**Model Based Statistics in Biology.**
**Part IV. The General Linear Model. Multiple Explanatory Variables.**
**Chapter 13.7 Hierarchical ANOVA Random within Fixed** 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 | Repeated Measures, within subject |
| 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)

**Introduction**.

A mixed model consists of at least one fixed factor of interest, and at least one random factor (defined unit or random block).
For nested designs, the main effect likelihood ratio is formed relative to the nested term
Here is an example of a mixed model nested design.
Mixed model Drug is fixed, Subj is random
Design is Random within Fixed

Mixed Model - Correct F-ratios

| Source | df | SS | MS | F ----> | p |
|---|---|---|---|---|---|
| Drug | 2 | 665.68 | 332.84 | 1.74 | 0.23 |
| Subj(Drug) | 9 | 1720.68 | 191.19 | 146.88 | $>10^3$ |
| Error | 12 | 15.62 | 1.3017 | | |
| Total | 23 | 2401.97 | | | |

Mixed Model - Incorrect F-ratios (all ratios relative to residual)

| Source | df | SS | MS | F ----> | p |
|---|---|---|---|---|---|
| Drug | 2 | 665.68 | 332.84 | 255.7 | $>10^3$ |
| Subj(Drug) | 9 | 1720.68 | 191.19 | 146.55 | $>10^3$ |
| Error | 12 | 15.62 | 1.3017 | | |
| Total | 23 | 2401.97 | | | |

The simple rule is test over the random within fixed term.
What about more complex designs?
This is addressed by writing out the expected mean squares.

Forming the likelihood ratios (and *F*-ratios) relative to the residual SS is not always correct. Here is a procedure for identifying the correct F-ratio. To illustrate it, we will use the Extra Sleep example.
Here is a step by step procedure to write out the correct LRs and F-ratios.

List the terms in the model vertically, as in the ANOVA table
List the same terms horizontally.
In each row, show the row term.

| | | Drug | Subj%in% Drug | Error | |
|---|---|---|---|---|---|
| EMS | Drug | Drug | | | Each EMS includes itself |
| EMS | Subj%in%Drug | | Subj%in% Drug | | |
| EMS | Error | | | Error | |

| | | Drug | Subj%in% Drug | Error | |
|---|---|---|---|---|---|
| EMS | Drug | Drug | | Error | Each EMS includes the fixed error term |
| EMS | Subj%in%Drug | | Subj%in% Drug | Error | |
| EMS | Error | | | Error | |

| | | Drug | Subj%in% Drug | Error | |
|---|---|---|---|---|---|
| EMS | Drug | Drug | Subj%in% Drug | Error | Each EMS includes crossed (or nested) random terms |
| EMS | Subj%in%Drug | | Subj%in% Drug | Error | |
| EMS | Error | | | Error | |

| | | Drug | Subj%in% Drug | Error | Correct Denominator MS | |
|---|---|---|---|---|---|---|
| EMS | Drug | Drug | Subj%in% Drug | Error | Subj%in% Drug | Identify the denominator MS for the F-ratio |
| EMS | Subj%in%Drug | | Subj%in% Drug | Error | Error | The denominator MS cancels all but the term of interest |
| EMS | Error | | | Error | | |

| | | Drug | Subj%in% Drug | Error | Incorrect Denominator MS | |
|---|---|---|---|---|---|---|
| EMS | Drug | Drug | Subj%in% Drug | Error | | Drug MS / Error results in *two* uncancelled terms. |
| EMS | Subj%in%Drug | | | | | |
| EMS | Error | | | Error | Error | The F-test is confounded. |

In this example each ratio is formed relative to the term below it in the ANOVA table. This will not always be true in more complex analyses. In these, the EMS table must be written out to identify the correct ratio.

