

## Use of CAT for D-Optimal design

The freeware **Chemometric Agile Tool software** CAT written in R can be downloaded from <https://www.gruppochemiometria.it/index.php/software/19-download-the-r-based-chemometric-software>.

If you have R on your computer, this may be another version of R and may cause trouble.

### 1. Find the runs to make using CAT

- 1) Create the *Candidate set* using the workbook *D-Optimal design* and save this workbook.
- 2) Open CAT (loading may take some time...wait for the menus to be populated) and load the data:
  - Choose *Data Handling* -> *Load* -> *XLS/XLSX*
  - In the box select the workbook with the *Candidate set* and press *Open*
  - In the box enter a short *Matrix Name* e.g. "cs" for candidate set.
  - Click *Row Names* and then *OK*
- 3) In the *R Console* enter the matrix name after the last > and the matrix will be displayed. Check that it looks ok.
- 4) Run the D-Optimal design routine:
  - Choose *MLR-DOE* -> *D-Optimal Design* -> *Independent var.* -> *D-Optimal by Addition*
  - In the box enter the matrix name and unclick *Higher terms* and click *Ok* (normally we do not have enough runs to include interactions)
  - In the next box change the *Upper number of experiments* to a reasonable number, e.g. 2 times the number of run you want to perform and click *Ok*.
- 5) Select the appropriate solution with a suitable number of experiments:
  - Look at the two diagnostic plots:
    - Maximum Inflation Factor should be less than 2 and at least < 4
    - Log(Normalized Determinant) should have a high value
  - Go to the *R Console* and copy the string with the number of runs below the heading *Selected Points*: Copy only the numbers, not including the [1]
- 6) Go back to the *D-Optimal workbook* sheet *Generate runs* and paste the string with the numbers to *Table 1* and press *Update Table 2*
- 7) Below *Table 2* there is diagnostics about confounding for the selected solution. Check these results and decide whether or not to try to modify the problem and test a new solution from CAT. It may be a good idea to try different plans.
- 8) When the solution has been accepted save the workbook *D-Optimal design* and use it for analysis of the results when the experimental responses are available – or use CAT to calculate results, see next section.

## 2. Calculate the results using CAT

The *D-Optimal design* workbook can in most cases calculate the results, but the present version of the workbook does not include calculation of models including interactions and higher degree terms. In such cases the results can be calculated in CAT.

Another thing is that CAT gives you more options for additional diagnostics and graphs, but this is outside the scope of this simple introduction.

To calculate results using CAT:

- 1) Copy the design matrix from the *D-Optimal design* workbook sheet *Results*:
  - Copy the upper left corner of *Table B* with the actual design matrix to an unused part of the same sheet including the factor names
  - Copy the responses *Y* from the right part of the table including the heading *Y* and paste it to the right of the design matrix from the previous point
  - Copy the whole matrix including headings and responses
- 2) Open CAT and calculate the results:
  - Choose *Data Handling* -> *Load* -> *Copy and Paste*
  - In the box enter a name for the matrix and press *Ok*
  - In the next box click on the “0” and paste the matrix into the area
  - A new box opens, check *Use Header Row* and click *Ok*
- 3) In the *R Console* enter the matrix name after the last > and the matrix will be displayed. Check that it looks ok.
- 4) Choose *MLR-DOE* -> *Model Computation* -> *Independent Variables*
  - In the box enter the following parameters and click *Ok*:
    - Matrix name
    - Rows to be selected, normally “All”
    - X-Variables...: The first *k* columns with X-values e.g. 1:5
    - Y-Variable: The column with the responses *Y*
- 5) In the box select *Residuals* and click *Ok*
- 6) The results are calculated and shown in the *R Console*. The results contain:
  - A table with the coefficients with statistics
    - Headings: *b0* is the intercept
    - Coefficients in the model
    - Standard deviations of the coefficients
    - Confidence interval for the coefficients
    - P-values. If  $P < 0.05$  the coefficient is statistically significant different from 0
  - Many other statistics, explained variance, residuals etc.
- 7) CAT may be used to generate many plots etc. but this manual does not cover that.
- 8) What you need for the course is the coefficients and maybe some of the statistical parameters. Copy them and use them in Excel and for reports and presentations.