Index of Lecture 14: Course wrap-up

Practical Information

News / Schedule:

- project presentations 1-2pm, followed by course evaluation and lecture part of session (approximately 2-4pm),
- projects with comments and mark to be returned to you on Monday,
- updated Instructions for home assignments and exam posted to webpage,
- small addition to aid interpretation posted to slide 13L−15.

Today's lecture:

- preparations for exam: Wednesday 16/4, 9am-12pm, AVC 278N,
  - exam practical remarks,
  - exam questions (types, calculations),
- a few review slides: new or partly new slides, 14L−5/6/(7)/8,
- review of some questions, suggested choice: final exam 2015,

Q1: experimental design and analysis,

Q2: repeated measures,

Q3: multiple linear regression.

Exam Practical Remarks

Only one version of the exam:

- "full" (Serem, Taylor): 9am−12pm, 3 questions.

Exam rules: open book (all aids are allowed), except a computer-like device, although cell phones and tablets can be used (in flight mode):

* as a calculator (with basic calculation functions, no statistical software),
* to access course notes and electronic textbook material.

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

- the questions have equal weight, unless stated otherwise − use your time sensibly!
- layout — essential requirements: readability, and a clear distinction between what is *in* your answer and what is not — don't write first a draft and then a final version,
- conclusions should be part of all analyses,
- statistical model should be part of all data analysis,
- 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,
- the exam finishes at 12pm (no extensions of time).

Exam Questions

The plan is for two questions on VHM 802 material and one question on regression (linear and/or logistic) in VHM 802/812 combined material, with all questions having equal weight (totalling 30%).

Typical elements of exam questions:

- choose between different analyses with results provided (a mix of Minitab and Stata listings per default),
- interpret the results of provided listings,
- supplement the provided listings with extra calculations for specific purposes (e.g., statistical inference),
- outline/sketch analyses based on description of data, specifically:

  i) how the suggested analysis would be done (indicate relevant Minitab menus or relevant Stata commands),

  ii) how you would use and interpret the results,
- planning of (simple) experimental designs,
- multiple choice (one or several correct answers).

Suggestions for Your Review

Check (naturally): the course syllabus and the Instructions for home assignments and exam at the website.

Suggested exercises to review:

- exams 2014−2016, 2019, 2021 and 2024 (no exam in 2023!),
  - not all questions will apply (the multivariate analysis questions can be ignored: 2021:2 and 2024:2),
  - always the regression question as the third question (when included),
- 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 (i.e., VER XX), except for 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 (Univariate) 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 or experimental designs? – e.g.
  - cross-over design,
  - hierarchical structure,
- random (instead of fixed) effects?
- blocks: do the data include 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).

Basic Overview of Model Types

Model type[1]	Characteristic	Topics for analysis
basic (VHM 801)	single outcome and explanatory variable	4-step approach 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, margins and least squares means, designs: Latin square, BIBD, cross-over
random effects models	right hand side random variables (in addition to ε)	variance components, extra residuals, more complex SEs, likelihood-based analysis
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)

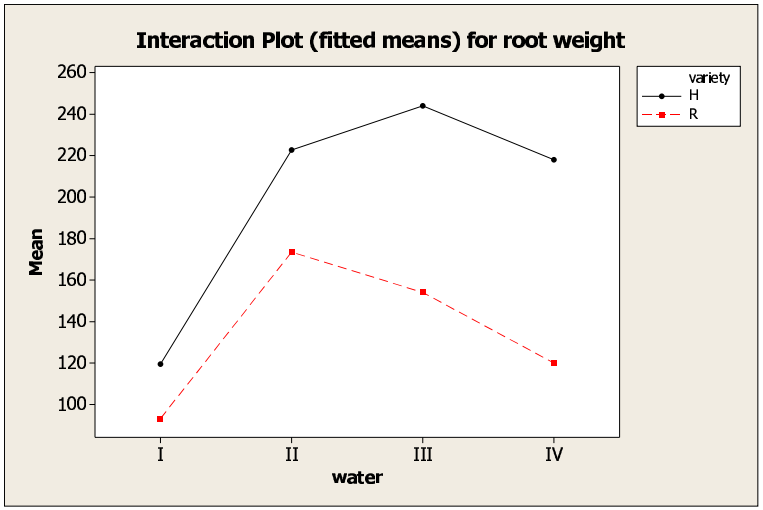
[1] Models for continuous outcomes with normal distribution errors.

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, for balanced designs maybe using LSD statistics,
- adjustment for multiple testing may be restricted to subset of all comparisons within combined factor; one major 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[2].

Example: root weights in greenhouse trial (Additional exercise 10.5)

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



[2] Beware that the parameter estimates may not represent the main effects when an interaction is present.

Hierarchical Structure and Nesting

Hierarchical structure is about experimental units:

- exists when there are different types/sizes/levels of experimental 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,
  - examples: litter nested within strain, sow nested within boar,
  - often corresponds to B being "applied to" or varying at the units of A,
- technical use — to help program recognize hierarchical structures:
  - declare units at a level (∼ random effects) as "nested" within factors at that level,
  - examples: `guinea_pig(dose)`; `herd(region)`; `dog(tx depl)`,
- technical use — to get parameter estimates for an *interaction* without one of the main effects: examples: `sex bodyw(sex)`; `time tx(time)`.