**Distinguishing crossed from nested designs.**

If the random units are nested within the fixed factor, the interaction terms cannot be estimated. A two-way table of the random and fixed factors will be incomplete. As a result, the interactive effect cannot be estimated. Instead, it is folded into a new factor, a random factor within another factor.

Here is an example of data that appears to be crossed, but is in fact nested.

Table 1 from Gutsell 1951. The effect of sulfamerazine on the erythrocyte and hemoglobin content of trout blood. *Biometrics* 7:171-179.

| | Treatment 0 | | Treatment 5 | | Treatment 10 | | Treatment 15 | |
|---|---|---|---|---|---|---|---|---|
| Fish | A | B | A | B | A | B | A | B |
| 1 | 6.7 | 7 | 9.9 | 9.9 | 10.4 | 9.9 | 9.3 | 11 |
| 2 | 7.8 | 7.8 | 8.4 | 9.6 | 8.1 | 9.6 | 9.3 | 9.3 |
| 3 | 5.5 | 6.8 | 10.4 | 10.2 | 10.6 | 10.2 | 7.2 | 11 |
| 4 | 8.4 | 7 | 9.3 | 10.4 | 8.7 | 10.4 | 7.8 | 9 |
| 5 | 7 | 7.5 | 10.7 | 11.3 | 10.7 | 11.3 | 9.3 | 8.4 |
| 6 | 7.8 | 6.5 | 11.9 | 9.1 | 9.1 | 10.9 | 10.2 | 8.4 |
| 7 | 8.6 | 5.8 | 7.1 | 9 | 8.8 | 8 | 8.7 | 6.8 |
| 8 | 7.4 | 7.1 | 6.4 | 10.6 | 8.1 | 10.2 | 8.6 | 7.2 |
| 9 | 5.8 | 6.5 | 8.6 | 11.7 | 7.8 | 6.1 | 9.3 | 8.1 |
| 10 | 7 | 5.5 | 10.6 | 9.6 | 8 | 10.7 | 7.2 | 11 |
| Mean | 7.2 | 6.75 | 9.33 | 10.14 | 9.03 | 9.73 | 8.69 | 9.02 |

Sulfamerazine is sulfa antibiotic found to be successful in treating furunculosisis a serious, septicemic, bacterial disease caused by *Aeromonas salmonicida*. The experiment investigated toxic effects as measured by reduced erythrocyte and hemoglobin content in fish blood. To the author's surprise the drug increased hemoglobin in brown trout, as evident in Table 1.

The table layout suggests that treatment, fish ID, and trough (A or B) are crossed, with repeated measurement of 10 fish.

However the layout was as follows:

> *Five pounds of brown trout from one large pool were placed in each of 8 adjacent upper troughs in the hatchery. Each of the 4 treatments was assigned to 2 troughs by random selection, so that the test was run in duplicate and was fully randomized.*

Thus there was 8 troughs x 10 trout/trough = 80 trout.

Consequently none of the two-way interactive effects can be estimated. For example the sample number in the two way table for Treatment x Trough is:

| | Trough | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H |
| Trt 0 | 10 | 10 | | | | | | |
| Trt 5 | | | 10 | 10 | | | | |
| Trt 10 | | | | | 10 | 10 | | |
| Trt 15 | | | | | | | 10 | 10 |

