Index of Lab 14

Practical Information

Today's session:

- project reports returned, with individual discussions,
- exam2011:1,2,3; exam2014:1,3; exam2012:3 (exam:2015,2016),[1]
- exam topics today:
  - exam logistics/practical remarks,
  - questions about exam assignments (types, calculations)?
  - suggestions for your review/practice,
- additional review topics we could include today:
  - model choice (new slide),
  - overview of model types (new slide),
  - tips for presenting interaction effects (new slide),
  - hierarchical structure and nesting (new slide).

Exam questions (2011, or others):

- do them (somewhat realistically) without computer,
- check out the solutions,
- extra Stata do-files at exercises page: maybe do some extra analyses in Minitab/Stata for the questions where data is provided.

What you should be able to do after this course...

- build models of the types encountered in the course, and validate them,
- analyse models to determine effects of importance, and interpret them,
- critically assess the usefulness of computer software output and presentation of analyses in papers.

[1] Note: Question 3 in exams is only for students not taking VHM 812 simultaneously!

Exam Practical Remarks

Final exam: Thursday 18/4, 9am-12pm, AVC 278N (we may want to start 5-10 minutes early because the room is booked by others at 12pm).

Two versions of exam:

- "reduced" (Daniel, Sarah): 10am−12pm, 2 questions,
- "full" (Eric): 9am−12pm, 3 questions.

All aids (books and notes and calculators) are allowed,
— except a computer or computer-like device (tablet or smartphone).

The questions have equal weight, unless specified otherwise − use your time sensibly!

Some hints and advices: (to use or not...)

- layout: essential requirements are
  - readability,
  - clear division between what is *in* the solution and what is not,
- conclusions should be part of all analyses,
- statistical model(s) should be part of all data analysis,
- explicit calculations may prevent loss of points due to typing errors (or the like),
- errors: if you realize an error and do not have time to correct it: write what is wrong, what should have been done and how the error would affect the result,
- sketch of computer analysis: specify i) how the suggested analysis would be done, and ii) how you would use/interpret the results.

Note: the exam finishes at 12pm (sharp!)

Suggestions for Your Review

Check: Course syllabus and Notes about exam on the web.

Suggested exercises to review:

- exams 2011–2016 (no exams for 2017–2018),
- all home assignments and their solutions (in particular, my comments for your answers),
- all regular exercises listed for the lab sessions,
- all VHM 812 exercises (VER XX), perhaps excluding VER 22,
- perhaps also extra exercises listed for the lab sessions,
- lots of exercises and problems in GO textbook (data files at Gary Oehlert website),
- don't skip the exercises on sample size...

Choice of Statistical Model

Some useful questions to ask about the data:

- purpose of study?
- response or explanatory variable?
- continuous or discrete/categorical variable?
- particular data structures? – e.g.
  - repeated measures / longitudinal data,
  - hierarchical structure,
  - split-plot units (some "treatments" on larger units than others) or subsampling,
- random (instead of fixed) effects?
- variable(s) of blocking type? (division of experimental units into homogeneous groups, with no intrinsic interest) – or obvious blocking schemes? (Latin square, BIBD etc.), versus "pure" replication,
- interactions between variables? (quantitative or categorical)
- continuous variable (explanatory or response) to be used for prediction of another variable? (regression)
- transformation? (to achieve normal distribution for residuals, homogeneity of variance, linear relation).

