Experimental design (DOE) - Design

Menu:	QCExpert	Experimental Design	Design	Full Factorial Fract Factorial

This module designs a two-level multifactorial orthogonal plan 2^{n-k} and perform its analysis. The DOE module has two parts, *Design* for the experimental design before carrying out experiments which will find optimal combinations of factor levels to gain maximum information at a reasonable number of experiments and part *Analysis* described in the next chapter 0 on page 4, which will analyze results of the planned experiment. The main goal of DOE is to find which of the factors included in the model have considerable influence on one outcome of the experiment. The outcome is called response and it can typically be yield, energy consumption, costs, rate of non-conforming product units, blood pressure etc. Factors are variables which will set for the purpose of the experiment to two values or levels. Factors must have two states („low“ and „high“, or –1 and +1) defined naturally (night – day, male – female) or defined by the user (low temperature = 160°C, high temperature = 180°C). Each state is assigned the value –1 or +1 respectively, regardless of the sign, i.e. formally high temperature may be defined as the „low“ state (–1) and low temperature as the „high“ state with no effect to the result of the analysis. Factors may typically be night and day, cooling off/on, smoker/nonsmoker, clockwise/counterclockwise mixer rotation, etc. The user defines number of factors *n*, fraction *k* of the full experimental plan and number of replications *m* of each experiment. The module will create a matrix of the experimental plan and stores it in a new data sheet in the form of plus and minus ones. Each row in the spreadsheet represents one experiment. The number of rows is $m2^{n-k}$. Factors are named by letters of the alphabet A, B, C, …. Columns defining order of an experiment and replication are also added for reference. The column *Response* is left empty – here the user will enter results *Y* of the carried out experiments for further analysis by the module *Design of Experiments – Analysis*. The result of the analysis will be a set of coefficients of a regression model with all linear and all mixed terms (main effects and interactions).

$$Y = a_0 + \sum a_i \text{comb}\left(A, B, C...\right),$$

for example, with 3 factors *A*, *B*, *C* we have a model with $2^3 = 8$ parameters a_0 to a_7.

$$Y = a_0 + a_1 A + a_2 B + a_3 C + a_4 AB + a_5 AC + a_6 BC + a_7 ABC$$

A, *B*, *C* are the factors, *AB*, *AC*, *BC* are second-order interactions, *ABC* is the third order interaction. The linear terms coefficients (main effects) reflect an influence of the factor level on *Y*. For example, the value $a_1 = 4$ suggests that the high level of factor *A* results in *Y* bigger by 8 units than at low level of *A*. However, to make a final conclusion about the influence of factors the statistical significance of the coefficients must be assessed either by the significance test when $m > 1$, or by the coefficient QQ-plot, see below. Coefficients at mixed terms like a_4AB are influences of one factor conditioned by the level of another factor (interactions). Great value of an interaction coefficient means that the factor influences *Y* differently in dependence on the level of the other factor.

Fractional factorial designs can significantly reduce the number of experiments needed to calculate the coefficients to a fraction 2^{-k} compared to the full fractional design. The fraction *k* can be an integer, generally $0 < k < n$. The number of experiments in such a design will then be $m2^{n-k}$. The price to be paid for such a reduction of the model is aliasing. Each coefficient represents the influence of more than one term of the model, for example a_1 may stand for combination of the influences of the factor *A* and the interaction *AB*, with no possibility to distinguish between there influences. Fractional version of the above model 2^{3-1} with $k = 1$ can thus be written as

$$Y = a_0\left(1 + ABC\right) + a_1\left(A + AB\right) + a_2\left(B + AC\right) + a_3\left(C + BC\right)$$

If the interaction *AB* is assumed to be negligible, we can take a_1 for the main effect of *A*. The summation of main effects and interactions is called aliasing. The goal of fractional design is to try to create a design in which a main effect is aliased only with interaction of the highest possible order, as it is generally known that high order interactions are often not present, therefore the respective coefficient represents indeed the influence of the factor. This goal is sometimes difficult to achieve, especially for high *k*. This module gives the best possible predefined designs in this respect.

