This is the last assignment – Turn in your homework in lecture or in your TA’s mailbox by 4pm on Monday May 14 (after classes are over). Note the unusual due date. The solution will be posted that Monday night right after it is due, to help everyone prepare for the final exam. For that reason, late extensions will not be granted. Make sure you write your name and your discussion section. Show your work for full credit.

0. Turn in form C for each project report you are evaluating by Th. May 10 (each total should be on 30, not 60 as indicated). Also turn in form D on May 10 (evaluation questionnaire, one per person).

1. Problem 13 on page 503 (chapter 17).

2. Problem 14. You may do this all in R. The data will be provided in a csv file on the course website. The key commands are:
   - `lm` to fit the regression line, `anova` to obtain the ANOVA table and associated F-test,
   - `summary` to obtain a bunch of information including standard errors,
   - `plot` to obtain a scatter plot of the data, and `abline` to add the regression line.

Refer to lecture notes for examples of how to use these functions (slides 15, 25, 29, 36).

3. Problem 15 p.503-504. Do this all by hand. In case you want to check your results with R (which you are encouraged to do), a data file will be provided on the course website. In this problem, add the following questions.

(c2) Build the ANOVA table to test the null hypothesis of no effect of the number of earthworm species on nitrogen content. Identify the residual standard deviation (standard deviation in nitrogen content around the regression line).

(d2) Provide a 95% prediction interval for your predicted nitrogen content in soil having 5 earthworm species.

(g) What are your assumptions in (c2)-(f)? List the assumptions and assess their validity for this particular data set.

4. What is the flaw? In a metabolic study, four male swine were tested three times: when they weighed 30 kg, again when they weighed 60 kg, and again when they weighed 90 kg. During each test, the experimenter analyzed feed intake and fecal and urinary output for 15 days, and from these data calculated the nitrogen balance, which is defined as the amount of nitrogen incorporated into body tissue per day (in g/day, see results in the table). Suppose the data are analyzed by linear regression with $x = \text{body weight}$ and $y = \text{nitrogen balance}$. Preliminary calculations yield $\bar{x} = 60 \text{ kg}$ and $\bar{y} = 18.7 \text{ g/day}$. The slope is $b_1 = .017 \text{ g/day/kg}$ with standard error $SE_{b_1} = .032$. The $t$ statistic is $t = .53$. On $12 - 2 = 10$ degrees of freedom, the p-value is 0.60, which is not significant at any reasonable level. According to this analysis, there is insufficient evidence to conclude that nitrogen balance depends on body weight under the conditions of this study.

The preceding analysis is flawed in 2 major ways. What are they? Hint: display the data. No calculations needed.

Reading: Interleaf 9 (it’s humbling) and Interleaf 10, Chapters 16 and 17. Sections not covered in lecture or on the final exam are: SE and CI for a correlation coefficient in 16.1, 16.2, 16.3, 16.5, 16.6. In 17.2, we do not cover confidence bands, but we do cover prediction intervals. Other sections not covered are 17.6, 17.7 and 17.8.

Cécile Ané