Index of Lecture 10

Practical Information

Today's lecture:
continuation of random effects models (from Lecture 9):

- model building & checking + software (9L−13/14),
- last example: nested design (pig breeding data, 9L−10),
- split-plot or hierarchical design:
  - ∗ a special type of design which is modelled/analysed by random effects models,
  - ∗ arguably the most important application of random effects models.

Textbook reading:

- GO Chapter 12 (excl. Sections 6,7,9) and Chapter 16 (Sections 1-5); skip discussions involving Hasse diagrams[1] and restricted random effects,
- supplementary notes on linear mixed models, including split-plot designs; part of course curriculum.

Other news / practical information:

- 4th home assignment returned today; 5th home assignment posted,
- course exam set for April 18.

[1] Hasse diagrams is a general graphical tool to compute degrees of freedom and assess the structure of the ANOVA table for random effects models; it's not really needed for simple models but very useful for complex models.

Split-Plot Designs

Consider a two-factor design with blocks or replications:

- treatments (factors) A at a levels, and B at b levels,
- ab experimental units in each of c blocks, or replicated c times (assuming a balanced design).

Randomized (block) design:

- $N = abc$ units allocated randomly to treatments (A×B),
- all treatments estimated with same precision.

Split-plot design:

- A varies between (is applied to) larger units than B (for convenience or of necessity),
- conceptual two-step design construction:
  - i) A is allocated randomly to large units ("whole plots"),
  - ii) units for A are split into sub-units ("split plots"[2]), onto which B is allocated randomly,
- A = "whole-plot factor" and B = "split-plot factor",
- implication: A estimated with less precision than B and the interaction A∗B (not intuitively obvious!).

Analysis of split-plot design:

- include random effects for A-units (whole plots) in model!

[2] GO uses the term "split plots", other authors use sub plots.

Split-Plot Example: Memory Trials

Anxiety, tension, and memory data (GO Example 16.5):

- memory errors recorded in each of 4 memory trials taken (in random order ∼ cross-over design) by 12 subjects allocated to one of four groups defined by different levels of anxiety (low/high) and muscular tension (low/high),

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Anxiety	1	1	1	1	1	1	2	2	2	2	2	2
Tension	1	1	1	2	2	2	1	1	1	2	2	2
Type 1	18	19	14	16	12	18	16	18	16	19	16	16
Type 2	14	12	10	12	8	10	10	8	12	16	14	12
Type 3	12	8	6	10	6	5	8	4	6	10	10	8
Type 4	6	4	2	4	2	1	4	1	2	8	9	8

  notation: y_{ijkl} = # memory errors for trial of type k for subject l in anxiety group i and tension group j,

  * whole plots = subjects (within anxiety×tension group),
  * whole-plot factors: (A) = anxiety, (B) = (tension),
  * split plots = trials (for each subject×type),
  * split-plot factor (C) = trial type,

- split-plot model with replication (instead of blocks):

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + D_{ijl} + \varepsilon_{ijkl},$$

  * $\text{Var}(\varepsilon_{ijkl}) = \sigma^2$ (variation between split plots),
  * $\text{Var}(D_{ijl}) = \sigma_D^2$ (extra variation between whole plots),
  * D_{ijl} ∼ whole-plot random effect for subject (i, j, l).

Memory Example Results

ANOVA table:

