

Model Based Statistics in Biology.

Part V. The Generalized Linear Model.

Chapter 18 Binomial 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 (Ch 13, 14)

18 Binomial Response Variables

18.1 Logistic Regression (Dose-Response)

18.2 Single Factor. Prospective Analysis

18.3 Single Factor. Retrospective Analysis

18.4 Single Fixed Factor.

18.5 Single Explanatory Variable. Ordinal Scale.

18.6 Two Categorical Explanatory Variables

18.7 Logistic ANCOVA

Ch18.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)

Count data. Distinction between binomial and ratio scale response variables.

Reasons for generalized linear model for count data.

Variance increases with fitted values, not fixed.

Analysis of multiplicative effects (changes in proportion) without having to resort to log transform.

Advantages of model-based statistics.

Important concepts

Goodness of fit - Chisquare, G

Improvement in fit

ΔG

Analysis of deviance table

Today: Binomial response variables

Wrap-up.

Binomial data are analyzed within the framework of the generalized linear model.

The response variable is the odds, computed from the proportion of cases p .

Preliminary calculations: proportions, odds, and binomial variances.

In this course we will adopt a modeling approach that includes logistic regression as a special case. Logistic regression refers to response variables that are binomial. These arise when we define a unit of analysis, then each unit is scored on a nominal scale: yes/no, present/absent, *etc.*

Here is an example of a binomial response variable – a dose- response curve.

```
18  0  0
22  2  1
22  1  5
21  4 15
25 20 50
28 28 100

N Ntmr Dose

N = number of experimental animals fed
    aflatoxin B1, a suspected carcinogen.
Ntmr = number developing liver tumors
Dose = amount fed to animals (ppb)

Data from D.W. Gaylor (1987)
Linear_nonparametric upper limits for
low dose extrapolation.
American Statistical Association: Proceedings
of the Biopharmaceutical Section 63-66.
```

We begin by computing the proportion of animals that develop tumors, the variance at each dosage, and the odds of having a tumor at each dosage.

Dose	Cases	Cases	Proportion	Variance
ppb	N	w/tumors Ntmor	w/tumors p=Ntmor/N	N*p*q
0	18	0	0.0000	0.0000
1	22	2	0.0909	1.8182
5	22	1	0.0455	0.9545
15	21	4	0.1905	3.2381
50	25	20	0.8000	4.0000
100	28	28	1.0000	0.0000

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

Instead of assuming a fixed error and homogeneous variances (as with a GLM), we are going to assume that the variance depends on the proportion p . To estimate the parameters of the logistic regression equation we further assume that the residuals arise from a binomial distribution.

Preliminary calculations: Odds ratios from expected odds.

Next we compute odds ratios.

In this case we cannot calculate the odds at $p = 0\%$.

We could instead calculate observed odds relative to 1 ppb.

Thus, at 5 ppb, we odds ratio relative to 1 ppb is $OR = 0.0476/0.1 = 0.476:1$.

To further demonstrate the calculation of odds ratios we can use the expected value from the logistic regression. The rate of increase in odds is 0.09009 %/ppb, as estimated by binomial (logistic) regression.

Equivalently the rate of increase is 90.09 %/ppm.

The units make evident that this is a powerful carcinogen at concentrations on the order of parts per million. The logistic regression equation to calculate expected odds is:

$$E(\text{Odds}) = \exp(-3.03604 + 0.09009 \text{ ppb})$$

This gives us an expected odds of 0.048:1 at 0 ppb.

Dose ppb	Cases N	Cases w/tumors Ntumor	Proportion w/tumors $p = N_{\text{tumor}}/N$	Variance $N \cdot p \cdot q$	Odds		Odds Ratios OR
					Observed $p/(1-p)$	Expected	
0	18	0	0.0000	0.0000		0.0480	
1	22	2	0.0909	1.8182	0.1000	0.0526	1.094
5	22	1	0.0455	0.9545	0.0476	0.0754	1.57
15	21	4	0.1905	3.2381	0.2353	0.1855	3.86
50	25	20	0.8000	4.0000	4.0000	4.34	90.4
100	28	28	1.0000	0.0000		393	8176

Expected odds ratios can now be calculated relative to odds at 0 ppb.

Thus, at a dose of 1 ppb the odds ratio is $0.0526/0.048 = 1.094$

At 50 ppb the expected odds rise to 4.3:1.

The odds ratio rises to $OR = 4.34/0.048 = 90$.