Index of Lecture 14: Course wrap-up

Practical Information

Major news:

- projects with comments and mark to be returned to you in Monday's session,
- course Syllabus and updated Instructions for home assignments and exam posted to webpage,
- request for questions and review topics for the last session (14—P),
- the EXAM..., for 30% of the course mark and conducted as a take-home exam, starting Thursday 22/4, 9am, in two versions:
  - "full exam" (Simon, Paul): 3 questions, completion time 30 hours,
  - "reduced exam" (everybody else): 2 questions, completion time 24 hours.

Today's lecture:

- exam practical remarks,
- exam questions (types, calculations),
- a few review slides: new or partly new slides, 13L—4/5/6/7,
- review of exam questions from final exam 2015:[1]
  1. ANOVA/experimental design,
  3. regression.

[1] Question 2 of the 2015 exam is not relevant for this year's version of the course.

Exam Practical Remarks

For the "rules" around the exam, see the updated Instructions for home assignments and exam at the course homepage.

Some notes on logistics:

- the exam questions will be available at Moodle and/or the course homepage in either `.htm` or `.pdf` formats, and with datafiles in `.csv` format,
- send any questions to me by e-mail at the usual address (`hstryhn@upei.ca`),
- submit your answers through Moodle as electronic files:
  - different files and file types are allowed, e.g. Word, `.pdf`, scanned text (`.pdf` or standard image formats), Minitab project files, Stata do-files, R programs,...
  - if multiple files are submitted, their roles as part of the answer must be clear.

Expected level of detail:

- statistical model (see below for methods when no specific model is involved) should be part of all data analysis,
- for analysis by software: the method/settings used should be described and justified (as per statistical reporting), and results should be presented in the text,
- assessment of model/method assumptions is understood to be part of all data analysis, and conclusions and interpretations should be part of all analyses.

Exam Questions

The plan is for two questions in the "reduced exam", and one extra question on regression (linear and/or logistic) in the "full exam", in both cases with all questions having equal weight (totalling 30%).

Question types:

A) a "standard" question similar to previous (in-class) exams[2],

B) a "small[3] home assignment"-type question focused on practical data analysis and interpretation,

   * no restrictions on your choice of software or software versions,
   * cannot guarantee that every analysis will be do-able in every software, but "advanced coding" will not be needed for major parts of the question,
   * the intent is to offer some choice (similar to the VHM 801 exam in the fall),
   * more focus on interpretations than on coding,

C) a regression question for the "full exam": most likely based on prints without extra analysis (so similar to previous exams).

[2] Such questions require no (major) software analysis, but are typically based instead on supplied Minitab and Stata listings (R output will most likely not be supplied).

[3] Such a question will typically be "smaller" than home assignments, by including means less sub-questions and -analyses than a typical home assignment.

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 2012−2016 and 2019,
  - not all questions will apply (e.g. no repeated measures or split-plot models),[4]
  - 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,

Review requests for Session 14−P are, as earlier mentioned, very welcome.

[4] The listing of exam questions for the lab sessions will include all relevant questions ⇒ those not listed can be skipped!

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).

Choice of Multivariate Method

Most important consideration = objective of analysis,

- analysis without a clearly stated objective will not count fully (and may be misunderstood),
- suggested to review Manly's datasets, and their use for different objectives.

Features of the data to look for:

- do the data include a grouping variable of interest?,
- what are the units on which multiple measures are taken?
  - are there hypotheses of interest related to those units?,
  - do the units have a relevant structure (e.g., hierarchical or factorial)?,
- are the multiple measures to be considered as outcomes or predictors?
- can distances of interest be defined?, and if so, should variables be standardized for these to be more meaningful?,
- do the data or the problem suggest certain graphical displays to be of particular interest?,
- the usual categorizations of and descriptors for individual variables, perhaps involving the distributional assumptions that can reasonable be made.

Basic Overview of Model Types

Model type[5]	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
multivariate analysis	multiple measures on same "subject"	wide range of techniques with different aims: dimension-reduction, classification, relations between subjects and/or variables, distances

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