Source	DF	SS	MS	EMS	F	P
Anxiety	1	10.08	10.1	$\sigma^2+4\sigma_D^2+\sigma_\alpha^2$	$MS_A/MS_D=0.98$	0.35
Tension	1	8.33	8.3	$\sigma^2+4\sigma_D^2+\sigma_\beta^2$	$MS_B/MS_D=0.81$	0.40
A ∗ Tn	1	80.08	80.1	$\sigma^2+4\sigma_D^2+\sigma_{\alpha\beta}^2$	$MS_{AB}/MS_D=7.77$	0.024
Type	3	991.50	330.5	$\sigma^2+\sigma_\gamma^2$	$MS_C/MS_E=152$	$<.001$
A ∗ Tp	3	8.42	2.8	$\sigma^2+\sigma_{\alpha\gamma}^2$	$MS_{AC}/MS_E=1.29$	0.30
Tn ∗ Tp	3	12.17	4.1	$\sigma^2+\sigma_{\alpha\gamma}^2$	$MS_{BC}/MS_E=1.87$	0.16
A∗Tn∗Tp	3	12.75	4.3	$\sigma^2+\sigma_{\alpha\beta\gamma}^2$	$MS_{ABC}/MS_E=1.96$	0.15
Whole plot	8	82.50	10.3	$\sigma^2+4\sigma_D^2$	$MS_D/MS_E=4.74$	$<.001$
Split plot	24	52.17	2.2	σ^2		
Total	47	1258.0				

- clearly significant effect of type (and no interactions),
- weakly significant interaction between anxiety and tension ⇒ interpretation in interaction plot,
- variance components:

$$\hat{\sigma}^2 = MS_E = 2.174, \quad \hat{\sigma}_D^2 = (MS_D - MS_E)/4 = 2.035,$$

Standard errors of means, and pairwise comparisons:

- whole-plot factors (+ interaction): usual formulae, using MS_D instead of MS_E (∼ analysis of whole-plot means),
- split-plot factor (type): pairwise comparisons using MS_E,
- interactions with split-plot factor: within whole-plot factor pairwise comparisons by usual formulae using MS_E,

– otherwise by approximation method (details in notes).

Correlations in Random Effects Models

Fact: random effects introduce (positive) *correlations* between the observations,

- observations are no longer independent, or
  – (same meaning) observations are clustered,
- observations *at the same level of a random factor* are correlated, whereas observations at different levels of all random factors are still independent,
- rule to compute *intra-class correlations* (ICCs):
  - ∗ compute the total variance, $\mathrm{Var}(y)$, of an (any) observation – as the sum of all variance components,
  - ∗ for the two observations, y_1 and y_2, in question, compute the covariance between them, $\mathrm{Cov}(y_1, y_2)$ – as the sum of all variance components which y_1 and y_2 *have in common*,
  - ∗ the correlation (ρ) between y_1 and y_2, $\mathrm{Corr}(y_1, y_2)$, is the ratio of these:

$$\rho \text{ (or ICC)} = \mathrm{Corr}(y_1, y_2) = \mathrm{Cov}(y_1, y_2)/\mathrm{Var}(y).$$

Example: split-plot model (with σ^2_{AC} and σ^2 as whole-plot and split-plot variances, respectively),

$$\mathrm{Corr}(y_1, y_2) = \begin{cases} \sigma^2_{AC}/(\sigma^2_{AC}+\sigma^2), & \text{(same whole plot)}, \\ 0, & \text{(different whole plots)}, \end{cases}$$

More about Nesting and Subsampling

Recall crossed vs. nested factors A and B:

- A and B are *crossed*, if every level of A and B have the same meaning across the entire experiment,
- A is *nested* within B, if levels of A are specific to levels of B and don't generalize across levels of B.

Subsampling: (also "pseudo-replication")

- multiple measures taken on the same experimental unit,
- often "identical" subsamples, e.g. duplicates,
- split-plot designs: split (sub) plots $\sim$ subsampling of whole plots and allocation of split-factor treatments,
- rarely realistic to assume independence between subsamples; two possible approaches/solutions:
  - average observations to level of experimental units,
  - include random effects $\sim$ units subsampled.[3]

Steps to build an ANOVA model (GO Section 12.3):

1. Determine the sources of variation (HS: or the units),
2. Decide which factors cross and which nest,
3. Decide which factors are fixed and which are random,
4. Decide which interactions should be in the model.

[3] Keeping the original observations but adding suitable random effects, is the equivalent approach to split-plot modelling.

