Assignment 4 — due February 18, 2019

1. Consider the following artificial data from a two-factor design. The observations were allocated in a completely random fashion. The data are available in `hw04.csv` and shown below.

y	trt	a	b	x1	x2	x3
17	ab	1	1	1	1	1
20	ab	1	1	1	1	1
18	ab	1	1	1	1	1
13	ab	1	1	1	1	1
7	a	1	0	1	0	0
9	a	1	0	1	0	0
12	a	1	0	1	0	0
8	a	1	0	1	0	0
25	b	0	1	0	1	0
17	b	0	1	0	1	0
19	b	0	1	0	1	0
19	b	0	1	0	1	0
29	1	0	0	0	0	0
28	1	0	0	0	0	0
32	1	0	0	0	0	0
24	1	0	0	0	0	0

   (a) Fit a one-way model, using `R`, to the factor called `trt`. By using the appropriate contrasts, determine the F-values (which are T^2 values here) to test the significance of the main effects for A and B and the AB interaction. Caution: The internal ordering of the treatment labels within R might be different from the ordering in the input file.

   (b) Fit the two-factor model including terms for A, B, and the AB interaction. Demonstrate that the F-values for contrasts in part (a) yield the same results as the two-factor F-values. Note that the `anova` funtion in `R` returns the sequential SSs, also called "type I" SSs, whereas the tests for the contrasts in part (a) use the "additional SSs", also called "type III" SSs. These two types of SSs are equal when the design is balanced.

   (c) Construct an interaction plot for these data. Include on your plot standard error bars, or standard deviation bars, or an LSD "bar" (and say what your bars mean).

   (d) In R, fit a regression of `y` on `x1`, `x2`, and `x3`, and show how to use the `anova` function in R to get the SS from parts (a) and (b). What happens if you change the fitting order in R to be `x3`, `x2`, and `x1`? Explain why the fitting order `x1`, `x2`, and `x3`, gives the desired SS, but the fitting order `x3`, `x2`, and `x1` does not.

2. An experiment was conducted to compare the effect of different soil pH and calcium additives on the increase in trunk diameters for orange trees. Annual applications of elemental sulfur, gypsum, soda ash, and other ingredients were applied to provide pH value levels of 4, 5, 6, and 7. Three levels of a calcium supplement (100, 200, and 300 pounds per acre) were also applied. All factor-level combinations of these two variables were used in the experiment; these were randomly assigned in a CRD with 3 observations (trees) per treatment combination. At the end of a 2-year period, the increase in diameter was measured for each tree. The data are stored in the file `orange.txt` .

   (a) State formally the appropriate statistical model for the design. Define each term in the model.

   (b) Construct an interaction diagram for this experiment and interpret the results.

   (c) Present the relevant ANOVA table for the given data and perform all the relevant tests based on the ANOVA table. Interpret the results.

   (d) Use Fisher's (protected) LSD to determine how the response differs (significantly) based on the levels of calcium and/or pH.

   (e) Produce a residual plot and comment.

3. This question is an example of a question that could appear on the in-class midterm. You are to answer it without using the computer.

   An experiment was conducted to assess the effect of various factors on the quality of vanilla ice cream. Three factors were of interest: amount of sugar (high or low), blending time (short or long), and use of an egg flavoring substitute (present or absent). The experiment was randomized as follows: for each of the 8 treatment combinations, two tickets were made. All 16 tickets were put in a hat, and each morning a ticket would be drawn from the hat at random and a batch of ice cream would be made according to the combination on the ticket drawn from the hat. The experiment took 16 days to complete. On each day, after each batch was produced, an assessment of flavor was made on a 100 point scale — the larger the score, the better the flavor.

   The data are given below:

Sugar:	Low	Low	Low	Low	High	High	High	High
Blending Time:	Short	Short	Long	Long	Short	Short	Long	Long
Egg Flavoring:	Yes	No	Yes	No	Yes	No	Yes	No
Results:	14, 21	60, 65	32, 39	35, 37	27, 17	71, 67	38, 38	49, 51

   (a) Write down the model for this experiment, defining all terms.

   (b) Using graphical methods, assess whether there is evidence of a three-way interaction between sugar amount, blending time, and use of egg flavoring. (No formal inferences are required for this part.)

   (c) By using an appropriate contrast, perform a formal test for the presence of a main effect for sugar amount. (You need only test the main effect for this question — you are not required to test interactions first.) You may use the fact that, in the ANOVA table, SSError = 123.2.

