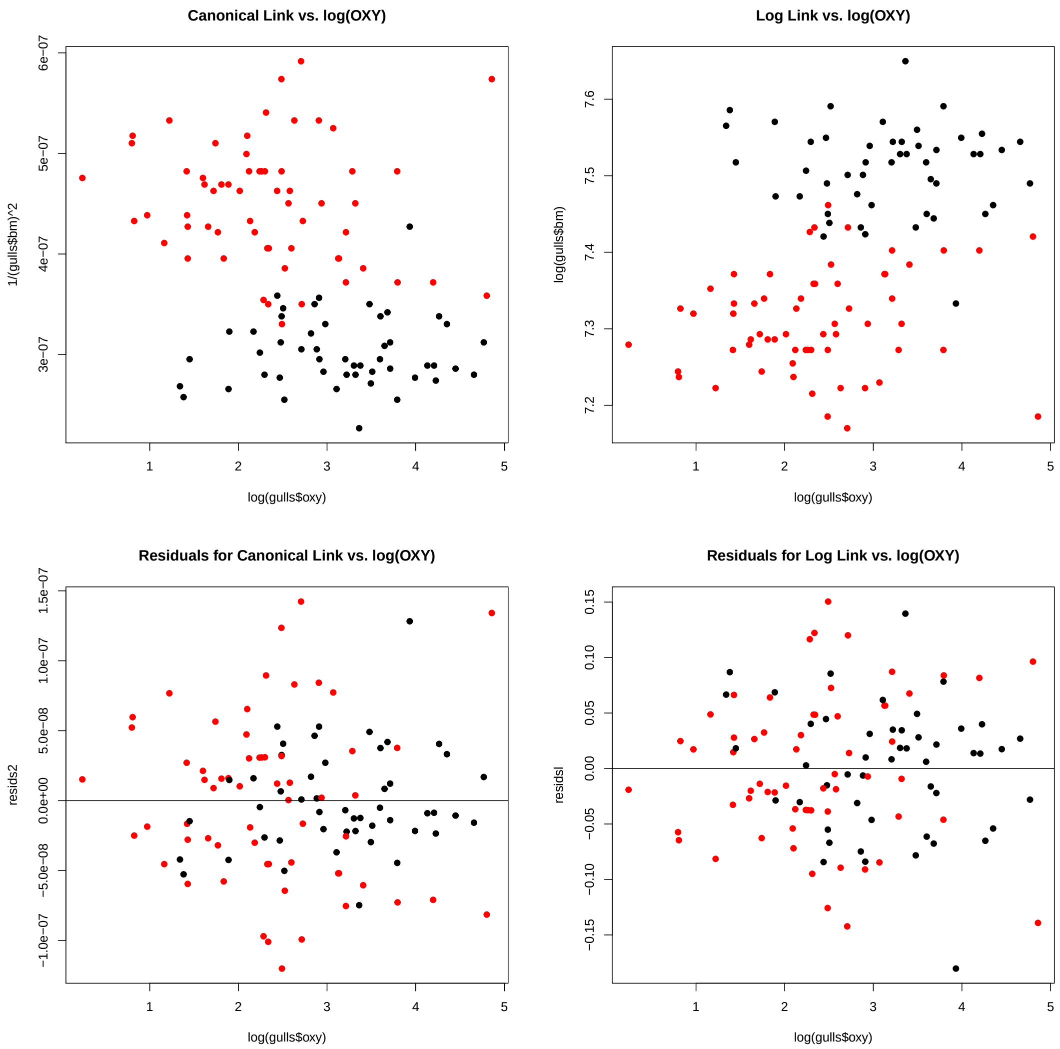
Let our random variable of interest be the body mass of the glaucous gulls, i.e. Y_i is the body mass of gull i for $i = 1, ..., 110$, where we have removed the 1 case with incomplete data. We note body mass cannot be negative, so that our sample space is $\Omega_Y = \{y : y \geq 0\}$. A number of covariates are available, however the data description and pairwise scatterplot matrix reveal that most are highly collinear and thus can be eliminated. We use only SEX (1=Male, 2=Female), and our covariate of interest OXY (oxychlordane concentration in blood [ng / g wet weight] ≥ 0) in our final analysis. Moreover, we can see that the relationshp for the relationship between Y and OXY is nonlinear, we use a log transformation for oxycholorodane concentration in our final model. The fit for log(OXY) is better than OXY.

