Biology 7932 Assignment and Take Home Exam 15 November 2005
Instructor: David Schneider

The generalized linear model (Nelder and Wedderburn 1972, McCullagh and Nelder 1983) allows appropriate choice of an error model when data do not meet the assumptions (homogeneous residuals with fixed variance) that underlie regression, ANOVA, and other special cases of the general linear model. The generalized linear model has come into increasingly wide use in the last decade. A web search in mid-November 2005 turned up 96,300 hits ('generalized linear model') and 17,000 hits ('generalised linear model').

Assignment and Exam.

Attached is a list of terms that are central to understanding the generalized linear model approach to data analysis. Develop a list of definitions, working as follows.

1. Assignment. Develop a preliminary list of definition sources, for each term assigned to you (each term has been assigned to two people). This can consist of notes on material in a book (with page number) and material pasted from a website (with web address). List as your source either a text or a website. Early texts include McCullagh and Nelder (1983, 1987), Agresti (1990, 1996), Dobson (1990), Crawley (1993), Fahrmeir and Tutz (1994), Lindsey (1997). Several texts have appeared since 2000: Hoffman (2003), McCulloch and Searle (2001) and Myers et al (2002). The Sage series, which are usually clear and well written, has several titles (Dunteman and Ho 2005, Gill 2001, Hutcheson and Sofroniou 1999, Liao 1994).
A preliminary list is due from each person on 25 November.

2. Take Home Exam. Working as a group, develop a consensus list of definition(s) for each term. Due 5 December.

References - The Generalized Linear Model.

Agresti, A. 1990. *Categorical Data Analysis*. John Wiley, New York.
Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. John Wiley, New York.

Dobson, A.J. 1990 (1st ed) 2001 (2nd ed). *An Introduction to Generalized Linear Models*. Chapman Hall, London.
Intermediate level text that assumes several course in statistics (including linear models) including familiarity with calculus and linear algebra.

Crawley, M.J. 1993. *GLIM for Ecologists.* Blackwell, Oxford.
Biologically relevant text with clear explanation of GLIM, which was long the only available software. User friendly software has replaced GLIM and the book is now out of print.

Dunteman, G.H. and M.R. Ho. 2005. *An Introduction to Generalized Linear Models*. Sage Publications Inc. Thousand Oaks, CA, USA.

Fahrmeir, L and Tutz G. 1994 (1st ed) 2001 (2nd ed). *Multivariate Statistical Modeling Based on Generalized Linear Models*. Springer-Verlag, New York.
Covers many type of multivariate data and models that do not fit the standard multivariate Gaussian (fixed and normal error) theory.

Gill, J. 2001. *Generalized Linear Models : A Unified Approach*. Sage Publications Inc. Thousand Oaks, CA, USA.

Hoffmann, J.P. 2004. *Generalized Linear Models*. Allyn and Bacon.
Elementary introduction with many examples using the STATA statistical package.

Hutcheson, G.D. Sofroniou, N. 1999. *The Multivariate Social Scientist : Introductory Statistics Using Generalized Linear Models*. Sage Publications Inc. Thousand Oaks, CA, USA.

Liao, T.F. 1994. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Sage Publications Inc. Thousand Oaks, CA, USA.
Good on interpreting results and coefficients.

Lindsay, J.K. 1994. *Applying Generalized Linear Models*. Springer-Verlag, New York.
Suitable for Master's level students in biostatistics. Numerous examples,

McCullagh, P. and Nelder, J.A. 1983 (1st ed) 1987 (2nd ed). *Generalized Linear Models* (1st edition). Chapman and Hall, London.
The first text on the topic, from those who invented it, still authoritative. Requires math stats background.

McCulloch, C.E. S. R. Searle 2001. *Generalized, Linear, and Mixed Models.* John Wiley, NY

..recent and authoritative treatment of classical parametric models, starting with the general linear model and extending to generalized linear models, linear mixed models and finally to generalized linear mixed models. It also has applications to longitudinal data analysis and prediction problems. Heavy on theory and matrix algebra but not loaded with applications. Good for a graduate course in statistics especially for Ph.D. students.

Myers, R.H., Montgomery, D.C. and Vining, G.G. 2002. *Generalized Linear Models*. John Wiley, New York.
Requires substantial background in statistics. Numerous examples and applications.

Nelder, J.A. and Wedderburn, R.W.M. 1972. Generalized Linear Models. *J. Roy. Stat. Soc. A* 135: 370-384.
The paper that started it all.

List of terms (assignments were made randomly with the sample routine in Minitab).

NE = Ngaire Eskelin
SM = Stephen Mayor
AO = Anna Olafsdottir
SR = Sarah Ross
TY = Todd Yurich

Give a definition, an example, a table (where appropriate), and explanation (where appropriate).

1	SR	NE	Taylor's Power Law, definition + connection to GzLM
2	SM	AO	Scale parameter, dispersion parameter, variance function (+DScale in SAS, PScale in SAS)
3	AO	TY	Binomial, poisson, and non-poisson count data.
4	NE	SM	Odds and Odd ratios
5	TY	SR	Overdispersion (definition, sources, why it matters, how diagnosed, how addressed).
6	TY	SR	Loglinear models
7	SM	NE	Saturated model.
8	SR	TY	Goodness of fit (definition of each measure, choice of measure).
9	AO	SM	Analysis of deviance
10	NE	AO	Maximum Likelihood
11	TY	SM	Scaled residuals (why residuals are scaled, types, definition of each type)
12	SM	AO	Link (at least 10 types)
13	NE	SR	Error distributions (types, deviance for each type, canonical link for each type)
14	SR	TY	Parameter estimation (distinguish OLS, WLS, and iteratively reweighted IRLS).
15	AO	NE	Model diagnosis (linearity assumption, error assumptions, interpretation of diagnostic plots)