# Short manual for workbooks for DoE with Factorial Designs and D-Optimal designs

## 1 Introduction

The workbooks are intended for the lab-course in sample preparation, where 4 to 8 groups named A-H make the design, experiments and analysis of the results. The workbooks can be used for other projects where some of the functions or inputs are not needed and can be neglected.

In general the purpose of DoE is to investigate the effect of a number of controlled factors each at a number of levels, with a reasonable number of experiments (runs). A full factor design – all factors at all levels – will in many cases require too many runs, i.e. 4 factors at 3 levels is $3^4 = 81$ runs, often more than time or resources allow, so the task is to find a good balance between number of runs and confidence in the results.

A basic understanding of DoE, resolution, confounding etc. is recommended. For 4 factors at 3 levels a half fractional factorial design $2^{4-1} = 8$ runs + 4 center points = 12 runs would be a good solution to the example with a resolution of IV meaning that main factors are not confounded with 2-factor interactions.

**What to do?**

1) Try to find a suitable design for all the factors and levels to be investigated (Factorial or D-Optimal design, see section 4-5 or 6-7 below).
2) If all factors from 1) cannot be included in the design at all levels, try to exclude some factors and/or levels or to split the design into more than one design.
3) If the design is less than the possible number of runs select conditions for extra runs. The extra runs can be used to check the effect of parameters excluded in 2) or to test a 3rd level of a parameter or for center points or repeat measurements.
4) Open the Excel workbook for the design, *R4, R8, R16 or D-Optimal design*, fill in the design in all tables and save a copy of the design with a new name. For D-Optimal designs the CAT software must also be used to find the runs to make.

## 2 Labelling of experiments

For other uses than at the course you can use your own labelling with the following exceptions:

- For R4, R8 and R16 designs the last 3 characters in the label must be the run-number with 2 digits (01 – 02 – 03 etc.) followed by a letter, e.g. 01a, 02b, 03c… etc.
- For D-Optimal designs, the labels must be in ascending number order when sorted with the Excel Sort function, i.e. run 1, 2, 3, 4….etc.

Example of a valid label at the course: **SPM1208D**

| | |
|---|---|
| **SPM**: | The 3 letter code for the technique. |
| **1**: | The design number. Normally only one design will be made, but sometimes 2 or 3 designs with 4 runs or 2 designs with 8 runs are used. Use a new workbook for each design. |

**2**: The replicate number. In most cases this will be 1 because only one run is used.
**08**: The run number in the DoE. Two digits in the range 01 to 36.
**D**: The group that performs this run (OBS: not the group that made the design).

Using these labels for the runs **and for the samples to be analysed** makes an automatic sort of the results easy. It is IMPORTANT to label all samples correctly in the lab.

## 3 Randomization and blocking:

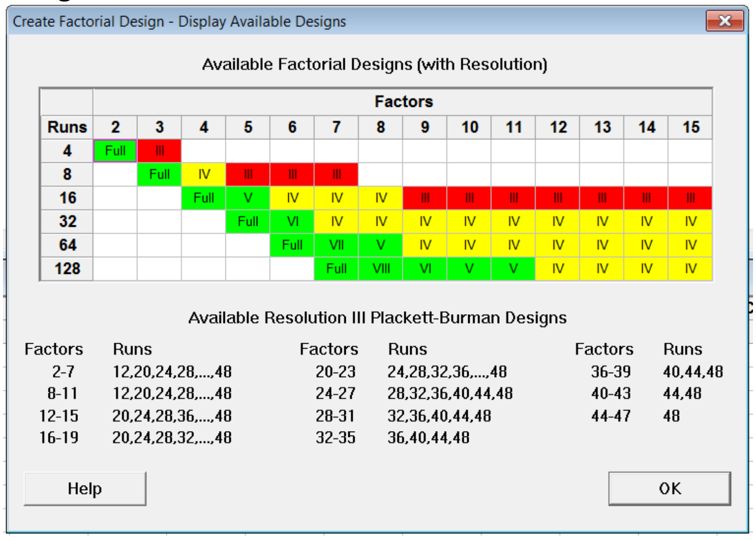
All the runs should be run in random order and under the same conditions as far as possible. This means that it is a bad idea to run all high levels one day and all low levels another day. On the other hand, it may be easiest to run several runs with similar settings. You must find the best compromise between a random run order and a rational run order.

If runs are made on two days: One way to divide runs between days is to use extra factors in the design for blocking between days. If you make a design $2^{6-2}$ with 6 factors and 16 runs, then you could make a $2^{7-3}$ design and use the $7^{th}$ factor as the day (block) number. The resolution will still be IV, and you can see if there is any effect of the day. For full factor designs ($2^2$, $2^3$ and $2^4$) you may use the interaction of all factors (**12**, **123** or **1234**) for blocking.

