Lecture 17. Design of Ecological Field Studies

First Critique Papers Brief Discussion

Nelson and Marler (1994) paper

Scarf et al. (2006) paper

Design of Telemetry Studies Discussion

Study of Relationships

Regression Methods

Multivariate Analysis (Later Lecture)

Nelson and Marler (1994) Selection- Based Learning in Bird Song Development

•The study involved 10 hand reared white-crowned sparrows which were kept in sound isolation chambers. The birds were either played "tutor songs" or novel song types and studied to see what songs they then sang later on.

•Experimental unit- an individual bird

•Treatments(2)

Experimental –Tutor songs

Control- Novel songs

•The design was basically a simple paired design or randomised complete block with 5 pairs of control and treated birds.

•Factors ignored: There were that the birds were of 2 different subspecies and were born in 2 different years.

Nelson and Marler (1994) Selection- Based Learning in Bird Song Development

- The design was basically a simple paired design or randomised block with 5 pairs of control and treated birds.
- A very small study because of the expense I suppose.
- Analysis uses a sign test and a signed rank test in different places. These are appropriate for the design.
- However, in one other analysis they used the Mann Whitney U test which is a rank test for unpaired data where there are two treatments. This seems inappropriate to me.

Scarf et al. (2006) Predation Paper Discussion

Treatment Factors

Factor 1 -2 levels –Predators summer flounder and striped searobin

Factor 2 -3 levels –Prey were juvenile winter flounder, black sea bass and scup.

Factor 3 -3 levels-Habitats were bare sand, bare sand and shell, bare sand sponges

Treatment design- A 2x3x3 factorial with 18treats

Experimental unit- a feeding trial of 1 predator on one prey type in one habitat type

Scarf et al. (2006) Predation Paper Discussion

Experimental Unit- a feeding trial of 1 predator on one prey type in one habitat type

Experimental Design- Completely Random Design. Not clear how many reps were run. 167 trials all together so roughly 9 per treatment.

Analysis- Seems kind of piecemeal to me. No where do we see an overall analysis of variance table. At first I thought they were just treating the predators separately but that is not the case (See for example Fig 3).

Scarf et al. (2006) Overall ANOVA Table (if 9 reps)

Source	df
Model	17
Pred	1
Prey	2
Habitat	2
Pred*Prey	2
Pred*Habitat	2
Prey*Habitat	4
Pred*Prey*Habitat	4
Residual	144
Total	161

Design of Telemetry Studies: Joe Hightower Lecture

- I really appreciate Joe taking over while I was away.
- Telemetry studies are very widely used by many ecologists.
- Many points relevant whether you are in fisheries or not.
- I will just quickly make a few key points.

Design of Telemetry Studies: Nature of Data

- Spatial locations over time (hourly, daily, weekly, …, yearly)
- Mortality sensor sometimes.
- Other information on activity depending on the nature of the tag.
- There is often right censoring (tags lost) and also animals may be tagged at different points in time.

Design of Telemetry Studies: Many Purposes

- Many “Home Range” related objectives (daily, seasonal, annual movements and activity etc)
- Many Habitat Utilisation and Preferences related objectives
- Survival and Mortality Information (ST 506 will cover in some detail)

Design of Telemetry Studies: Types of Studies

- Descriptive (Measure Home range, survival rate etc on a population using telemetry)
- Correlational (Spawning Locations vs River Flow, Survival vs Age of animal)

Design of Telemetry Studies: Types of Studies

- Manipulative

- Before vs After No Controls-Quaker Neck Dam Study

- Ideally Before-After-Control-Impact

Design of Telemetry Studies: What is the Sampling Unit?

- A single relocation of an animal
- An animal
- A group of animals

Design of Telemetry Studies: Habitat Preference

- Need to have a good sample of telemetered relocations by habitat.
- Need to also have a random sample of habitats available to the animal
- Then one can judge if the animals are using the habitat proportional its availablity or not.