Next, we pick a random component for our model. Our intuition tells us that $Y_i > 0$ implies either gamma or inverse gaussian distribution for our random component. We bin the data by deciles and then plot the log group SDs vs. the log group means to find the slope. In general, the fit of this model is poor, even after separating the groups into male and female and plotting separately. The Box-Cox diagnostic plots, despite poor fit, suggest that $\theta > 3$ and thus we model $Y \sim InvGaussian$ as our random component in our analysis. An alternative model with gamma random component or a normal random component with log link provide similar results, and our conclusions about the scientific question of interest do not change.





For the systematic component, we compare the canonical link function of $1/\mu_i^2$ and an alternative link function of $log(\mu_i)$. In both cases, the scatterplots and residual plot of the transformed Y_is vs. log(OXY) after accounting for sex looks reasonable. We use the log link function in the final analysis, although the canonical link provides similar results.



Now we fit our model, which is an Inverse-Gaussian GLM with log link function, with body mass as the response and sex and log(OXY) as the covariates. Results are provided below.

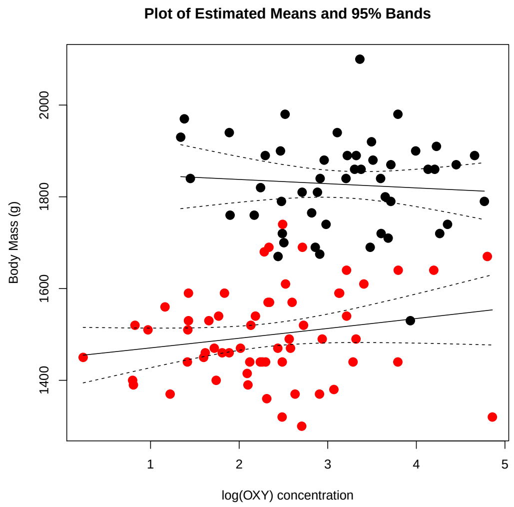
Overall, the fit is poor and most of the variation in body mass can be explained by sex. We can see that the coefficient for log(OXY) is not significantly different from 0 after accounting for sex, and thus we can conclude at a 95% level that the oxychlordane content does NOT have an effect on body mass in glaucous gulls.One important note is that we have not accounted for years captured.

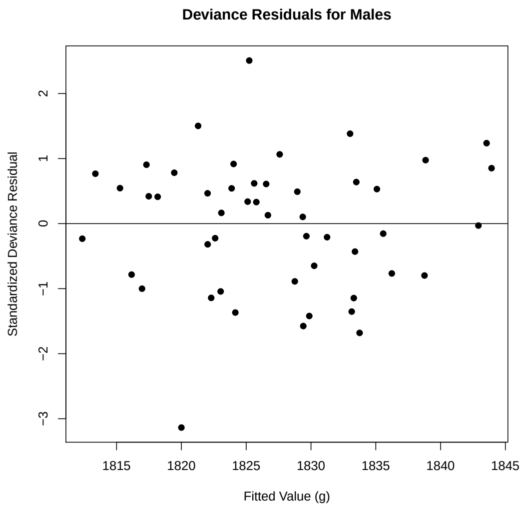
The plots of standardized deviance residuals show a nearly constant variance across fitted values, which in addition to a goodness-of-fit p-value of 0.54 indicates that our Inverse-Gaussian random component is justified in this case. We also provide a summary of results for the model with Gamma random component to show that the estimates, intervals and scaled deviances are all similar. We cannot compare likelihoods because our model for the mean is not the same.

