Experimental Designs: Project 2

Instructor: Xin Qi

Due on Thurs., Nov. 29

General Descriptions

A special type of cell is incubated in a biology lab. The amount of cells are measured after a specific time period. Given the same initial amount of cells, the final amount of cells may vary due to factors such as temperatures, diluted concentration, and so on. The experimenters want to choose a combination of the levels of these factors such that they can obtain the maximum final amount of cells. They choose five factors which may affect the amount of cells. The factors are denoted by A, B, C, D and E. Two levels are selected for each factor and are denoted by "+" and "-".

First, the one-half fraction of the 2^5 experiments were performed. The data of this experiment was stored in the file named "project_data_1.csv". Then the follow-up experiments with 16 runs were performed. The full fold over technique was used in the design of the follow-up experiments. The data of this experiment was stored in the file "project_data_2.csv"

After obtaining some financial support, they wanted to do a full factorial experiment with two replicates. However, due to the limitations of incubators, only 16 experiments can be run in the same day. Hence, the 64 runs were assigned to four days. The assignment of the first replicate was based on the effect ABCDE which was selected to be confounded with block effects. The assignment of the second replicate was based on ABCD. The data of this experiment was stored in the csv file named "project_data_3.csv".

Finally, they purchased a new incubator which allows 320 experiments simutaneously. Hence, they designed a full factorial experiment with 10 replicates. The data was stored in the csv file named "project_data_4.csv".

The purpose of this project is to analyze the data and find the best combination of these factors which can produce the maximum amount of cells.

Specific Requirements

(1). Analyze the data from the fractional design (in "project_data_1.csv"). Find the defining relation and all aliased effects. Make appropriate assumptions, identify important effects and fit the final model by using the coded variables. Make the model checking.

(2). Combine the data from the follow up experiments (in "project_data_2.csv") with the original data (in "project_data_1.csv"). Analyze the combined data, fit the final model by using the coded variables and make the model checking. Compare the model with that in (1).

(3). Analyze the data from the block design (in "project_data_3.csv"), fit the final model by using the coded variables and make the model checking. Compare the model with that in (1) and (2).

(4). Analyze the data from the full factorial design (in "project_data_4.csv"), fit the final model by using the coded variables and make the model checking. Compare the model with that in (1), (2) and (3).

(5). Based on the analysis in (1)-(4), choose a good final model. Then based on this model, recommend the best combination of these factors which can produce the maximum amount of cells.

Project Evaluations

This project is worth 20 points. I will evaluate your project on the basis of the following elements:

- **Everybody must do the project independently. Discussion with other students is not allowed**.
- Your data representation. You can use any graphic methods to explore the data such as boxplots, scatter plots and so on. Make comments on the plots (do not just attach the plots without any explanations).
- Your choice of the defining relation in (1). Choose the correct defining effect.

- Your data analysis. You can use any techniques such as ANOVA, normal probability plot, Lenth's method and so on, to analyze the data. Different people may choose different final models. As long as you can justify your choice, you will get the credits.
- Your interpretation on the final model and your model checking. Make reasonable interpretations, draw reasonable conclusions and make reasonable recommendations based on the analysis.
- Your final written report. Make clear presentation of your report. Remember that this is a project report intead of a homework assignment. It should be understood by experimenters. Attach all your R codes in Appendix (do not put them in the main part of your report).