Design of Telemetry Studies: Logistic Regression

- **Another topic that Joe and Julie discussed was the use of logistic regression models**
- **Note that we will discuss logistic regression models in some detail in the next couple of lectures.**

Study of Relationships

- **May be a crucial analysis element in observational studies**
- **Remember that these explorations do not in any way imply causality. These are not experiments!**

Study of Relationships

- Generalised Regression Methods

  One variable of primary importance (dependent variable Y) and you would like to see you it changes with other variables(independent variables X's)

- Multivariate Analysis

  Multiple (X) variables that you would like to study jointly .

  No one variable of predominant Interest

Exploratory Study of Relationships

- Generalised Regression Methods

Single Explanatory Variable

Linear Regression and Covariance Methods

Nonlinear Regression Methods

Multiple Explanatory Variables

Multiple Linear Regression Methods

Logistic Regression Methods

Exploratory Study of Relationships

- Multivariate Analysis

Principal Component Analysis

Canonical Correlations Analysis

Discriminant Analysis

Cluster Analysis

Other Methods

Simple Linear Regression Method

$$Y_i = \mu\{Y \mid X) + \varepsilon_i$$

$$\mu\{Y \mid X) = \beta_0 + \beta_1 X$$

Y -Dependent Variable

X- independent variable

$\mu\{.\}$ –is the mean function and is conditional on the X’s.

β_0 is intercept and β_1 is slope of line. ε-independent Normal errors with constant variance is the standard form.

Simple Linear Regression: Least Squares

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Find β's which Minimise

$$\sum_{i=1}^{n} (Y_i - \beta_0 + \beta_1 X_i)^2$$

It is called least squares because we are minimizing the sum of the squared deviations

Simple Linear Regression

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

β_0 is intercept and β_1 is slope.

Exact inferences based on standard regression theory (ST 511). ANOVA Table, F Tests, T tests all exact.

Proc REG in SAS can be used.

In many cases there are concerns about linearity as an assumption.

Sometimes the variance may increase as X increases we can then use weighted regression methods.

Analysis of Covariance Method

- **Even though we talked about this earlier I thought that it was worth another look.**
- **Wanted to emphasize again that it is a combination of regression and anova methods**

Analysis of Covariance Method

$$Y_{ij} = \mu + \tau_i + \delta_j + \beta X + \varepsilon_{ij}$$

Y_{ij}-Dependent Variable

τ_i is the treatment effect and δ_j is the block effect

β is slope and is constant over all treatments

X_{ij}- independent continuous variable

This model mixes ANOVA and Regression. Goal to adjust for a covariate so one can test a "treatment" effect at a constant level of the covariate.

ANCOVA: Example

Knapp et al (2003). Effects of clearcutting and alternative silviculture practices on terrestrial salamander abundance. Conservation Biology

17, 752-762.

Randomised Complete Block Experiment

5 Study Blocks each containing 7 silviculture treatments randomly allocated to 2 ha plots

The year Before Cutting was used as a covariate in an ANCOVA. They considered 1-4 years after harvest.

Analysis of Covariance Example

Results show clearly that the covariate was very important

Results also show that after controlling for the covariate (pre harvest abundance) the abundance of salamanders was greatly reduced under all the canopy removal treatments except for one called shelterwood1 where some of the major trees which form the canopy are left.

One weakness I found in this paper was disregard for possible measurement errors due to potentially different detection probabilities between treatments so their estimates of relative abundance may be biased and misleading

Nonlinear Regression Methods

Concern with the linearity assumption-We should strive for biological realism.

Simple Non Linear Regression

$$Y_i = \mu\{Y \mid X\} + \varepsilon_i$$

$$\mu\{Y \mid X\} \text{ some nonlinear fn of X}$$

Proc NLIN in SAS can be used.

No longer exact inferences and standard regression theory does not apply. Use maximum likelihood to fit the parameters.

Very important to ecologists because so many of the functions from theory are nonlinear.

