

## Model Based Statistics in Biology.

### Part V. The Generalized Linear Model.

#### Chapter 17 Poisson Response Variables

ReCap. Part I (Chapters 1,2,3,4), Part II (Ch 5, 6, 7)

ReCap Part III (Ch 9, 10, 11), Part IV (Ch13, 14)

17 Poisson Response Variables

17.1 Poisson Regression

17.2 Single Categorical Explanatory Variable  
(Log-linear Model)

17.3 Single Categorical Explanatory Variable  
(Sensitivity Analysis)

17.4 Two or More Categorical Explanatory Variables

17.5 Poisson ANCOVA

17.6 Model Revision

Ch17.xls

on chalk board

**ReCap** Part I (Chapters 1,2,3,4) Quantitative reasoning

**ReCap** Part II (Chapters 5,6,7) Hypothesis testing and estimation

**ReCap** (Ch 9, 10,11) The General Linear Model with a single explanatory variable.

**ReCap** (Ch 12,13,14) GLM with more than one explanatory variable

**ReCap** (Ch 15) GLM review

**ReCap** (Ch 16) The generalized linear model.

The General Linear Model (errors normal) is a special case of the Generalized Linear Model (errors normal or non-normal).

We use the improvement in fit, in the Analysis of Deviance table, to declare statistical decisions.

Today: Analysis of Poisson Response Variable with the  
Generalized Linear Model

#### Wrap-up.

Poisson response variables consist of counts in predefined categories, such as quadrats. The variance increases with the mean count. The result will be heterogeneous residuals when we use the general linear model. The solution is to use the generalized linear model, which allows Poisson response variables.

## Poisson Response Variables

Poisson response variables arise when we define a unit of analysis, and the resulting counts range from zero upward. In contrast, binomial counts result when each unit is scored on a nominal scale: yes/no, present/absent, etc.

Counts are readily analyzed within the framework of the generalized linear model. Here is a table showing the names of models, depending on the response variable and type of explanatory variable.

<u>Response Variable</u>	<u>Explanatory Variable</u>	<u>Canonical Link</u>	<u>Name of Error Model</u>
Continuous	Continuous	Identity	Normal
Continuous	Categorical	Identity	Normal
Continuous	Mixed	Identity	Normal
Success/Trial		Continuous	Logit
Success/Trial		Categorical	Logit
Success/Trial		Mixed	Logit
Count	Ratio scale	Log	Poisson
Count	Categorical	Log	Poisson
Count	Mixed	Log	Poisson
			Regression (GLM)
			ANOVA (GLM)
			ANCOVA (GLM)
			Binomial
			Logistic regression
			Binomial
			Logistic ANOVA
			Binomial
			Logistic ANCOVA
			Poisson regression
			Log-linear models
			Poisson ANCOVA

Generalized linear models have 4 parts: the response variable, the explanatory variables, the link function, and the error function.

The generalized linear model for a Poisson response (count) variable is:

$$Count = e^{(\eta)} + Poisson\ error$$

$$\eta = \sum \beta_i \cdot X_i = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots$$

The response variable appears on the left side of the equation,

The structural model  $\eta$  appears on the right side, as does the error model.

The structural model consists of one or more terms, one for each explanatory variable or interaction term in the model.

The response variable is linked to the structural model by a function.

For count data and Poisson error structure, a log link is typically used.

The link function is:  $Count = e^{\eta}$

This is the preferred or 'canonical' link function. We are free to use other links, however, such as the identity link

Poisson regression is used to estimate the rate of change in count with respect to some other variable. An example is number of deaths in some defined area, such as a province. If we know only the number of deaths, and do not know the population number, we have a Poisson count. The regression of the count against year gives us the change, if any, in death rate.

## Poisson Response Variables

Why do we use a log link with Poisson counts?

- ▶ The models that arise in analyzing counts compare proportions. For example, we might expect that toxins will have a multiplicative rather than additive effect on biological rate, such as number of offspring per unit time.
- ▶ The log link puts multiplicative effects on an additive scale, which is convenient for statistical analysis.
- ▶ Using a log link avoids the inconvenience of log transforming the response variable. We avoid having to add some arbitrary number to zero counts so that we can take the log.
- ▶ Using a log link avoids negative fitted values and negative confidence limits, which make no sense for counts.

A number of statistical packages will undertake analysis of Poisson response variables in a model based format. These include SAS, Stata, R/S+, Matlab, and SPSS.

References (Poisson response variables).

Agresti, A. 1996. Introduction to Categorical Data Analysis. NY: John Wiley and Sons.

Knoke, D., and P.J. Burke. Log-linear models. London: Sage Publications.

Lindsey, J.K. Applying Generalized Linear Models. NY: Springer Texts in Statistics.

Exact confidence interval for Poisson count. Tomas Aragon and Travis Porco.  
<http://www.medepi.org/epitools/rfunctions/cipois.html>

Confidence Intervals for the Mean of a Poisson Distribution. P.D. M. Macdonald.  
<http://www.math.mcmaster.ca/peter/s743/poissonalpha.html>