Model Based Statistics in Biology.

Part IV. The General Linear Model. Multiple Explanatory Variables.

Chapter 13.5 Repeated Measures To be completed

on chalk board

Ch13.xls

ReCap Part I (Chapters 1,2,3,4) Quantitative reasoning is based on models, including statistical analysis based on models.
ReCap Part II (Chapters 5,6,7)
Hypothesis testing uses the logic of the null hypothesis to declare a decision.
Estimation is concerned with the specific value of an unknown population parameter.
ReCap (Ch 9, 10,11) The General Linear Model with a single explanatory variable.
ReCap (Ch 12) GLM with more than one regression variable (multiple regression)
ReCap (Ch 13) GLM with more than one categorical variable (ANOVA).
Two fixed factors (Ch 13.1, Ch13.2)
One fixed and one random factor (Paired t-test, Randomized block)

Today: Special case of two factor ANOVA: Repeated Measures

Wrap-up.

Introduction. A very common design is to make measurements of the same variable on the each of several units on two or more occasions. Conditions differ on each occasion, such as treatment on one occasion, control (no treatment) on another. The major advantage of this design is that variation among units can be removed from the analysis, allowing a more sensitive test of the factor of interest.
The repeated measures design thus resembles the randomized blocks design, having the same motivation–a more sensitive analysis with lower Type II error (erroneously accepting the null hypothesis). If a factor is measured repeatedly on a unit, rather each level of a factor being measured on one unit, then fewer units can usually be used because variability among units can be controlled in the analysis.

Repeated measures designs arise if we have clearly defined units, then measure all levels of a factor on each unit.

If the factor has two levels (e.g., treatment versus control) then the design is also called a paired comparison.

Some definitions.

<u>Repeated measures</u>. A factor is measured at all levels in each of several clearly defined units.

Examples.

Unit is person. Treatments applied in random order to each subject.
Unit is an animal. Size of animal on several occasions to obtain growth rate.
Unit is a tank. Size of fish on several occasions to obtain growth rate.
Unit is a plot. Treatments assigned randomly to adjacent subplots (2 or more).
Treatments are applied in random order, if possible, to eliminate carry-over.

<u>Randomized blocks.</u> A factor is applied to units that can be matched by a random factor, such as propinquity in time or space.

Examples.

Experiment carried out in 3 different labs (blocks).
Experiment carried out on 4 different occasions (blocks).

<u>Paired comparisons</u>. Two levels of a factor within a unit (repeated measures), or in pair of units matched according to some random factor (randomized blocks).

The general linear model will be written the same way for all of these. There will be one factor of interest, and a second or control factor (defined unit or random block). The interaction term is assumed to be absent and hence omitted from the model. This is based on random assignment of treatments within a block or unit (if possible) or on prior knowledge of response of units or subunits of a block.

In each case, the defined unit or random block term is estimated in order to reduced the magnitude of the error term in the model, to achieve a more sensitive analysis with lower Type II error.

Example Data from

1. Construct model

Verbal model.

Graphical model. Plot of

Formal Model

Write GLM:	T	=	β_o	+	$\beta_U \cdot X_U$	+	$\beta_E \cdot X_E$	+	$\beta_{D \times S} \cdot X_D \cdot X_S$	+	residual
S&R95	T_{ijk}	=	μ	+	A_i	+	β_j	+	$(A \cdot \beta)_{ij}$	+	ε_{ijk}

2. Execute analysis.

Place data in model format:

Column labelled L, with response variable

Column labelled X_U, with explanatory variable X_U =

Column labelled X_E with

Code model statement in statistical package according to the GLM

$$L = \beta_o + \beta_U \cdot X_U + \beta_E X_E + \varepsilon$$

```
MTB > anova 'wlngth' = 'XU'   'XE';
SUBC> fits c4;
SUBC> residuals c5.
```

Here are the parameters of the model: the overall mean ($\hat{\beta}_o$), the mean for...
and the fitted values -- the mean for each fly ($\hat{\beta}_o + \hat{\beta}_U \ \hat{\beta}_E$).

$\hat{\beta}_o$ = mean(L) = $24^{-1} \cdot$

$\hat{\beta}_o + \hat{\beta}_E$ = mean($L_{cage=I}$)
mean($L_{cage=II}$)
mean($L_{cage=III}$)

The residuals for each measurement are computed from the means.

3. Evaluate the model

Plot residuals versus fitted values.

a. No line fitted in model, so skip evaluation of straight line assumption.
b. Homogeneity of residuals.
c. If n small, evaluate assumptions for chisquare (t, F) distributions.
n = 24

Homogeneous?
Sum(res) = 0?
Independent?
Normal ?

4. State population and whether sample is representative.

5. Decide on mode of inference. Is hypothesis testing appropriate?

6. State H_A H_o pairs, test statistic, distribution, tolerance for Type I error.
Interaction term.

Are there more specific hypotheses about parameters? No

State test statistic	F-ratio
Distribution of test statistic	F-distribution
Tolerance for Type I error	5% (conventional level)

7. ANOVA table: set up.

GLM:	Y	=	$\$_o$	+	$\$_U X_U$	+	$\$_E X_E$	+ res
Source:	total	=			Unit		Effect	res

Source	df	SS	MS	F	----> p
Unit					
Effect					
Error					
Total					

ANOVA. Calculate df, partition according to model.

GLM:	Y	=	$\$_o$	+	$\$_U X_U$	+	$\$_E X_E$	+ res
Source:	total	=			Unit		Effect	res
	24 ! 1	=			3 ! 1	+		+ 12

Source	df	SS	MS	F	----> p
Unit	2				
Effect	9				
Error	12				
Total	23				

7. ANOVA. Calculate variance, partition according to model.
Calculate SS_{total} = 23* Var(L) = 23*104.43 = 2401.98

Here is the partitioning of the SS_{total} produced by any statistical package.

GLM:	Y	=	$\$_o$	+	$\$_U X_U$	+	$\$_E X_E$	+ res
	2401.97	=			665.68	+	1720.68	+ 15.62

Source	df	SS	MS	F	---->	p
Unit	2	665.68				
Effect	9	1720.68				
Error	12	15.62				
Total	23	2401.97				

Table SS, MS, F-ratio.
Calculate MS = SS/df

Source	df	SS	MS	F	---->	p
Unit	2	665.68	332.84			
Effect	9	1720.68	191.19			
Error	12	15.62	1.3017			
Total	23	2401.97				

Calculate F ratios.

Calculate Type I error from F-distribution.

8. Decide whether to recompute p-value.

9. Declare decision about terms.

10. Report and interpret parameters of biological interest.