Overview of Model Types

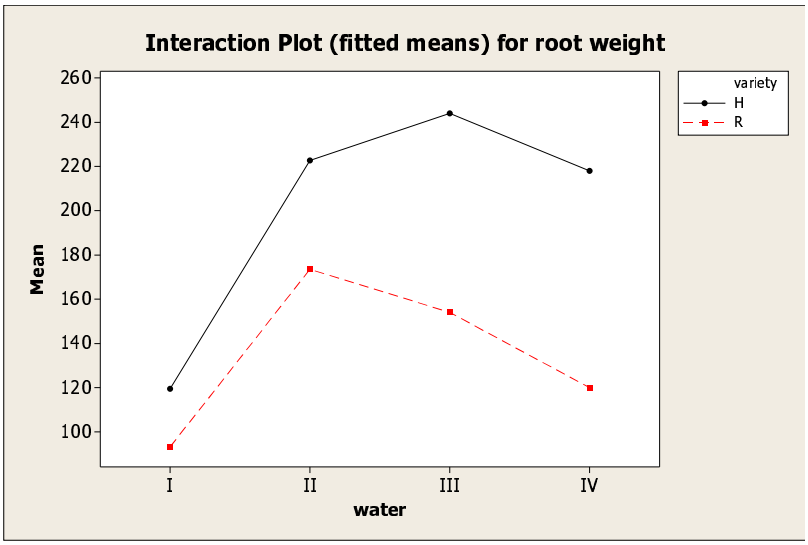
Model type[2]	Characteristic	Topics for analysis
basic (VHM 801)	single explan. variable	4-step appr. for CI and test, ANOVA table, F-statistics, transformation e.g. Box-Cox
multiple linear regression	quantitative explanatory variables	residuals, diagnostics, outlier test, collinearity, test reduced/ full model, variable selection
ANOVA models, (general) linear models	categorical explanatory variables ("factors")	replications, blocks, interactions, contrasts, dummy variables, multiple comparisons, least squares means, designs: Latin square, BIBD, Youden, cross-over
random effects models	right hand side random variables (in addition to ε)	variance components, extra residuals, more complex SEs, 2 methods of analysis: ANOVA-based, likelihood-based
repeated measures, longitudinal data	repeated observ. on same "subject" over "time"	different approaches: separate times, response features, hierarchical/ split-plot, (ε-correction), (mixed w. correlation structure)

[2] models for continuous outcomes ( ) $\sim$ not in syllabus

Tips for Presenting Interactions

- most helpful tool: interaction plot (2-factor interaction),
- focus on combined factor for estimates (least squares means), confidence intervals and pairwise comparisons[3], for balanced designs often most conveninently using LSD statistics,
- adjustment for multiple testing may be restricted to subset of all comparisons within combined factor; main example,
  - within levels of factor A: compare factor B, and vice versa,
- what about "uninteresting" interactions?
  - if non-significant: disregard/drop, and look at main effects,
  - if significant: maybe take interaction as random effect,

Example: root weights in greenhouse trial (water & variety)



Mean	Watering type			
Variety	I	II	III	IV
H	119.50	222.75	244.00	218.00
R	93.25	173.50	154.00	120.00

[3] Special formulae apply to models with random effects, e.g. split-plot designs.

Hierarchical Structure and Nesting

Hierarchical structure is about experimental units:

- exists when there are different types/sizes/levels of exper. units,
- use diagram to display hierarchical structure and the level(s) of the predictors (where they were applied or where they vary),
- assumption for hierarchical structure:
  every unit appears only with one value at levels above (e.g., a cow is in one herd, a rat is from one litter) = *nesting of units*.

Nesting is *also* about factors/predictors:

- previous definition:
  - a (random) factor B is nested within A, written as B(A) (in Stata as B|A), if there is no relation between the levels of B across the levels of A,
  - example: litter nested within strain, sow nested within boar (in a reproduction experiment)...,
  - often corresponds to B being "applied to" or varying at the units of A,
- technical use in Minitab (and other software) to help program recognize hierarchical structures:
  - declare units at a level ($\sim$ random effects) as "nested" within all factors at that level,
  - examples: `guinea_pig(dose)`; `herd(region)`; `dog(tx depl)`,
- technical use in Minitab (and other software) to get parameter estimates for an *interaction* without one of the main effects:
  - examples: `sex bodyw(sex)`; `time tx(time)`.