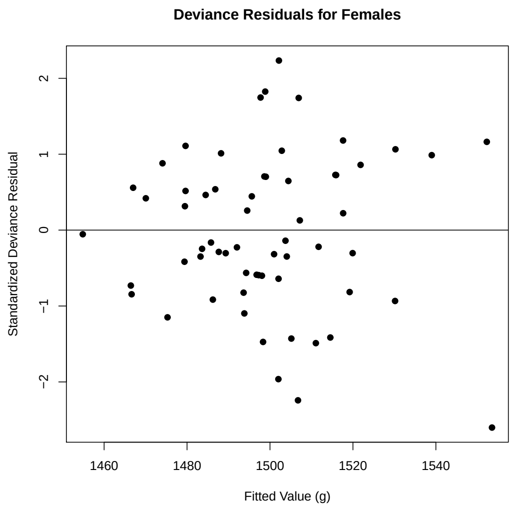
Model Estimation Summary - INVERSE GAUSSIAN random component		
Coefficient	**Estimate**	**95% Wald Interval**
Intercept	7.488	$(7.44197, 7.53399)$
log (OXY)	0.007	$(-0.0062, 0.0205)$
Sex (F)	-0.192	$(-0.2188, -0.1658)$
Dispersion $= 2.444542 \text{ x } 10^{-6}$		
Unscaled Deviance $= 0.0002635328$		
Scaled Deviance $= 107.8045$		
χ^2 p-value $= .5399963$ for 107 d.f.		
Maximized Log-Likelihood $= -665.0813$		

Model Estimation Summary - GAMMA random component		
Coefficient	**Estimate**	**95% Wald Interval**
Intercept	7.4907	$(7.4518, 7.5410)$
log (OXY)	0.0063	$(-0.00875, 0.01766)$
Sex (F)	-0.1929	$(-0.2199, -0.1687)$
Dispersion $= 2.444542$ x 10^{-6}		
Unscaled Deviance $= 0.42231$		
Scaled Deviance $= 107.3621$		
χ^2 p-value $= 0.528$ for 107 d.f.		
Maximized Log-Likelihood $= -664.0467$		

Plots of Means and Deviance Residuals for the INVERSE GAUSSIAN model - separated by the factor sex:



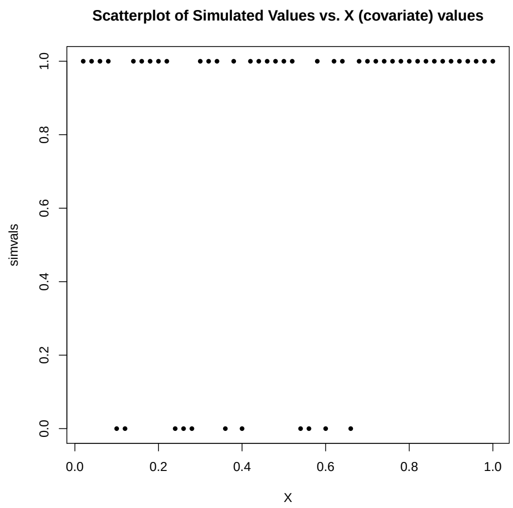




2 Part II

2.1

[Code is provided at the end]. First, a scatterplot of the data for our simulated model with c-log-log link function:

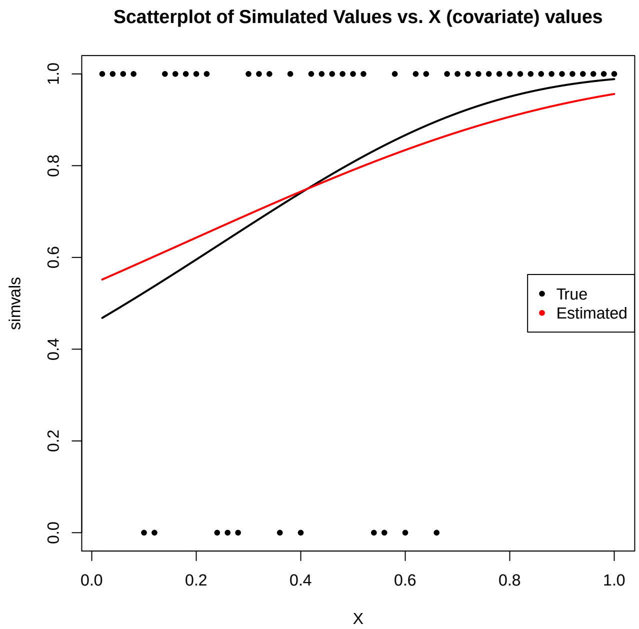


2.2

Now, results of model estimation:

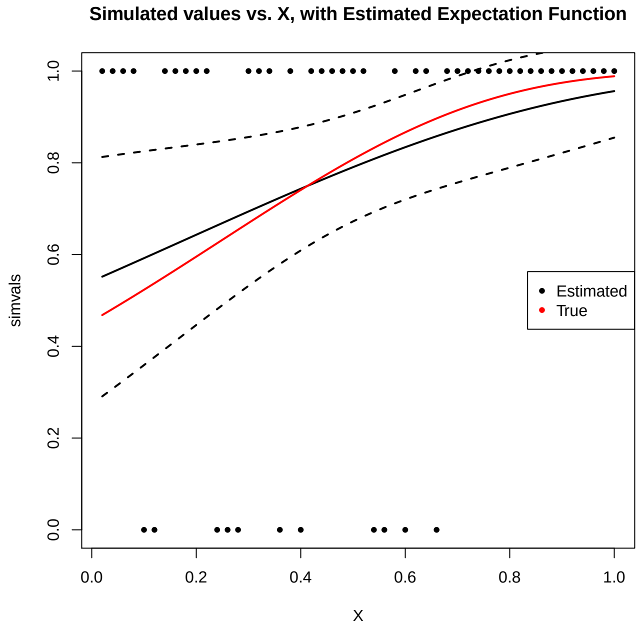
Regression Coefficients		
Coefficient	**Estimate (Std. Error)**	**95% Wald Interval**
Intercept	-0.2478 (0.3817)	$(-0.9959, 0.5004)$
Covariate (X)	1.3889 (0.6646)	$(0.0862, 2.6915)$
Dispersion = 1		
Unscaled Deviance = 48.799		
Scaled Deviance = 48.799		
p-value = .559284 for 48 d.f.		
Maximized Log-Likelihood = -24.3997		

Our scatterplot with response curves added in:



2.3

Finally, a scatterplot with our 95% Confidence Bands overlaid:



3 Part III

We return to Part 1 of Homework 2, which was concerned with the effects of silver iodide injections on rainfall in clouds. We formulated the model as a gamma distribution:
For cloud $i \in \{1, ..., 26\}$ in group $j \in \{1, 2\}$ (where $j = 1$ is the control), we define $Y_{ij} \sim Gamma\{\alpha_j, \beta_j\}$, where the sample space Ω_y is $[0, \infty)$. For our first model, we consider no shared gamma parameters between groups, so we are concerned with $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2)$, with parameter space $\Theta : \{\alpha_j > 0, \beta_j > 0\}$. We can write the distribution for Y_{ij} in exponential dispersion family form as:

$$f(y_{ij}|\theta) = exp\{\phi_j\{y_{ij}\theta_j - b(\theta_j)\} + c(y_{ij}, \phi_j)\} \tag{1}$$

where

$\phi_j = \alpha_j, \qquad \theta_j = \frac{-\beta_j}{\alpha_j},$
$b(\theta_j) = log\left(\frac{-1}{\theta_j}\right), \qquad c(y, \phi_j) = -log(\Gamma(\alpha_j)) - log(y_{ij})$

In homework 2, we considered a common scale parameter ($\alpha_1 = \alpha_2$). Now, we consider a common shape parameter: $\beta_1 = \beta_2$. As formulated in model (1), this corresponds to saying that $\phi_1(\theta_1) = \alpha_1(-\beta_1/\alpha_1) = -\beta_1 = -\beta_2 = \alpha_2(-\beta_2/\alpha_2) = \phi_2(\theta_2)$. That is, our two groups do not share a common mean or dispersion parameter in (1), but $\underline{\phi_1\theta_1 = \phi_2\theta_2}$.

With this, we are now modeling our random variables (rainfall from cloud i, group j) as $Y_{ij} \sim Gamma\{\alpha_j, \beta\}$. We can use silver iodide injection (0 if injected, 1 if not) as a covariate. Our scientific question of interest is still related to the mean after accounting for silver iodide injection. We are still interested in the group means, which are still given in this form by $\mu_j = -\frac{1}{\theta_j}$ and variance of group j given by $V_j = \frac{\phi_j}{\theta_j^2}$. In homework 2 we considered a model with shared dispersion parameter ϕ, here the dispersion parameter is no longer shared and we thus need to pay more attention to group variances as well.