In the theory for DoE there are systematic methods for blocking of runs in order t avoid confounding of effects and run conditions.

## 4 Factorial Experimental designs

**Designs to choose from:**



Create Factorial Design - Display Available Designs

Available Factorial Designs (with Resolution)

| | Factors | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Runs** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| **4** | Full | III | | | | | | | | | | | | |
| **8** | | Full | IV | III | III | III | | | | | | | | |
| **16** | | | Full | V | IV | IV | IV | III | III | III | III | III | III | III |
| **32** | | | | Full | VI | IV | IV | IV | IV | IV | IV | IV | IV | IV |
| **64** | | | | | Full | VII | V | IV | IV | IV | IV | IV | IV | IV |
| **128** | | | | | | Full | VIII | VI | V | V | IV | IV | IV | IV |

Available Resolution III Plackett-Burman Designs

| Factors | Runs | Factors | Runs | Factors | Runs |
|---|---|---|---|---|---|
| 2-7 | 12,20,24,28,...,48 | 20-23 | 24,28,32,36,...,48 | 36-39 | 40,44,48 |
| 8-11 | 12,20,24,28,...,48 | 24-27 | 28,32,36,40,44,48 | 40-43 | 44,48 |
| 12-15 | 20,24,28,36,...,48 | 28-31 | 32,36,40,44,48 | 44-47 | 48 |
| 16-19 | 20,24,28,32,...,48 | 32-35 | 36,40,44,48 | | |

Help OK

**Interactions:** Factors are likely to interact. Especially in chemistry, it is often unlikely that one factor (e.g. temperature) has the same effect for two levels of another factor (e.g. solvent).

- Use only green or yellow designs, i.e. resolution IV or better. This allows investigation of:
  - 4 runs: 2 factors
  - 8 runs: 3 or 4 factors
  - 12 runs: 3 or 4 factors

- 16 runs: 4 to 8 factors
- 32 runs: 5 to 15 factors.

**Additional runs:** Most designs have a core of 4, 8 or 16 runs. The additional runs can be used for center points, for repeat measurements of one or more of the runs, or for additional measurements at a setting that was not included in the design itself, i.e. for a factor that was kept constant (e.g. the 2 additional runs without acid in addition to 6 runs with acid). You can use the additional runs for:

- Center points (to determine standard deviation and check linearity)
- Repeat runs (to determine standard deviation and check repeatability)
- Extra runs to check an assumption or variation or a new level not included in the design itself.

**Center points:** It is recommended to add center points if there are enough runs. For some factors, you cannot make one center point, i.e. for the factor "add acid", because either you add acid or not. In such cases you can either:

- Make 2 center point levels, e.g. 2 points with acid and 2 points without acid.
- Partial design: Make a design for the expected best level for this factor and use a few extra runs to check that the other level indeed was worse e.g. 6 runs with acid and repeat 2 runs without acid.
- Split the design in two: one with low and one with high level for this factor.

## 5 Excels workbooks for Factorial designs

1) Open the workbook with the basic design you will use (*R4*, *R8* or *R16*).

2) Go the **sheet *RX Design*** with the appropriate number of factors.

3) Fill in or correct the white cells in Table 1 to 5:

   **Table 1:** Basic information in free format.
   **Table 2:** Parameters that must be valid, i.e. name of technique, number of factors etc. For other use than at the course the name of the technique is optional and may be in error without problems.
   **Table 3:** Factors and levels. This is the design itself. Enter the name, unit, low, high and eventually center levels for each factor. The levels for each run will be calculated in the table below. You can enter values for extra runs in the last part of the table. If you want to include factors which are not part of the main design, you can enter information about these factors and their levels in the comment field, *Table 5*. Note that you may have two center points at different levels.
   **Table 4:** Assign group A-D to runs (and if relevant also groups E-H). For a normal design with only one run for each setting enter group names in this table in the first column. If you have 2 replicates enter group names in both columns. Be careful to enter group names for all runs in the design, also for center points or extra runs. The design can be randomized after you entered the group names, but you should only randomize, if you do not use blocking of the runs. Max 6 runs for each group A-H.
   **Table 5:** Comments in free format (1 line of any length). This can be used to give extra information about the specific run, such as constant factors.

4) If you want to randomize the runs click on *Randomize design* now.

5) Finally click on *Make design* and tables with run conditions for all groups will appear in the associated sheet *R4-Runs*, *R8-Runs* or *R16-Runs.* This is only needed at the course.

6) Save the workbook with a new name (Save As…).

