

## Model Based Statistics in Biology.

### Part V. The General Linear Model.

#### Chapter 17.6 Model Revision

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 12,13,14,15) GLM with more than one explanatory variable

**ReCap** (Ch 16,17)

Today: Model revision for Poisson regression

**Wrap-up.**

Example: biological assay.

Does a suspected toxin reduce the fecundity (brood size) of the assay organism *Ceriodaphnia dubia* ?

Assay data by Bailer and Oris taken from:

Nicholas Lange, Louise Ryan and Lynne Billard, David Brillinger, Loveday Conquest, Joel Greenhouse. 1994 *Case Studies in Biometry* John Wiley & Sons, Inc.

Preliminary computations: proportions and variances.

We begin by computing the mean and variance in fecundity (brood size) at each dose.

Brood 3 statistics					
Conc	0	80	160	235	310
Mean	13.90	14.80	11.50	6.70	0.00
Var	4.77	3.07	0.94	8.68	0.00
Var/Mean	0.34	0.21	0.08	1.30	

It is evident that the variance is not a fixed value.

Instead of assuming homogeneous variances (as with GLM), we are going to assume that the residuals arise from a Poisson distribution (GzLM with Poisson response).

## 1. Model

Verbal model.

Count is the number of offspring in the third brood of each *C. dubia*.

Conc is the concentration of the some substance (micrograms/L)

Graphical model

Plot of fecundity versus toxin concentration shows a curvilinear relation.

Response variable: Count

Explanatory variable: experimentally fixed concentration of substance Conc

Explanatory variable has multiplicative effect fecundity.

Write formal model as a multiplicative effect  $Count = e^{(\mu)} + :Poisson\ error$

$$\mu = \beta_o + \beta_{Conc} \cdot Conc$$

The link between the Count and the structural model  $\mu$  is:  $\ln(Count) = \mu$

This avoids negative predicted values of fecundity

## 2. Execute analysis.

Place data in model format:

Column labelled Count, with response variable # of animals

Column labelled Conc, with explanatory variable concentration

```
data A;
  input Conc Count;
cards;
0      10
0      15
0      17
0      15
0      15
0      15
0      15
0      15
0      12
0      11
0      14
80     16
80     16   etc
;
```

SAS command file

In a package with spreadsheet format, there will be two columns (variables) and 50 rows for this data set.

Code the GzLM model statement in statistical package

$$\text{Count} = e^{(\mu)} + : \text{Poisson error}$$

$$\mu = \beta_0 + \beta_{\text{Conc}} \cdot \text{Conc}$$

```
Proc Genmod;
  Model Count = Conc/
  Link=log dist=poisson type1 type3;
```

SAS command file

```
MTB >
```

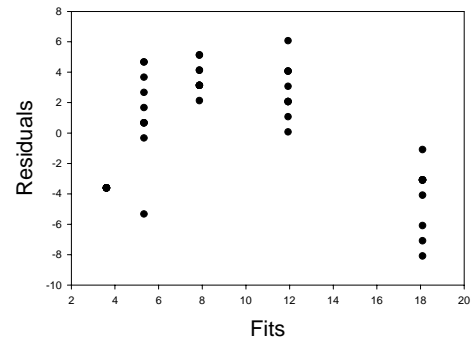
Minitab command lines

```
Click Stat
```

Minitab sequence to produce line commands

**3. Evaluate model**

Conclusion: Revise model.  
 Use concentration as a categorical variable.  
 Return to step 2.



**2. Execute analysis.**

```
Proc Genmod;
Class Conc;
Model Count = Conc/
Link=log dist=poisson type1 type3;
```

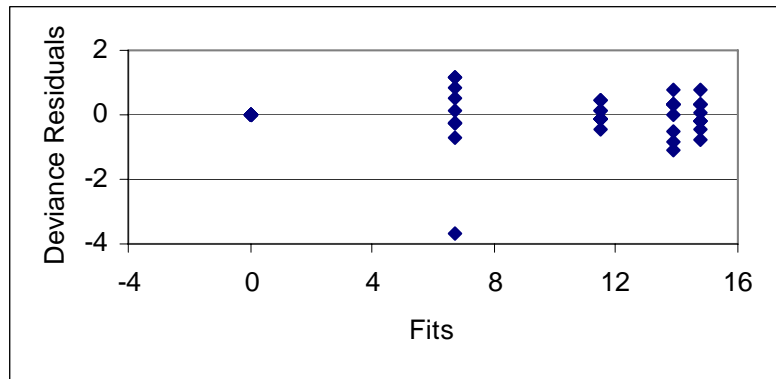
SAS command file

**3. Evaluate model**

a. No need to evaluate straight line assumption because no straight line was fit.

b. Residuals homogeneous ?  
 Yes, except for outlier.

