Notes for Exercises in Session 11

- 13:3,4,19,**31**; 26:7,8,33,**37**; 28:4,**14**; 11:15,16,17; **final2013:2 home2003:4**; (11:6,45; 13:15,16; 26:17,18,19),
- follow-up from lecture: multiple regression (11L−17/18/19),
- no summary worksheet at 3:30pm, discuss home asgnm 3 instead?,
- individual work on exercises + time for questions on 4th home assignment (and on 3rd home assignment returned today).

Minitab for 2-way ANOVA and multiple regression[1] (with demos):

- 2-way table of descriptive statistics using
  `Stat-Tables-Descriptive Statistics`,
  enter outcome in submenu `Associated Variables`,
- brief analysis for balanced data: `Stat-ANOVA-Balanced ANOVA`, and `Main Effects` and `Interaction Plots` menus based on means,
- detailed analysis: `Stat-ANOVA-General Linear Model`, giving
  - estimates and means with SEs (choose in `Options` submenu),
  - extra menus for `Comparisons`, `Predict`, `Factorial Plots`,
- multiple regression (also) from `Stat-Regression` menus.

Notes and questions for specific exercises:

- 13.3, 13.4: we did part (a) in lecture 11, maybe limit to (b),
- 26.33: analyze also on square-root scale: compare the results and include a full post-ANOVA table analysis,
- home assign.2003:4: most detailed 2-way ANOVA examples in course,
- final2013:2 (recommended): exam question involving 2-way ANOVA.

[1] Stata: brief analysis by `Stat-Linear Models-ANOVA-Analysis` menu, using model formulae, or directly with `regress` and `anova` commands; for detailed analysis (Stata/R) see solution files.

Notes on Home Assignment 3

Biggest problems:

i) getting the data/tables right for Q2 + Q3:

- in order to explore association/dependence between categorical variables, you need to cross-tabulate them,
- it is crucial to realize that the two tests are applied to the same samples (this should be obvious from both the text and the layout of the data file),

ii) to explore the association in Study B without splitting,

- one would need to look at conditional probabilities in the full table, and develop statistical inference for those,
- the cells with highest X^2-contributions are helpful to focus attention, but don't offer quantitative conclusions.

Other comments:

- very many answers that the 2-sample z-test for independent proportions would be valid; *it's not*, because the samples are paired (2 results from the same fish/sample), not independent,
- dependence/association for binary variables cannot be explored by methods for quantitative variables (scatterplot, correlation),
- some serious confusion around (of unknown origin, not coming from me!) about comparing confidence intervals (CIs) for parameters, such as means and proportions:
  - just *invalid*, if the estimates are not independent,
  - the optimal way of comparing two groups is by a CI for the parameter difference (e.g., $p_1 - p_2$), and "overlapping CIs" does not in itself provide any conclusions about significance!