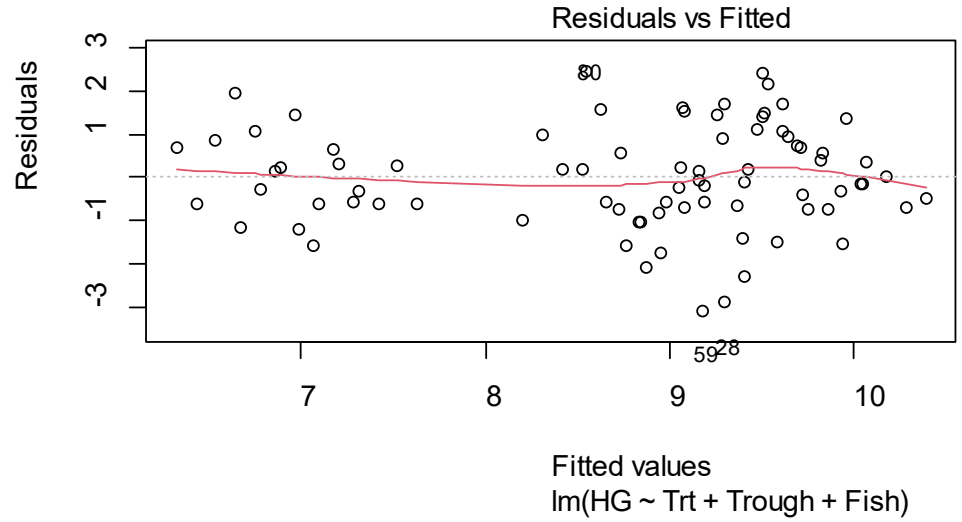
Similarly sparse displays appear in the two way tables for Fish x Trough and Fish x Treatment.

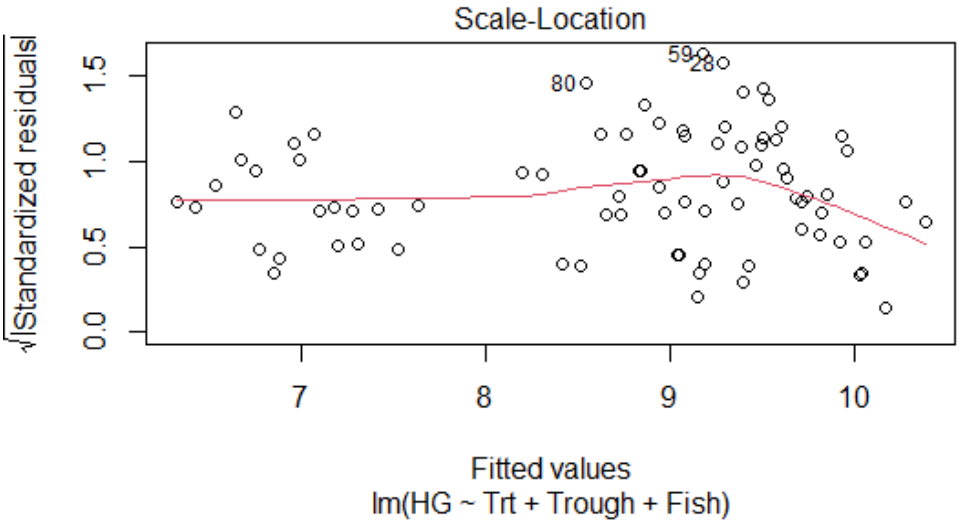
For nested design the interactive effect variance that cannot be estimated appears as the sum or the variance due to main and interactive effects.
Thus var(Trt) + Var(Trt x Trough  become var(Trough(Treatment)), where the nested term Trough(Treatment) is read left to right, as “Trough within Treatment.”

