Model Based Statistics in Biology.

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

Chapter 13.5 Repeated Measures

To be completed

ReCap. Part I (Chapters 1,2,3,4), Part II (Ch 5, 6, 7)
ReCap Part III (Ch 9, 10, 11)
ReCap Multiple Regression (Ch 12)
13.1 Fixed Effects ANOVA (no interactive effects)
13.2 Fixed Effects ANOVA (interactive effects)
13.3 Fixed*Random Effects (Paired t-test)
13.4 Fixed*Random Effects (Randomized Block)
13.5 Sequential within Random (Repeated Measures)
13.6 Nested Random Effects (Hierarchical ANOVA)
13.7 Random within Fixed (problem of confounding)
13.8 More Than Two Factors (to be written)

Ch13.xls

on chalk board

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)

Today: Special case of two factor ANOVA: Repeated Measures

One fixed and one random factor (Paired t-test, Randomized block)

Wrap-up.

Introduction. A very common design is to make repeated measurements of the response variable within a unit of a random factor. Repeated measures can be longitudinal-repeated measures on the same individual assigned to one level of a fixed factor. Repeated measures can be cross-sectional. That is, repeated measures on an individual, with re-assignment of the individual to another level of the fixed factor. This is called a cross-over design. The major advantage of this design is that variation among units can be removed from the analysis, allowing a more sensitive test of the fixed factor of interest. Adding measurements within a unit costs less than adding units.
The repeated measures design has the same motivation as randomized complete block design RCBD –a more sensitive analysis with lower Type II error (erroneously failing to reject the null hypothesis). The concern is that repeated measurements in time (or within a unit in space) are autocorrelated—they cannot be taken as independent estimates of the response variable.

Some definitions.

Repeated measures. An experimental unit (one level of a random factor) is measured repeatedly.

Examples.

Treatments applied in random order to each subject.
Size of fish in a tank on several occasions to obtain growth rate.
Treatments assigned randomly to adjacent subplots (2 or more).
Treatments are applied in random order, if possible, to eliminate carry-over.
A paired t-test is a special case of repeated measures and of RCBD

Randomized blocks. In a standard RCBD, the blocks are groups of similar experimental units. Subjects are grouped into blocks based on a characteristic, and then treatments are randomly assigned within each block to different subjects.

Examples.

Experiment carried out in 3 plots at three locations in the intertidal zone.
Experiment carried out on several units on 4 different occasions (blocks).

Paired comparisons. 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 have a fixed factor of interest, and a random factor (defined unit or random block). Interaction terms cannot be estimated because fixed and random factors do not pass the cross test.

Example Data from

1. Construct model

Verbal model.

Graphical model. Plot of

Write GLM:	T	=	$\exists_o$	+	$\exists_U \cong X_U$ +		$\exists_E \cong X_E$ +		$\exists_{D \times S} \cong X_D \cong X_S$ +		residual
S&R95	T_{ijk}	=	:	+	A_i	+	$\exists_j$	+	$(A \cong \exists)_{ij}$	+	$_{ijk}$

Formal Model

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

$$L = \exists_o + \exists_U \cong X_U + \exists_E X_E + ,$$

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

Here are the parameters of the model: the overall mean ($_o$), the mean for...
and the fitted values -- the mean for each fly ($_o + {}_{U\ E}$).

$$\hat{\beta}_o = \quad \text{mean}(L) = 24^{!1} \cong$$

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

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	=	$\exists_o$	+	$\exists_U X_U$	+	$\exists_E X_E$	+	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	=	$\exists_o$	+	$\exists_U X_U$	+	$\exists_E X_E$	+	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	=	$\exists_o$	+	$\exists_U X_U$	+	$\exists_E X_E$	+	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.