

## LAB SESSION 11

### Outline of lab session:

- brief extra review of model building and checking (11L–17/19),
- Minitab for 2-way ANOVA and multiple regression<sup>1</sup> (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 + Main Effects, Interaction Plots` menus based on means,
  - \* detailed analysis: `Stat-ANOVA-General Linear Model`, for
    - estimates and means with SEs (in `Options` submenu),
    - extra menus for `Comparisons, Predict, Factorial Plots`,
  - multiple regression (also) from `Stat-Regression` menus.
- summary worksheet (3:00pm): S.11:3 (`satscore`), S.10 (`sales`),
- individual work on the exercises, and time for questions...:  
13:3,4,19,31; 26:7,8,33,37; 28:4; 11:15,16,17; 28:14; **final2019:2**;  
**home2003:4** (11:6,45; 13:15,16; 26:17,18,19),

### Notes and questions for specific exercises:

- 11.15, 11.16, 11.17: description of dataset in solution for 11.15,
- 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,
- final2019:2 (recommended): exam question involving 2-way ANOVA.

---

<sup>1</sup> 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.

## HOME ASSIGNMENT III

### Major discussion points:

- choice of statistical model (model I or II?):
  - \* all variables can be seen as response variables  
 $\Rightarrow$  model II,
  - \* primary focus is on support responses, so other variables can be considered as explanatory (fixed, no interest in their distribution)  $\Rightarrow$  model I,
- discussion/analysis after significant  $X^2$ -test:
  - \* comparison of conditional and marginal distributions,
  - \* focus on table cells with highest  $X^2$ -contributions,
  - \* significance in itself is of **limited interest** without some description and exploration of differences,
- nonparametric test:
  - \* tests different hypotheses than the  $X^2$ -test (focus is on ordering, not on individual categories),
  - \* avoids assumptions about the scale (one reason why a normal distribution analysis is questionable),
- concerns about separate analyses for age and gender:
  - \* strong lurking/confounding variable (e.g. gender, for the association age  $\leftrightarrow$  support)  $\sim$  Simpson's paradox,
  - \* inconsistent effects across categories of ignored variable  $\sim$  interaction.