AMS 582, FALL 2021

Computer Project

Each student will work individually. Your report is due on Tuesday, November 30, at 11:59 pm, Stony Brook time.

Each model has ten independent variables (called *A* through *J*, including *I*). The value of each variable, x_i , satisfies $-1 \le x_i \le 1$, for $i = 1,2,\ldots,10$. The dependent variable *Y* is given by $Y = f(a,b,\ldots,j) + \varepsilon$. That is, $f(a,b,\ldots,j)$ is potentially a function of up to ten variables. There is a belief that the model will involve five or fewer independent variables (but of course, that belief may not be correct). Your assignment is to determine $f(a,b,\ldots,j)$ by designing an experiment, requesting the values of the dependent variable generated by the model for a specified set of runs, and analyzing the data. The experiment that you design consists of a collection of runs. A run is one setting of $(a,b,\ldots,j)$

Design

You will create a spreadsheet that contains your Stony Brook identification number, and a series of lines specifying the runs to be generated. You should send the spreadsheet (call it request1ID123456789) to the TA at the e-mail account ziheng.chen@stonybrook.edu. For example, the spreadsheet for student 1 (who has SB ID number 123456789) has the format below:

Student 123456789										
Setting	A	B	C	D	E	F	G	H	I	J
	-1	1	1	-1	1	-1	1	1	-1	1
2	1	-1	1	1	-1	1	-1	1	1	-1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

A spreadsheet with the values of the independent variable and a realization of the dependent variable for those settings will be emailed to you within two business days after you submit your spreadsheet. For your TA's convenience, please save your spreadsheet in a .csv file. You can achieve this by clicking save as .csv file in Excel.

You will be charged for data using the following cost structure:

1. If the values of the independent variables in a run are either -1 or 1, the cost of the run is 1 point.

2. If there is at least one 0 value in the run and all other independent variables have a value -1 or 1, the cost of the 0 runs is 3 points.
3. If at least one of the values of an independent variable x_i are such that $-1 < x_i < 1, x_i \neq 0$, the cost of the run is 5 points for each value different from -1, 0, or 1.
4. The total cost of your first request is the sum of the costs for the runs requested.
5. The cost of a second request for runs is 100 points plus the costs for the runs requested. The cost of a third request is 400 points plus the sum of the costs for the runs requested. The cost of a fourth request is 600 points plus the sum of the costs for the runs requested. You may not make more than four requests for data.
6. There is no limit to the number of runs you can specify in each request. Of course, you must pay for each run, and there may not be enough time to generate and analyze more than 3 sets of data.

Report

Your report is due on December 4 and should include

1. An Introduction Section in which you specify the rationale of your design.
2. A Method section in which you precisely specify the design and conditions for getting supplemental data if any and your protocol for validating your model.
3. A Results section that includes: 1. your estimated model; 2. a table that specifies whether each variable does or does not affect the setting of the dependent variable; 3. your analysis of variance table; and 4. your documentation of the validity of the model that you report. Include all the data that you requested in your final analysis. Not using all the data requested will be significantly penalized.
4. Your Conclusions and Discussion section. This section includes the limitations of your study.

More specifically, you must indicate whether each of the fourteen variables is in or out of your estimated model.

VARIABLES	Model
A	In or Out
B	In or Out
C	In or Out
D	In or Out
E	In or Out
F	In or Out
G	In or Out
H	In or Out
I	In or Out
J	In or Out

Grading: You get 150 points for each correct assessment of variables. If an interaction is present, you also get points for detecting an interaction. You must specify the variables in the interaction. You lose points for each incorrect variable specified. The cost of your runs will also be deducted from your score. Your point score will be transformed to a scale that ranges from 0 to 300 points.

Reports that are submitted by November 30 will be graded and returned at the final examination.