.
Biology 4605/7220 - Model Based Statistics in Biology.
Generic Recipe for Data Analysis with the Generalized 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 structural model (explanatory variables)
2. Execute model   Place data in model format, code model statement.
                       Specify error structure and link function.
                       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, revise error structure and link (step 2)
4. State population and whether the sample is representative
5. Decide on mode of inference. Is hypothesis testing appropriate?
    If yes step 6, otherwise, skip to step 9.
6. State H_o/H_A pair (some analyses may require several pairs).
    State test statistic, its distribution (Chisquared), and tolerance of Type I error.
7. ANODEV: Partition df according to model.
                   Table Source,df, ΔG, p-value from Chisquared distribution.
8. As needed, evaluate sensitivity to deviations from assumptions, or compute p-values
    and estimates by randomization.
9. Declare decision about model terms: If $p < \alpha$ then reject H_o and accept H_A
                                                 If $p \geq \alpha$ then accept H_o and reject H_A
    Report conclusion with evidence: Either the ANODEV table or
        p-value (not α) for terms of interest.
10. Report and interpret parameters of biological interest (means, slopes)
    along with one measure of uncertainty (st. error, st. dev., or conf. intervals).
    Use appropriate distribution to compute confidence limits.

Notes on Generic Recipe

1. **Construct model**
    Define variables with symbols, type of measurement scale, and units if possible.
    Response variable is...   Explanatory variable is ...
    Graph response relative to explanatory variables.
    Write formal model (GLM) with graph sketched over individual terms.

2. **Execute model**. Place data in model format, one column per variable.
    Computer routines use model statement to estimate parameters and to
    partition df and SS according to model. Routines produce ANOVA table
    with p-values, fitted values, residuals, and plots of residuals.

3. **Evaluate model.**
    a. If regression used but straight line inappropriate (bowls or arches),
        revise the model (back to step 1).
    b. If errors expand (cones in residual plot) specify new error structure (step 2)

4. **State sample, population**. Statistical inference is restricted to the population for
        which the sample is representative.

5. **Hypothesis testing ?** Hypothesis testing not mandatory.
    Appropriate to binary (yes/no) research questions.
    Required in some contexts (e.g. journal policy, high cost of Type I and II errors).
    If needed, state H_o/H_A pair about terms in model.
    State test statistic (ΔG), its distribution (Chisquared), and tolerance of Type I
    error.

7. **ANODEV**:   GLM routines produce ANOVA table and Type I error from
                appropriate distribution (t or F).

8. **Evaluate tabled results.** Assess sensitivity of p-values and parameter estimates
    to deviations from assumptions of homogeneous residuals.   Re-compute
    estimates and p-values by randomization if needed.

9. **Declare decision about model terms.**
    If H_o accepted, decide whether Type II error matters. If so report an appropriate
     measure such as minimum detectable difference, minimum sample size.
    If interaction term significant, then decisions about component terms are not
    possible.
    Report conclusion with evidence.

10. **Examine parameters of biological interest**.
    For simple models, report parameters with confidence intervals.
    These are more informative (which hypotheses excluded?) than hypothesis
    testing.
    One-way ANOVA: Planned (*a priori*) comparisons are more informative than
        unplanned (*a posteriori)* comparisons.
    If interaction term significant report contrasts (*e.g.* compare slopes).