## 6 D-Optimal designs

In D-Optimal design, a number of runs to be execute are selected from a *Candidate set* of all possible runs for the selected parameters. There are 4 steps:

1. Create a *Candidate set* with all possible runs to be performed. Use the Excel workbook *D-Optimal design,* see below.
2. Use the software CAT to create a list of solutions to the problem with different number of runs.
3. Select the best solution that is possible with the available number of runs. If no good solution is found, simplify the problem and try again.
4. Copy the run numbers for the selected solution from CAT and paste it into the *D-Optimal design* workbook and finish the workbook as described below.

The following points should be taken into account:

- The maximum number of factors that can be estimated is *N*-1, where *N* is the number of experimental runs (because one degree of freedom is used for the *mean.*)
- The number of factors for a categorical parameter is *k-1*, where *k* is the number of levels, i.e. for the parameter *solvents* with 3 solvents there will be 2 factors (called dummy factors)
- The *Inflation factors* in the output from CAT do only take terms in the model into account, i.e. confounding with interactions not included in the model is not included!
- Diagnostic information about confounding are given in the *D-Optimal design* workbook (see below) and should be considered before the final decision about which solution from CAT to use is taken.

Hints:

- If you include 3 levels in a continuous (numerical) factor, D-optimal designs will only find runs with the low or high level, i.e. the 3$^{rd}$ center level will not be included in the solutions.
- If you want to investigate 3 levels of a numerical factor, e.g. *Add salt (0, 300 or 600 mg)*, this can be done by entering it as a categorical factor with the levels 300, 0 and 600 mg. In this case 2 factors will be generated, one for 0 mg and one for 600 mg, and the result will be for 0 or 600 mg relative to 300 mg (the center).

## 7 Excel workbook for D-Optimal designs

1) Open the workbook *D-Optimal design* and go to sheet *Design*
2) Fill in the parameters and levels in Table 1 and 2 (Table 3 is optional)
3) Press *Create CAT Input*
4) Save the workbook (Save As…)
5) Use the software CAT to create a list of solutions to the problem with different number of runs (see separate manual for CAT).
   - Select the best solution that is possible with the available number of runs.
   - If no good solution is found, simplify the problem and try again.
   - Copy the run numbers for the selected solution from CAT.

6) Paste the string with the run numbers (numbers separated by blanks) into Table 1 in Sheet *Generate runs* in the Excel workbook for D-Optimal design
7) Press *Update Table 2*

8) The runs to be performed are now shown in Table 2. Before you decide to use this solution, **you must check** the information about confounding in 2 tables below Table 2. An example is given here:

| Factor | Res >III | Max | Confounded with | |
|---|---|---|---|---|
| 1 | Yes | 0 | | |
| 2 | No | 0.5 | Some | |
| 3 | No | 1 | 4*5 | |
| 4 | No | 1 | 3*5 | |
| 5 | No | 1 | 3*4 | |

1st column Factor: The factor #

2nd column Res >III: **No** means that this factor is confounded with 2-factor interactions
**Yes** means that this factor is independent of 2-factor interactions

3rd column Max: The maximal off diagonal term divided by the diagonal term:
**0** means that there is no confounding with 2-factor interactions
±**0.XXX** means that there is some confounding with 2-factor inter.
±**1** means that there is 100 % confounding with 2-factor inter.

4th column: The name of 2-factor interactions with 100 % confounding. Or “Some”

In the example Factor 1 is without confounding with 2-factor interactions, Factor 2 has some degree of confounding and Factor 3, 4 and 5 are 100 % confounded with 2-factor interactions.

**It will in general be impossible to avoid confounding but try to minimize it**. Hints:

- If a factor is expected to have strong interactions try to assign this factor to the # with a low degree of confounding (#1 in the example).
- If you include an interaction on the model for CAT, this interaction will have the smallest possible confounding with the other factors (but the whole confounding pattern will shift around and may be worse than without this interaction in the model).

The *Information matrix X’X* is also shown below Table 2 for your information.

9) Fill in or correct the white cells in Table 3, 4 and 5:
**Table 3:** Parameters that must be valid, i.e. design no and name of technique
**Table 4:** Assign group A-D to runs (and if relevant also groups E-H). For a normal design with only one run for each setting enter group names in this table in the first column. If you have 2 replicates enter group names in both columns. Be careful to enter group names for all runs in the design, also for center points or extra runs. The design can be randomized after you entered the group names, but you should only randomize, if you do not use blocking of the runs. Max 6 runs for each group A-H.
**Table 5:** Comments in free format (1 line of any length). This can be used to give extra information about the specific run, such as constant factors.

10) Press *Make runs* and tables with run conditions for all groups will appear in the sheet *Runs.* This is only needed at the course.

11) Save the workbook.