Assignment: Regression Analysis

You have been asked to develop a response surface model that describes microbial growth rate as a function of environmental conditions: temperature, water activity, and pH.

* The upper and lower limits on temperature are 16C and 24C,
* The upper and lower limits on pH are 5.6 and 6.4.
* The upper and lower limits on water activity are 0.982 and 0.99.

A full response surface model for three input variables is given by:

Analyse the regression model to determine whether all the input variables and corresponding interaction terms are significant. The best model is the simple possible model.

**Given:** A script file called dataGeneration will be used to generate the data. You need to specify your group number in this file. The experimental conditions must be provided in the excel spreadsheet called ExperimentalConditions. Each row in this spreadsheet corresponds to one experiment.

Make sure all the files are in the same folder. Do not change the headers in the excel sheet. You might not be able to run the script file if the excel sheet ExperimentalConditions is open. Once you run the dataGeneration script, a excel sheet with your group number and the time at which you run this script will appear. This excel sheet contains the data you will use for modelling. The first three columns are the experimental conditions, and the last column is the growth rate at these conditions.

**Part 1:** For the first part, you can choose up 30 different experimental conditions. Each experimental condition is a triplet {Temperature, pH, Water Activity}. Choose the conditions carefully. The model depends on the experiments you perform. Once you generate the data for these conditions, you will develop the response surface model. Analyse it to see which terms are significant (95% significance). Report the final model along with the 99% confidence bounds on the parameters. You can run the dataGeneration script as many times as you like with different experimental conditions. Report only the best model you find.

**Part 2:** For the second part, you will perform experiments designed using a full-factorial experiment design. Each experimental condition is a triplet {Temperature, pH, Water Activity}. The design is provided in Table 1 below. The experimental conditions provided in this table are normalised between -1 and 1. To use this design, you will first have to convert this normalised table to experimental conditions mentioned above. Once you generate the data for these conditions, you will develop the response surface model. Analyse it to see which terms are significant (95% significance). Report the final model along with the 90% confidence bounds on the parameters.

**Part 3:** For the third part, you will perform experiments designed using a central composite experiment design. Each experimental condition is a triplet {Temperature, pH, Water Activity}. The design is provided in Table 1 below. The experimental conditions provided in this table are again normalised between -1 and 1. To use this design, you will first have to convert this normalised table to experimental conditions mentioned above. Once you generate the data for these conditions, you will develop the response surface model. Analyse it to see which terms are significant (95% significance). Report the final model along with the 95% confidence bounds on the parameters.

**Part 4:** Write a report in which you detail the steps taken to analyse the regression. How did you ensure that the assumptions are satisfied? You have developed three models using three different experimental designs. In the report also answer the following questions:

* Which of these models is the most accurate?
* How do you determine this?
* If you had to choose one of the three designs, which one would you choose and why?
* In the Central Composite Design, you see the conditions {0,0,0} repeated 3 times. What purpose do you think this repetition serves?
* Use to best model to estimate the growth rate at Temperature of 17C, pH of 6, and water activity of 0.985. Then perform an experiment at these conditions and get the experimental growth rate. Are these two the same. Explain why or why not. Repeat the same experiment multiple times (100 or more). Check the statistics, plot histograms. What do you notice?

**The report must not be more than 6 pages** (including figures). Make sure the figures are legible, well labelled, and clear. The report **must be a PDF file** (not .docx or anything else). You will submit the MATLAB code and the data files you generated (only the files you use for regression) along with the report. Make sure the code is well commented.

The final submission is a .zip file which contains the PDF, the .m file which generates all the figures used in the code, and all the datasets you generated. **Check the folder before you zip it. Running the script should work. If any errors pop up you will be marked down.**

You can read more about Design of Experiments in Chapter 4, Section 3 of Q-Safe book.

**Use of GenAI**

Ofcourse several GenAI (e.g., ChatGPT, Microsoft Copilot) tools now exist which can help you solve this assignment. These tools exist to be used. However, they are meant to help you fulfil a task, not do the task for you. As such you are allowed to use GenAI to, for example, make code snippets (not the full code), proofread your report, help you make better plots, etc. The blind use of these tool is forbidden. The report will be handed in to KULeuven’s plagiarism detection tool which can indicate use of GenAI (even if re-written by, e.g., quilbot). Detection of plagiarism leads to an automatic fail on the module. Same with the code, use of GenAI is easily noticed. If suspected of using GenAI, you will be called in for an Oral exam.

**Collaboration**

In general, you are encouraged to collaborate with other groups on the assignment. However, of the assignment must be yours. Exchanging code is forbidden and counts as plagiarism.

**DEADLINE: December 1, 2024 23.59h**

**Full Factorial Design**

|  |  |  |  |
| --- | --- | --- | --- |
| Experiment # | Temperature | pH | Water Activity |
| 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 0 |
| 3 | 1 | 1 | -1 |
| 4 | 1 | 0 | 1 |
| 5 | 1 | 0 | 0 |
| 6 | 1 | 0 | -1 |
| 7 | 1 | -1 | 1 |
| 8 | 1 | -1 | 0 |
| 9 | 1 | -1 | -1 |
| 10 | 0 | 1 | 1 |
| 11 | 0 | 1 | 0 |
| 12 | 0 | 1 | -1 |
| 13 | 0 | 0 | 1 |
| 14 | 0 | 0 | 0 |
| 15 | 0 | 0 | -1 |
| 16 | 0 | -1 | 1 |
| 17 | 0 | -1 | 0 |
| 18 | 0 | -1 | -1 |
| 19 | -1 | 1 | 1 |
| 20 | -1 | 1 | 0 |
| 21 | -1 | 1 | -1 |
| 22 | -1 | 0 | 1 |
| 23 | -1 | 0 | 0 |
| 24 | -1 | 0 | -1 |
| 25 | -1 | -1 | 1 |
| 26 | -1 | -1 | 0 |
| 27 | -1 | -1 | -1 |

**Central Composite Design (Inscribed)**

|  |  |  |  |
| --- | --- | --- | --- |
| Experiment # | Temperature | pH | Water Activity |
| 1 | -0.594603558 | -0.59460356 | -0.594603558 |
| 2 | -0.594603558 | -0.59460356 | 0.594603558 |
| 3 | -0.594603558 | 0.594603558 | -0.594603558 |
| 4 | -0.594603558 | 0.594603558 | 0.594603558 |
| 5 | 0.594603558 | -0.59460356 | -0.594603558 |
| 6 | 0.594603558 | -0.59460356 | 0.594603558 |
| 7 | 0.594603558 | 0.594603558 | -0.594603558 |
| 8 | 0.594603558 | 0.594603558 | 0.594603558 |
| 9 | -1 | 0 | 0 |
| 10 | 1 | 0 | 0 |
| 11 | 0 | -1 | 0 |
| 12 | 0 | 1 | 0 |
| 13 | 0 | 0 | -1 |
| 14 | 0 | 0 | 1 |
| 15 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 |