COMPREHENSIVE WRITTEN EXAMINATION, PAPER III
PART 2: FRIDAY AUGUST 18, 2023 1:00 P.M.–5:00 P.M.
STOR 664 Data Analysis Question (50 points)

Format of the Exam. You have 4 hours to complete this exam, which consists of parallel questions set by Professor Smith (STOR 664) and Professor Li (STOR 665). You are expected to use your own laptop using R; other computer languages are also allowed but the questions are designed to be completed in R. Answers may be written in R Markdown, MS-Word, Latex or any other word processing system but it is important that the answer you hand in shows clearly what your answer is and how it was derived; unannotated computer output will earn very low marks. If you wish to write out part of your answer by hand, that is also allowed but in that case the solution should either be handed in in person or else scanned and emailed. Your completed solution to this problem should be emailed to Professor Smith at rls@email.unc.edu. You are allowed to use the help features in R or the World Wide Web if it is for the purpose of looking up the syntax of a command in R (or some other computer language). You should not use the Web for help with this specific dataset and use of AI tools (e.g. ChatGPT) is strictly forbidden. No communication is allowed during the exam with any other individual whether inside or outside the exam room; however, questions of clarification may be addressed to Professor Smith at the above email address or by phone or text at the number provided. Answer all parts of the question.

The "pizza dough" experiment consists of 28 measurements of the expansion coefficient of pizza dough (i.e. how much the crust expands when the dough is baked). The experiment is carried out over 7 days and using 15 different recipes for the dough (the "treatment"). The data are given in Table 1 and may be read into R through the command
`read.csv('https://rls.sites.oasis.unc.edu/s664-22/PizzaDough.csv')`.

Observation	Day	Treatment	Response	Observation	Day	Treatment	Response
1	1	1	15	15	4	14	11.4
2	1	8	14.8	16	4	10	11.2
3	1	9	13	17	5	11	13
4	1	9	11.7	18	5	15	11.1
5	2	9	12.2	19	5	3	10.1
6	2	5	14.1	20	5	13	11.7
7	2	4	11.2	21	6	1	14.6
8	2	9	11.6	22	6	6	17.8
9	3	2	15.9	23	6	4	12.8
10	3	3	10.8	24	6	7	15.4
11	3	8	15.8	25	7	2	15
12	3	5	15.6	26	7	9	10.7
13	4	12	12.7	27	7	7	10.9
14	4	6	18.6	28	7	9	9.6

Table 1. Response in pizza dough experiment for each of days 1–7 and treatments 1–15

(a) Do the responses appear to follow a normal distribution? Use both graphical and formal tests of fit to state your answer. [7 points.]

(b) Analyze the data as an analysis of variance experiment in which both Day and Treatment are treated as factor variables and there is no interaction. Use appropriate tests to decide whether these effects are statistically significant. What are your conclusions? [11 points.]

(c) Which Treatment has the highest response after correcting for the Day effect? Using Tukey's Honest Significant Difference or some other statistical test of your own choosing, comment on whether this treatment is significantly better than the alternatives. [10 points.]

(d) Do the residuals from this model appear to follow a normal distribution? If not, what is the problem? What does this tell you about the design of the experiment? (In other words, based on the information so far, do you think it was a good or a bad design, and why?) [7 points.]

In fact, the "treatment" in this experiment is really three separate treatments, labelled `x1, x2, x3`, each of which is a factor variable with three levels labelled –1, 0 and 1; see Table 2. You can read in this table by `read.csv('https://rls.sites.oasis.unc.edu/s664-22/tab2.csv')`.

Treatment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x1	–1	–1	–1	–1	1	1	1	1	0	–1	0	0	1	0	0
x2	–1	–1	1	1	–1	–1	1	1	0	0	–1	0	0	1	0
x3	–1	1	–1	1	–1	1	–1	1	0	0	0	–1	0	0	1

Table 2. Combination of x1, x2, x3 factor variables in each of the 15 treatments

(e) Reanalyze the data with an additive analysis of variance model in which `x1, x2, x3` and Day are all considered as factor variables. What are your conclusions now? In particular, what are the optimal values of each of `x1, x2, x3`, and can you say whether these are significantly better than the alternatives? [15 points.]