1. Write the model.

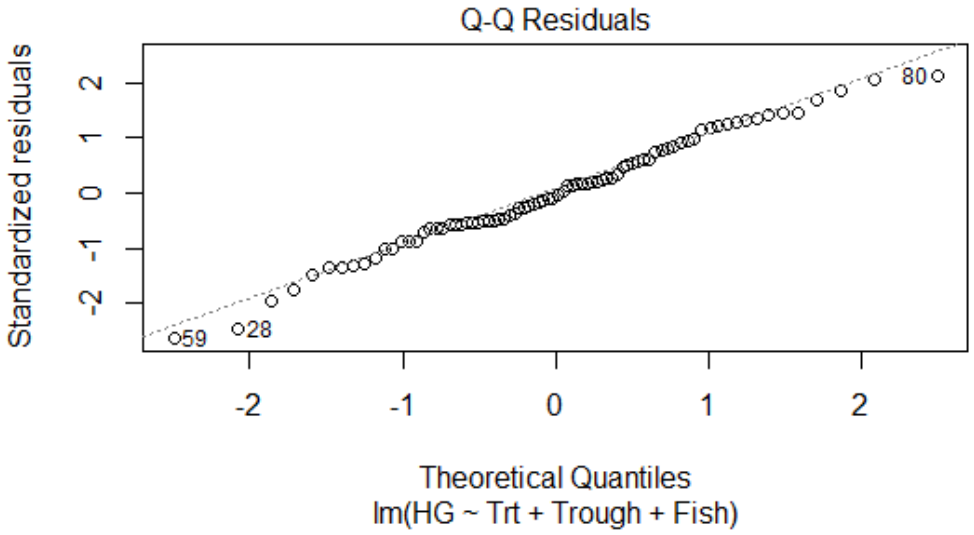
2. Execute, obtain fitted values (means) and residuals.

3. Use residuals to check model
Straight line?  Not applicable.
Homogeneous errors?  Only slightly heterogeneous in first plot



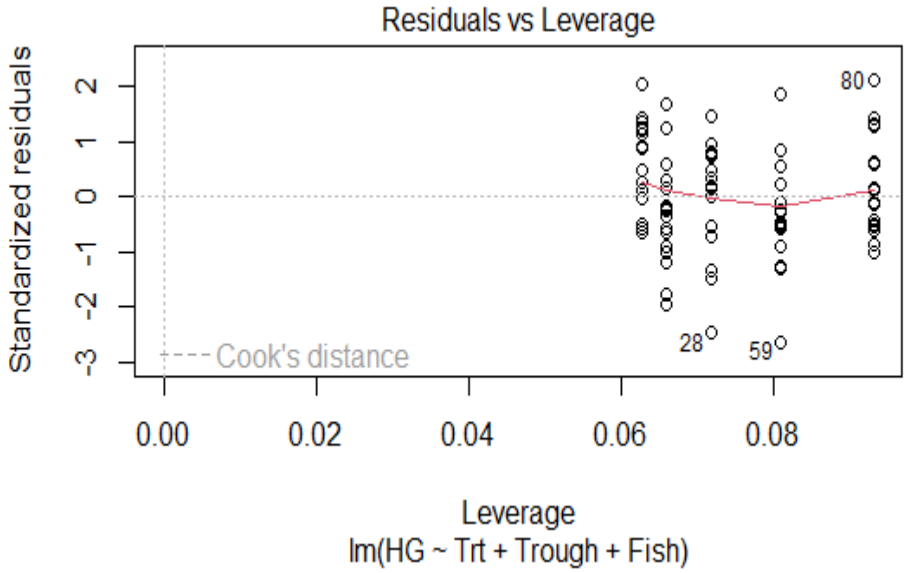

Second plot shows reduction in heterogeneity at large values, suggesting an upper limit to the response variable.  This is consistent with lack of increase in hemoglobin with increase in dose beyond the recommended level.




Normal errors?
Yes

Influential outliers with high leverage?

No.




Conclusion. Normal error model acceptable

Anova Table

```
Response: HG
          Df  Sum Sq Mean Sq F value    Pr(>F)
Trt        3  90.560 30.1868 20.4958 9.066e-10 ***
Trough     1   2.415  2.4151  1.6398   0.20435
Fish       1   7.520  7.5200  5.1058   0.02679 *
Residuals 74 108.989  1.4728
```

Df correct? Yes
Df correct? No
Df correct? No

Model has been executed as crossed design, with 2 troughs instead of 8.
Fish is regression variable. Somewhat disturbingly Hg depends on fish order.
LR = $\{(7.53+109)/109)^{80/2}$ = 14.5 Adequate evidence.

Run model as crossed design to show allocation of interaction term to nested terms.

o.