Data and parameters

Full factorial design creates a design matrix from the given number of factors *n* and replications *m*. Number of generated rows will thus be $m2^n$. Each row correspond to one experiment. Therefore this design is appropriate for lower number of factors, as the number of experiments needed may get quite high, eg. 1024 experiments for 10 factors without replications ($n = 10$, $m = 1$). In the dialog window (Fig. 1) select the target data sheet in which the design will be written. NOTE: Any contents of this sheet will be deleted, so you should create a new sheet (Menu: *Format – Sheet – Append*). Fill in number of factors and the desired number of replications of each experiment. If the checkbox at *No of replications* is not checked, the number of replications is ignored, $m = 1$ is taken as default. Check the box *Basic information* if you want to basic description of the design in the Protocol sheet. If the *Randomize order* box is checked, the column Order in the target sheet is randomized and after sorting the rows by this column we can obtain rows of the design in a random order, which may help to avoid possible deformation of response from the systematic sequence of similar experiments.

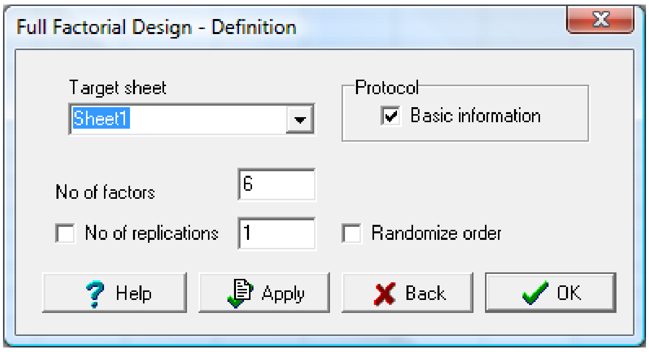


Fig. 1 Full factorial design dialog

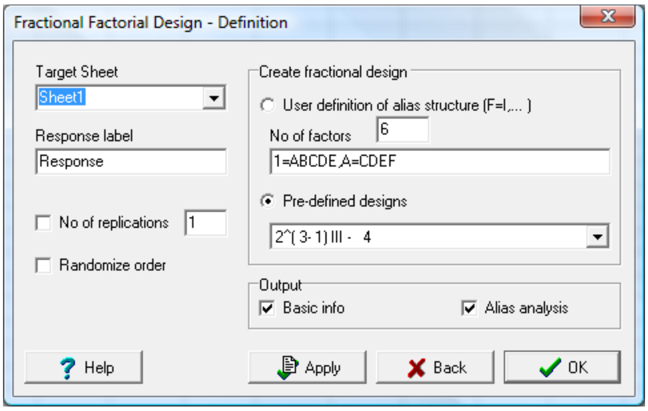


Fig. 2 Fractional factorial design dialog

Fractional factorial design is derived from the full factorial design, but needs much less experiments to estimate coefficients with the drawbacks mentioned above. In the dialog window (Fig. 2) slect the target data sheet in which the design will be written. NOTE: Any contents of this sheet will be deleted, so you should create a new sheet (Menu: *Format – Sheet – Append*). The field Response label will be used to label the response column in the design table. If replications are required fill in the desired number of replications of each experiment. If the checkbox at *No of replications* is not checked, the number of replications is ignored, $m = 1$ is taken as default. Check the box *Basic information* if you want to basic description of the design and Alias analysis if the analysis is to be performed in the Protocol sheet. If the *Randomize order* box is checked, the column Order in the target sheet is randomized and after sorting the rows by this column we can obtain rows of the design in a random order, which may help to avoid possible deformation of response from the systematic sequence of similar experiments. The fractionation is based on the design defining relationships in the form of sufficient alias equalities. They can be written in the *User definition of alias structure* field. The number of relationships is equal to *k*, relationships are separated by comma. There is no easy way to find optimal design definition, as the defining relationship implies other aliases, some of which may disqualify the design. For example, if we attempt to define a 2^{4-1} design for 4 factors A, B, C, D by a defining relationship A = ABD, we will get the alias B = D and will not be able to separate influence of main effects! DO NOT use user definitions unless you are sure they are correct, otherwise they will most probably lead to an unusable or non-optimal plan, with aliases of main effects, such as A = C. It is highly recommended to use predefined designs in the drop-down list *Pre-defined designs* field. The designs are ordered by the number of factors *n* and the fraction *k*. The design descriptions have the following meaning

2	^(	3	-	1	)	III	-	4
		Number of factors *n*		Fraction order *k*		Design resolution		Number of experiments needed

Design resolution is the information gain parameter related to the alias structure. The designs with aliases between main effect and high order interaction are more informative and have high resolution value. The design should be a compromise between the number of experiments and the design resolution.