Non Linear Regression: Asymptotic Regression Example

Asymptotic Regression Function

$\mu\{Y \mid X) = \alpha - \beta\gamma^{X}$

$0 < \gamma < 1, \alpha$ and β positive

$\mathrm{X} = 0 \; \mu\{Y \mid X) = \alpha - \beta$

X large $\mu\{Y \mid X) = \alpha$

No Turning Point - See Curve Next Lecture

Non Linear Regression: Logistic Growth Example

Logistic Growth Function

$\mu\{Y \mid X\} = \alpha / (1 + \beta\gamma^{X})$

$0 < \gamma < 1, \alpha$ and β positive

$\mathrm{X} = 0 \; \mu\{Y \mid X\} = \alpha / (1 + \beta)$

X large $\mu\{Y \mid X\} = \alpha$

Turning Point – See Curve Next Lecture

Multiple Regression Methods

$$Y_i = \mu\{Y \mid X_1, X_2, .., X_p) + \varepsilon_i$$

$$\mu\{Y \mid X_1, X_2, .., X_p) = f(X_1, X_2, .., X_p)$$

Y-Dependent Variable

X- independent variables-how to chose them is a key question?

μ{.} –is the mean function and is conditional on the X's

f- often a linear function of X's but not always.

ε-often independent Normal errors with constant variance.

Multiple Linear Regression

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon_i$$

Least Squares Estimation

Find β's which Minimise

$$\sum_{i=1}^{n} (Y_i - \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)^2$$

Exact inferences, F tests, T Tests to test hypotheses.
See ST 512

Multiple Regression Methods

Specification of Mean Function

-typically separate X variables with each linear function

-can add quadratic or cubic functions by creating a derived variable (X^2). This is still a linear regression in the parameters.

-can add interaction terms by using product of two separate X's.($X_4=X_1*X_3$) Still a linear regression in the parameters

-Proc GLM can run linear regressions of above kinds

-for truly nonlinear regressions use Proc NLIN as noted earlier.

Multiple Regression Methods

Specification of Error Structure

–ε-usually independent Normal errors with constant variance.

-However in spatial processes which are common then have spatial autocorrelation with a decaying structure with distance between points. That is the correlation gets smaller as the points get further apart.Model this as various functions.

-Time Series- in special case of modelling how some variable changes over time there are temporal autocorrelations also with a decaying structure with time between points.

Multiple Regression: Key Issues

Spatial Autocorrelation

-Usually ignored but can cause problems and makes us think our samples are bigger and better than they are.

-We are thus assuming the points are independent when they are not.

-Precision of estimates and power of tests are then over stated.

-Use Spatial model software or else design study so that points are "far apart".

Multiple Regression: Key Issues

Measurement Error in the X's-

We are assuming X's measured accurately.

- If not there is attenuation bias in the estimates of the β's.
- Measurement error may arise due to subsampling or other reasons.

Multiple Regression Methods:Key Issues

Model Selection

- Strive for parsimony in choice of X variables using a model selection procedure like backward elimination, forward selection or stepwise procedures. All possible regressions only used in small problems!
- Most users now consider something like a stepwise procedure or use of the AIC.

Multiple Regression Methods:Key Issues

Model Validation

Use of plots and residual plots to see where model is inadequate(Mean Function, Error Structure, Variance). Note that some violations like pseudo-replication not detectable this way.

Multiple Regression: Key Issues

Sampling Issues- assumes random sampling (Use weighted regressions?), be careful of pseudo replication in space and in time.

Sample Size- be careful especially if you try and include a lot of X variables

Multiple Regression: Key Issues

Multi Collinearity

High Correlations between X's. This is a very important problem. The estimated regression coefficients become very unstable.

Possible Solutions

- Eliminate one of two X's if very high correlation.
- Consider use of principal component regression. Involves running Principal Components Analysis on the X's and use new derived orthogonal X's in regression that are linear functions of original X's .