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 Random 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 are 2 examples of binomial response variables.
1. Foraging success (captures per 100 attempts).
2. Dose- response curves.

```
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 B_1, 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 w/ tumors	Proportion w/tumors	Variance	Odds
ppb	N	Ntumor	p=Ntumor/N	N*p*q	p/(1-p)
0	18	0	0.0000	0.0000	
1	22	2	0.0909	1.8182	0.1000
5	22	1	0.0455	0.9545	0.0476
15	21	4	0.1905	3.2381	0.2353
50	25	20	0.8000	4.0000	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 depend on the proportion *p* and that the residuals arise from a binomial distribution (GzLM with binomial response).