

Index of Lecture 15

Page	Title
1	Practical information
2	Exam practical remarks
3	Suggestions for your review
4	Choice of statistical model
5	Overview of model types
6	Tips for presenting interactions
7	Hierarchical structure and nesting

PRACTICAL INFORMATION

Last information:

- final exam: Thursday 21/4, 9am-12pm, AVC 278N,
- new webpage items: course syllabus and evaluation.

Today's session:

- exam topics today:
 - * questions about exam assignments (types, calculations)?
 - * exam practical remarks,
 - * suggestions for your review/practice,
- formal course evaluation (approx 10:40am) by GSR Office,
- additional review topics in today's lecture:
 - * model choice (new slide),
 - * overview of model types (new slide),
 - * tips for presenting interaction effects (new slide),
 - * hierarchical structure and nesting (new slide).
 - * follow-up from last home assignment,
 - * discussion of Exam 2011:2 and GO Problem 16.8.

EXAM PRACTICAL REMARKS

Two versions of exam:

- “reduced” (Danielle, Dennis, Emily, Joan): 10am–12pm, 2 questions,
- “full” (Mary-Ann, Suppada): 9am–12pm, 3 questions.

All aids (books and notes and calculators) are allowed,
– everything except a personal assistant or computer.

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,
- sketches of computer analyses: specify i) how the suggested analysis would be done, and ii) how you would use and 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 2008–2015 (no exams for 2009–2010),
- 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

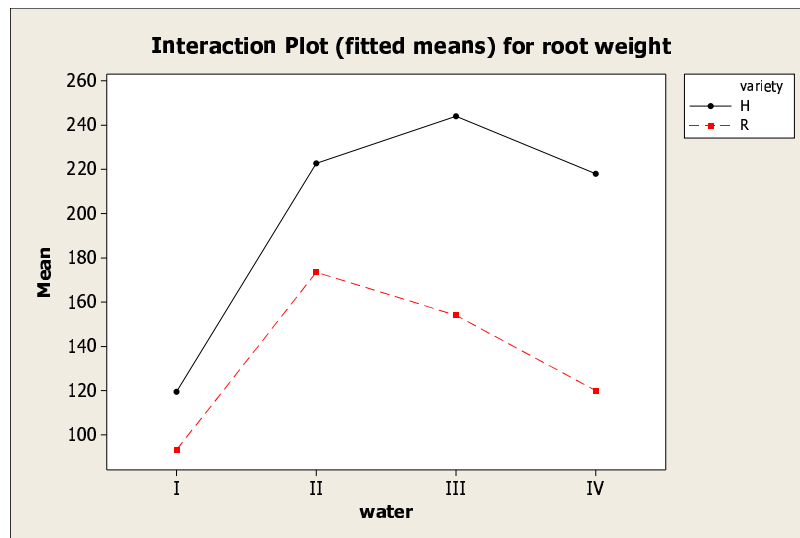
Model type ¹	Characteristic	Topics for analysis
basic (VHM 801)	single explan. variable	4-step appr. for CI and test, ANOVA table, F -statistics, transformation e.g. Box-Cox
multiple linear regression	quantitative explanatory variables	residuals, diagnostics, outlier test, collinearity, test reduced/full 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)

¹ 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 comparisons², for balanced designs often most conveniently 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

² 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)`.