Table 1 List of pre-defined optimal designs

No	Type of design	Fraction	Resolution	Experiments needed
1	2^{3-1}	3-1	III	4
2	2^{4-1}	4-1	IV	8
3	2^{5-1}	5-1	V	16
4	2^{5-2}	5-2	III	8
5	2^{6-1}	6-1	VI	32
6	2^{6-2}	6-2	IV	16
7	2^{6-3}	6-3	III	8
8	2^{7-1}	7-1	VII	64
9	2^{7-2}	7-2	IV	32
10	2^{7-3}	7-3	IV	16
11	2^{7-4}	7-4	III	8
12	2^{8-2}	8-2	V	64
13	2^{8-3}	8-3	IV	32
14	2^{8-4}	8-4	IV	16
15	2^{9-2}	9-2	VI	128
16	2^{9-3}	9-3	IV	64
17	2^{9-4}	9-4	IV	32
18	2^{9-5}	9-5	III	16
19	2^{10-3}	10-3	V	128
20	2^{10-4}	10-4	IV	64
21	2^{10-5}	10-5	IV	32
22	2^{10-6}	10-6	III	16
23	2^{11-5}	11-5	IV	64
24	2^{11-6}	11-6	IV	32
25	2^{11-7}	11-7	III	16
26	2^{12-8}	12-8	III	16
27	2^{13-9}	13-9	III	16
28	2^{14-10}	14-10	III	16
29	2^{15-11}	15-11	III	16

Table 2 Examples of 2^(5-2) designs

(A) Optimal design	(B) Unusable design, since A=D and B=absolute term
Design definition: D = AB, E = AC A = BD = CE = ABCDE B = AD = CDE = ABCE C = AE = BDE = ABCD D = AB = BCE = ACDE E = AC = BCD = ABDE BC = DE = ABE = ACD BE = CD = ABC = ADE ABD = ACE = BCDE = 1.0	Design definition: A = AB, B = AD A = D = AB = BD B = AD = ABD = 1.0 C = BC = ACD = ABCD E = BE = ADE = ABDE AC = CD = ABC = BCD AE = DE = ABE = BDE CE = BCE = ACDE = ABCDE ACE = CDE = ABCE = BCDE

Protocol

Design type	Full factorial, 2^*n* or Fractional factorial 2^(*n*-*k*).
Design definition	Only for Fractional factorial, defining relationships, e.g.: E = ABC F = BCD
Design description	Only for Fractional factorial design 2^(*n*-*k*), resolution, number of experiments (without replications). For example „2^(3- 1) III - 4“ means 2-level factors, 3 factors in design, half – fraction of the full design, resolution III, 4 distinct experiments.
No of factors	Number of factors
No of replications	Number of replications of each experiment
No of experiments	Number of distinct experiments
Alias-structure analysis	Only for fractional design. Complete listing of all aliases, of grouped combinations of undistinguishable factors and interactions, Aliases described by one coefficient are on one row. For example, if the alias row contains „B AD CDE ABCE“, then the coefficient for the factor „B“ will also include effects of interactions AD, CDE a ABCE. Number „1“ represents the absolute term a_0 in the model. Aliases between factors such as A = C are undesirable, as in that case we have no information about the influence of the factors A and C.

Graphs

This module does not generate any plots.

Experimental design (DOE) - Analysis

Menu:	QCExpert	Experimental Design	Analysis

This module analyses data prepared by the previous module (Experimental Design). It can analyze both full factorial and fractional factorial designs 2^n a 2^{n-k}, with filled in results (responses) of the experiments in the *Response* column.

The main purpose of a designed experiment analysis is to determine which of the factors have significant influence on the measured response. Based on these responses, the module computes the coefficients of the design model using the multiway ANOVA model. If the design does not contain replicated experiments, the resulting model has zero degrees of freedom. In consequence, coefficient estimates do no allow for any statistical analysis, all residuals are by definition zero and significance of factors and/or interactions can only be assessed graphically using the coefficient QQ plot. With replicated experiments the analysis is formally regression analysis, so we can obtain estimates with statistical parameters (variances) and test the significance of factors statistically. It is therefore recommended to replicate experiments where possible.

Data and parameters

An example of the data for the module Design of Experiments – Analysis is shown in Table 3. All data except the *Response* column were generated by the previous module. After setting factors according the design and carrying out all 16 experimental measurements (or responses), the response values are written to the data table and whole table is submitted to analysis.