   (d) Using graphical methods, assess whether there is evidence of a two-way interaction between sugar amount and use of egg flavoring. (No formal inferences are required for this part.)

   (e) In an experiment involving the production of *chocolate* ice cream, only use of sugar and blending time were varied. There were three levels of sugar: low, medium, and high, and there were three times used: short, medium, and long. The combinations used are listed in the table below, along with the resulting data. The experiment was randomized by placing 9 tickets in a hat representing the nine combinations listed below, and then each day a ticket was drawn from the hat and the appropriate batch of ice cream was made.

Sugar:	Low	Low	High	High	Medium	Medium	Medium	Medium	Medium
Time:	Short	Long	Short	Long	Medium	Medium	Medium	Medium	Medium
Results:	35	37	49	61	45	51	44	39	49

   Using only these data on chocolate ice cream (ignore the data on vanilla ice cream) perform a formal test to see whether there is evidence of a two-way interaction between sugar amount and blending time.

4. (Use `R` for this problem.) There are several different treatments that might be useful in apple orchards in terms of increasing yield. This experiment is concerned with two different treatment factors. First, it is thought that planting certain flowers near an apple tree can have a positive, although indirect, effect on yield. This is thought to work because the flowers attract insects to the neighborhood of the tree, and then these insects prey upon and kill other insects that damage the tree. Second, it might be useful to mow the grass around the tree. If this is useful, the effect is probably also indirect — it is thought that the grass around an apple tree competes with the tree for nutrients and water. Therefore, cutting the grass allows the tree to be less stressed, and thus the tree can produce more and better fruit.

   This experiment was designed to examine the effects of these factors. Six treatments were used:

   (a) Control: no special treatment is used.

   (b) Mowing: the grass around the tree is mowed once every two weeks. No flowers are planted.

   (c) Flowers: a ring of flowers is planted around the tree. The ring of flowers is about 2 feet wide, and has a radius of about 8 feet. The grass is not mowed.

(d) Mowing and Flowers: a ring of flowers is planted, as above, and, as well, the grass around the tree is mowed once every two weeks.

(e) Mowing, variation 1: the grass around the tree is mowed once every three weeks. No flowers are planted.

(f) Mowing, variation 2: the grass around the tree is mowed once every four weeks. No flowers are planted.

The experiment was conducted as a randomized complete block design. There were four blocks. Each block contained six trees, spaced well apart. Within each block the treatments were randomly assigned to the trees in the early spring. At the end of the season (in the fall) the data were collected. For each tree, two branches were randomly chosen. From each branch, 10 apples were randomly sampled and weighed. The data are presented below and are also in file .

Note that the quantity "branch total" refers to the total weight of the 10 apples for a given branch. The last two columns of the table give the branch total for the first and second branches sampled from a tree. For this analysis you should analyze the sum of the weights for the two branches as the dependent variable.

Because of limits of time, the data on blocks 1 and 3 were sampled on the first day of data collection, and the data from blocks 2 and 4 were sampled on the second day. You should also be aware that blocks 1 – 3 are on elevated ground, while block 4 lies in a low region. Finally, the trees in block 1 are 12-year-old trees and are therefore slightly younger than the trees in blocks 2 – 4, which are 13-year-old trees.

(a) Conduct an overall F test for the 6 treatment means, make appropriate residual plots, and comment.

(b) Of the six treatments, which would you recommend in terms of maximizing the yield? Do you think that there is more than one treatment that would do a good job?

(c) To what extent is there evidence of an interaction between planting flowers and mowing?

(d) Briefly explain why it is appropriate to analyze the sum of the weights for the two branches as the dependent variable.

		Branch Total	
Block	Treatment	First Branch	Second Branch
1	1	2072	2383
1	2	2195	2368
1	3	1909	2022
1	4	2329	2583
1	5	1972	1796
1	6	1876	2082
2	1	1912	2073
2	2	2508	2094
2	3	1677	1824
2	4	2529	2261
2	5	1739	2066
2	6	1381	1955
3	1	1647	1994
3	2	2028	1437
3	3	1659	2171
3	4	2420	2315
3	5	1753	2275
3	6	1728	1992
4	1	1673	1803
4	2	1523	1810
4	3	2543	1675
4	4	2310	2358
4	5	1681	1618
4	6	999	1456