SPLIT-PLOT EXAMPLE: WETLAND WEEDS

Weed biomass in wetlands (GO Example 16.7):

- percent of nonweed biomass after 8 weeks of growth of seeds in clipping-treated halves of weed-treated wetlands placed in 8 trays, onto which nitrogen treatments were applied, positioned on 2 separate tables (in greenhouse),

Table	N	Weed = 1		Weed = 2		Weed = 3	
		C = 1	C = 2	C = 1	C = 2	C = 1	C = 2
1	1	87.2	88.8	70.4	75.7	75.0	80.6
	2	80.5	83.8	59.2	61.5	59.5	62.5
	3	76.8	80.8	47.8	49.5	48.4	52.9
	4	77.7	81.5	35.7	37.3	38.3	42.4
2	1	78.2	80.5	65.1	68.3	65.3	66.6
	2	79.8	85.2	57.6	61.4	58.5	61.6
	3	82.4	83.1	50.5	54.0	51.6	54.7
	4	75.7	78.7	39.0	43.9	41.9	45.1

C=clipping tx

- a split-split-plot design with blocks:
  * whole plots = trays (8),
  * whole-plot factor = N levels (4),
  * split plots = wetlands (24, three in each tray),
  * split-plot factor = weed treatment (3),
  * split-split plots = wetland halves (48),
  * split-split-plot factor = clipping treatment (2),
  * blocks = tables (2),

- notation: y_{ijkl} = biomass percentage at clipping tx k, weed tx j and nitrogen tx i in block l.

Wetland Example Results

Model:

$$\begin{aligned} y_{ijkl} &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \\ &\quad + (\alpha\beta\gamma)_{ijk} + \delta_l + (AD)_{il} + (ABD)_{ijl} + \varepsilon_{ijkl}, \end{aligned}$$

with the whole-plot errors $AD_{il} \sim \mathrm{N}(0, \sigma^2_{AD})$, the split-plot errors $ABD_{ijl} \sim \mathrm{N}(0, \sigma^2_{ABD})$, and the split-split-plot errors $\varepsilon_{ijkl} \sim \mathrm{N}(0, \sigma^2)$, and all variables independent.

Analysis:

- exact model specification difficult for ANOVA-based estimation, but possible for likelihood-based estimation (both Minitab[4] and Stata),
- variance components (from REML-estimation): $\hat{\sigma}^2_{AD} = 14.45$, $\hat{\sigma}^2_{ABD} = 2.62$, $\hat{\sigma}^2 = 1.07$,
- residuals: split-split-plot level ok, split-plot and whole-plot levels from analysis of respective means also ok,
- interaction `n*weed` clearly significant ⇒ interaction plot: less biomass (i.e., more weeds) with increasing N, especially when weeds have been seeded,[5]
- strong main effect of clipping: slight increase of biomass,
- presentation of results: separate for `n*weed` interaction and clipping main effect (because additive), by graphs or (least squares) means with standard errors.

[4] Minitab model specification in `Mixed Effects Model` menu: random `tray` effect and `tray*weed` interaction.

[5] The interaction may be explored in the split-plot model for wetland means.

Final Remarks on Split-Plot Designs

Split-plot vs. hierarchical data structure:
much the same thing...:

- split plots constitute a level below the whole plots:

  whole plot └── split plot   similarly to   herd └── animal

- in a hierarchical structure, upper level units $\sim$ whole plots and lower level units $\sim$ split plots,
- explanatory variables may vary at upper or lower levels (possibly not randomised as in a split-plot design).

Two types of split-plot designs:

- with blocks (wetland weeds example),
- with replicates instead of blocks (memory example).

How to identify a "split-plot" structure?:

- *measurement* units are smaller than *experimental* units (for some treatment factors),
- expl. variables vary between "units" of different size,
- GO: follow the randomization!,
- think about whether subsampling was done,
- ask yourself: how many exper. units for this factor?
- draw a sketch of the experimental layout!