Model Based Statistics in Biology.

Part III. The General Linear Model.

Chapter 9 Regression

ReCap.	Part I (Chapters 1,2,3,4)
ReCap	Part II (Ch 5, 6, 7)
ReCap	Part III
8	GLM. Component Concepts Generic recipe
9	Regression –ratio scale explanatory variable
9.1	Explanatory Variable Fixed by Experiment
9.2	Explanatory Variable Fixed into Classes
9.3	Explanatory Variable Measured with Error
9.4	Exponential Functions

on chalk board

```
Examples chosen for interest, historical importance, and to illustrate common
problems.

Other examples worked up as lectures, but not retained (1992-2002)
Simpson,Roe,Lewontin.1960. EX65,p222
  not broad interest (snake tails),  part-whole correlation, n large
Sokal&Rohlf.1995. Box 14.1
  non-normal residuals, n small, used in lab
Sokal&Rohlf.1995. Box 14.12
  residuals not normal, n small,  x-variable measured with error
Daniel.1995. Exercise 9.22
  broad interest, non-normal residuals, n small, x-variable with error
```

ReCap Part I (Chapters 1,2,3,4)
Quantitative reasoning: Example of scallops,
which combined models (what is the relation of scallop density to substrate?)
with statistics (how certain can we be?)
ReCap Part II (Chapters 5,6,7)
Data equations summarize pattern in data as a series of parameters (means, slopes).
Frequency distributions, a key concept in statistics, are used to quantify uncertainty.

Today: Introduction to the General Linear Model
Begin with brief introduction to component concepts in a generic recipe.
Then work through an example, using a generic recipe.

Wrap-up
Going to use the General Linear Model.
The GLM consists of familiar and new components.
Regression, a simple case of the GLM, presented as an example.

Table 8.3 Generic Recipe for Statistical Inference with the General Linear Model.

1. Construct model. Begin with verbal and graphical model.
    Distinguish response from explanatory variables
    Assign symbols, state units and type of measurement scale for each.
    Write out statistical model.
2. Execute model Place data in model format, code model statement.
    Compute fitted values from parameter estimates.
    Compute residuals and plot against fitted values.
3. Evaluate the model, using residuals.
    If straight line inappropriate, revise the model (back to step 1).
    If errors not homogeneous, consider using generalized linear model (step 1)
    If n small, evaluate assumptions for using chisquare, t, or F distribution.
        residuals homogeneous ? (residual versus fit plot)
        residuals independent ? (plot residuals versus residuals at lag 1)
        residuals normal ? (histogram of residuals, quantile or normal score plot)
    If not met, empirical distribution (by randomization) may be necessary
4. Partition df and SS according to model. Calculate likelihood ratio for omnibus model.
    State the full (null) and reduced (alternative) model pair
5. Choose mode of inference: evidentialist, frequentist, priorist. If priorist, see recipe.
    If sufficient evidence for omnibus model Step 6, otherwise step 10.
6. State test statistic, its distribution (*t* or *F*). If frequentist, fixed or categorical Type I error?
7. ANOVA: Table Source, SS, df, MS, F-ratio.
    Obtain Type I error (p-value) from distribution (*F* or t).
8. Recompute p-value if necessary.
    If assumptions not met compute better Type I error by randomization if:
        sample small ($n < 30$) and if Type I error near fixed α.
9. Report Type I error and statistical conclusion about model terms:
    Report either the ANOVA table or F-ratio (df1,df2)
    or t-statistics (df) and p-value (not α) for terms of interest.
10. Report science conclusions. Interpret parameters of biological interest (means, slopes)
    along with one measure of uncertainty (st. error, st. dev., or conf. intervals).
    Use appropriate distribution (step 8) to compute confidence limits.

Regression is used to estimate the relation of a response variable (the dependent variable) to an explanatory variable (the independent variable. Regression is a special case of the general linear model where the explanatory variable is on a ratio type of scale. The relation between the response and explanatory variable is expressed as a line or curve. The equation links one or more response variables to one or more explanatory variables, via parameters (means and slopes).

To complete a thorough statistical analysis we will use a generic recipe that applies to any general linear model. This recipe incorporates the Generic Recipe for Hypothesis testing. The pattern is stated as an equation; the summary statistic for hypothesis testing is the F or t-ratio.

The regression model has two parts—the structural model (the line or curve) and the error model. The error model will be normal error for the GLM. The overall approach will be to write the model, evaluate the model assumptions, then use the model to:

set up hypotheses concerning the relation of response to explanatory variables.
partition the degrees of freedom
partition the total sum of squares: $SS_{total} = (n-1) * Var(Y) = (n-1) * s^2$
calculate the omnibus likelihood ratio.
carry through from likelihood ratios to hypothesis testing where warranted.
set up the ANOVA table

We will be learning a procedure instead of a series of named tests.

Learning a procedure (riding a bicycle, using a computer, learning a script) is very different than recalling events or facts (episodic memory).

“ Episodic memory, we know, develops relatively late in childhood and is dependent on a complex brain system involving the hippocampi and medial temporal-lobe structures, the system that is compromised in severe amnesiacs. “

“The basis of procedural or implicit memory is less easy to define, but it certainly involves larger and more primitive parts of the brain–subcortical structures like the basal ganglia and cerebellum and their many connections to each other and to the cerebral cortex. The size and variety of these systems guarantee the robustness of procedural memory and the fact that, unlike episodic memory, procedural memory can remain largely intact even in the face of extensive damage to the hippocampi and medial temporal-lobe structures.”
Oliver Sacks. *The Abyss*. The New Yorker September 24, 2007.