Final Exam. Biology 7932. Fall term 2009. Due date 7 December 2009

The generalized linear model, which was developed in the last decades of the 20th century, has come into wider use in the current century. The advantages of the generalized linear model, compared to 20th century analysis within the framework of normal errors and the identity link, are better estimates of Type I error, better estimates of parameters (means, slopes, odds, and odd ratios), and better estimates of standard errors.

Evaluate the degree of improvement by re-analyzing data from the following sources:
--published examples from textbooks
–published examples from the refereed literature where the data can be obtained
–your own data

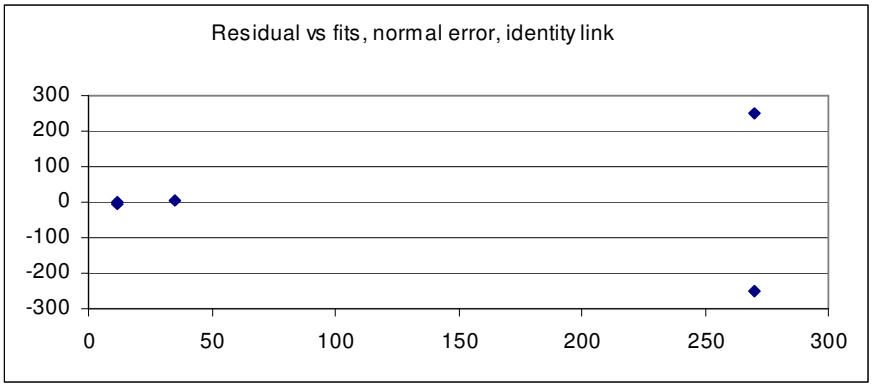
Each analysis (confined to 1-2 pages) should include:
-source of data, with citation for published results
-GLM model used or implied in the published analysis (or GLM if your own data).
-residual plot diagnosis of homogeneity and normality assumptions under
GLM (normal error, identity link).
-final revised model, where re-analysis is warranted by diagnosis of residuals
-comparison of Type I error via GLM and via GzLM (if revised model is used)
-comparison of parameter estimates and standard errors (if revised model is used)

The group report should be structured as a short review paper, such as in *Oikos.*
Points to cover in the results can include:
-frequency of re-analysis, based on evaluation of residual plots
-the degree of change in estimates of Type I error
-the degree of change in estimates of coefficients
-direction of change (if any) in standard errors
The discussion can include evaluation of GzLM vs GLM for AIC global model.

The report should include an appendix with each analysis, as above.

Analysis of algal biomass B in relation to tank volume V.
Data from Figure 5 in Chen et al. 1995 *Mar.-Ecol. Progr. Ser.* 155:1-15.
Model: $B = \mu$ + normal error $\mu = \alpha + \beta V$

Biomass	Vol
520	10
20	10
40	1
11	0.1
7	0.1

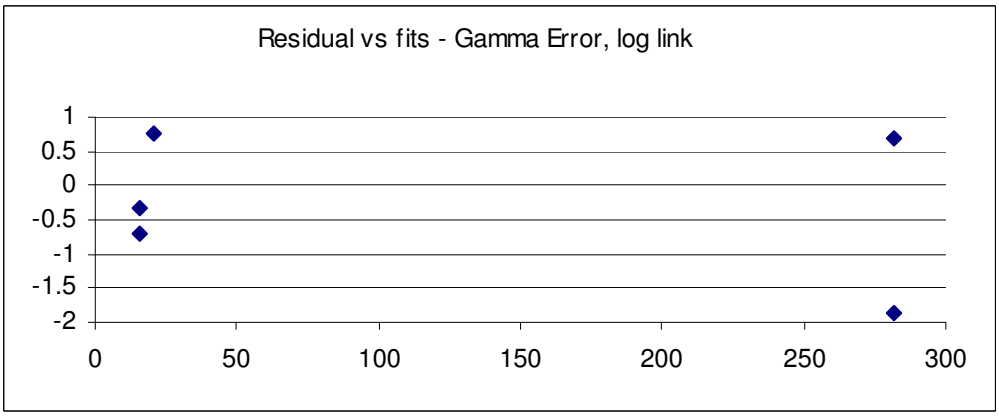


Coefficients:

Value		Std. Error	t value	Type I erro
Intercept	8.717	122.797	0.071	
Vol	26.152	19.367	1.350	0.270

(Dispersion Parameter for Gaussian family taken to be 41681.77)
Residual Deviance: 125045.3 on 3 degrees of freedom

Revised Model: $B = e^{\mu}$ + gamma error $\mu = \alpha + \beta V$



Coefficients:

	Value	Std. Error	t value	Type I erro
(Intercept)	2.731	0.588	4.641	
Vol	0.291	0.093	3.137	0.052

(Dispersion Parameter for Gamma family taken to be 0.9572717)
Residual Deviance: 5.091739 on 3 degrees of freedom

Summary:
GLM assumptions not met because residuals strongly heterogeneous.
Best revised model (most homogeneous errors) was Gamma error with log link
(exponential relation of Biomass to Volume)
Parameter estimates cannot be compared because of change in model structure.
Estimate of Type I error reduced from 27% to 5.2%.