

Stat 511, HW 4

1. The data in protein.txt are made up based on an experiment done in the Food Science and Human Nutrition Department. This study evaluated two chemical solutions used in extracting a protein meal from beans. The beans are ground, then soaked in the chemical solution for a set length of time. Four soaking times were used: 5 min, 10 min, 15 min, and 20 min for each of the two chemicals. There was also a 0 min (no soak) treatment. The response variable is the %protein in the extract.

The first few questions concern the usual 2 way ANOVA model fitted with terms in the order: time solution time*solution.

- (a) Report the df, SS, and p-value for the type I (sequential tests).
- (b) Report the df, SS, and p-value for the type III (partial tests).
Note: SAS and R will give different answers here.
- (c) Explain why the df for time is different in the type I and type III analyses.
Hint (R):, look at the names on the coefficients, or head(model.matrix(lm object)). drop1() will remove all the columns for each factor. Think about what R's choice of interaction columns means when they are included in the model.
Hint (SAS): Think about what including a column for solution = None means for time = 0. It also may help to fit the model with terms in the order: solution time time*solution.
Note: If you're a SAS user, you might want to add a LSMEANS time solution; statement to see what SAS does with missing cells.

The following parts are based on a 1-way ANOVA model. Your choice of cell means or factor effects model. The investigators want to answer four questions.

- How big is the difference between no soaking (0/None) and the average of the other 8 treatments?
 - What is the linear slope (protein vs time) for solution A?
 - What is the linear slope (protein vs time) for solution B?
 - Are these two slopes different?
- (d) Estimate the difference between no soaking and the average of the other 8 treatments. Report your estimate and its s.e.
 - (e) Estimate the linear slopes for solution A and B. Report your estimates and their s.e.'s.
Hints: The coefficients for the linear orthogonal polynomial for 5 groups are -2 -1 0 1 2. For this data set, with 5 reps per group, you can convert that estimate into the actual slope by dividing by 50. (i.e. by $5 * \sum l_i^2$)
Be careful to get coefficients on the right group. The order will probably not be -2 -1 0 1 2, but that depends on whether you reorder the groups. The two contrasts will not be orthogonal.
 - (f) Test whether the two slopes are equal. Report your test statistic and p-value.

2. The file doughnut2.txt contain data from an experiment on fat adsorbision by doughnuts. It compares four types of cooking fat in a completely randomized design. There are six replicates of each fat. The response is the amount of fat in the cooked doughnut. The question was whether the type of cooking fat mattered, i.e. did cooking in certain fats lead to doughnuts with more fat?

The sum-to-zero constraint says that in the model $Y_{ij} = \mu + \beta_i + \varepsilon_{ij}$, where i indexes treatments and j indexes observations, $\sum_{i=1}^4 \beta_i = 0$.

- (a) Write out the expressions relating the treatment means, μ_i , to the intercept and 3 unconstrained β 's.

A factor effects model with sum-to-zero constraints is equivalent to a multiple regression with indicator variables (B1, B2, and B3) that have the following values:

Fat	B1	B2	B3
1	1	0	0
2	0	1	0
3	0	0	1
4	-1	-1	-1

These values are in the doughnut2.txt file.

- (b) Fit a multiple regression using this set of indicator variables (B1-B3), including the default intercept. Use the estimated regression coefficients to compute the estimated means for each of the four treatments.