Later, we will evaluate the sensitivity of the results to this outlier.



**4. State population and whether the sample is representative.**

Population.

All possible outcomes if the experiment were repeated on *C. dubia* with the same toxin and experimental protocol.

**5. Decide on mode of inference. Is hypothesis testing appropriate?**

Yes, as we wish to declare overall yes/no decision about effect of aflatoxin on brood size.

**6. State Ho/HA pair (some analyses may require several pairs).**

State test statistic, its distribution (t or F), and tolerance of Type I error.

HA:  $\beta_{Conc} \neq 0$       hence:  $Count = e^{(\beta_o + \beta_{Conc} \cdot Conc)} \neq constant$

H0:  $\beta_{Conc} = 0$       hence:  $Count = e^{(\beta_o)} = constant$

Tolerance for Type I error.       $\alpha = 5\%$

### 7. Calculate change in fit ( $\Delta G$ ) due to explanatory variables.

For the Generalized Linear Model, step 6 is modified: we calculate the change in deviance  $\Delta G$  rather than the SS for each term in the model.

Here is the output from the SPlus package.

	Df	Deviance	Resid.	Df	Resid.	Dev
NULL				49		269.7944
Conc	4	246.108		45		23.6864

SPlus output

The improvement in fit is  $\Delta G = 269.79 - 246.11 = 23.69$  on 4 df.

Calculate p-value from Chisquare distribution.

Is the change in fit  $\Delta G$  better than by chance ?

The p-value reported for  $\Delta G = 23.69$  is  $p = 0.000092$

The p-value is small, hence  $\Delta G$  is too large to be due to chance.

For generalized linear models, we compute a p-value on  $\Delta G$ , not on the deviance itself  $G$ .

### 8. Evaluate results if assumptions not met. Recompute estimates and p-values if necessary.

The p-value is far from the criterion of significance so the presence of the outlier is unlikely to lead to an incorrect decision. We check this judgement by running the analysis without the outlier.

	Df	Deviance	Resid.	Df	Resid.	Dev
NULL				48		250.8443
Conc	4	241.2769		44		9.5674

SPlus output

The decision was unchanged but there was a substantial change in the p-value. Our experience with p-values from F-ratios, which rarely change by a factor of 5, did not apply here.

### 9. Declare decision. $p = 0.048$ hence reject $H_0$ and accept $H_A$

The substance is toxic, causing a reduction of fecundity as concentration increases.

## 10. Analysis of parameters.

Here is the SAS output.

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-22.9644	0.1222	-23.2039	-22.7250	35333.4	<.0001
conc	0	25.5963	0.1487	25.3048	25.8878	29619.4	<.0001
conc	80	25.6590	0.1472	25.3704	25.9476	30365.4	<.0001
conc	160	25.4068	0.1537	25.1055	25.7080	27327.5	<.0001
conc	235	0	0.0000	24.9719	24.9719	.	.
conc	310	0	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000	.	.

SAS output

The estimates and confidence limits are converted back to predicted values via the link function. For example:

$$\exp(-22.9644 + 25.5963) = 13.9$$

This is the predicted value of brood size at zero concentration of the toxin.

Here are the expected or predicted values for each concentration.

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept	1	-22.96	0.1222	-23.20	-22.73
conc	0	13.90	0.1487	8.17	23.64
conc	80	14.80	0.1472	8.73	25.09
conc	160	11.50	0.1537	6.70	19.75
conc	235	7.44	0	5.86	9.46
conc	310	0	0	0	0

SAS output

The estimates are statistically indistinguishable at 0, 80, and 160 micrograms/L.

Brood size drops at 235 micrograms/L.

Brood size at 310 micrograms/L differs substantially from brood size at low concentrations of the toxin.