In the dialog window Factorial Design – Analysis (Fig. 3) the response column is pre-selected. The significance level is applicable only in case of replicated experiment, where statistical analysis is possible. The user can select items to be included in the text protocol output and plots in the graphical output. An advanced user can also write a design manually using the required notation: –1 for low and 1 for high factor level, first 2 columns in data sheet will be ignored, names of factor columns are ignored, factors are always named A, B, C,…, last column is expected to contain measured responses. Incorrect or unbalanced designs are not accepted and may end with an error message. It is recommended however to use designs created by the Experimental design module.

Table 3 Example of data for analysis of a designed fractional factorial experiment 25-2 with 5 factors and 2 replications

Order	Replication	A	B	C	D	E	Response
1	1	-1	-1	-1	1	1	14.6
2	2	-1	-1	-1	1	1	14.5
3	1	-1	-1	1	1	-1	13.6
4	2	-1	-1	1	1	-1	13.6
5	1	-1	1	-1	-1	1	15.1
6	2	-1	1	-1	-1	1	14.7
7	1	-1	1	1	-1	-1	13.2
8	2	-1	1	1	-1	-1	13.3
9	1	1	-1	-1	-1	-1	16.4
10	2	1	-1	-1	-1	-1	16.4
11	1	1	-1	1	-1	1	15.3
12	2	1	-1	1	-1	1	15.1
13	1	1	1	-1	1	-1	14.7
14	2	1	1	-1	1	-1	14.6
15	1	1	1	1	1	1	17.1
16	2	1	1	1	1	1	16.7

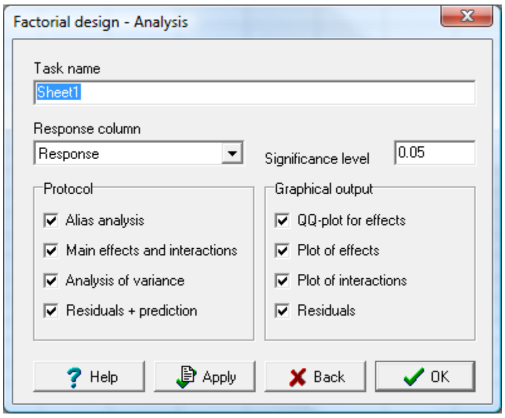


Fig. 3 Dialog window for Factorial design – Analysis

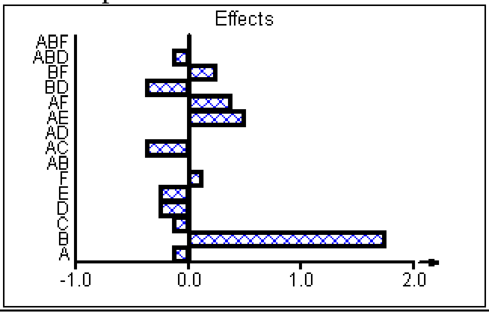
Protocol

Designed experiment analysis	
Design type	Factorial, full design, or fractional design with description in the form 2^(*n*-*k*), e.g. 2^(5-2).
No of factors	Number of factors in the design
No of replications	Number of replications
No of experiments	Total number of experiments (number of data rows)
Design is / IS NOT orthogonal	Information if the design is or is not orthogonal. Orthogonality is one of the requirements for a stable and effective design. All designs generated by QC.Expert™ are orthogonal.
Alias-structure analysis	Only for fractional design. Complete listing of all aliases, of grouped combinations of undistinguishable factors and interactions, Aliases described by one coefficient are on one row. For example, if the alias row contains „B AD CDE ABCE“, then the coefficient for the factor „B“ will also include effects of interactions AD, CDE a ABCE. Number „1“ represents the absolute term a_0 in the model. Aliases between factors such as A = C are undesirable, as in that case we have no information about the influence of the factors A and C.
Main effect values and interactions	Computed values of influences for factors and interactions.
Effect, interaction	Factor or interaction, remember that in fractional design, each factor or interaction listed here is aliased with one or more other interaction and the values are a sum of all aliased influences.
Coefficient	Estimates of main effects, interactions and the absolute term. The absolute term is the expected value of the response when all factors are at the low level. These coefficients are the actual effect of the factors and interactions.
Value	Estimates of parameters of the regression model. As here the factors are represented by values –1, +1, the parameter values are half the effects.
Std Deviation	Standard deviations of regression coefficients can be computed only for replicated experiments. Otherwise, the deviations are zero.
Analysis of variance	Analysis of variance table.
Source	Source of variability.
Total	Total variability of the response $Y - a_0$.
Explained by model	Variability explained by the model.
Residual	Residual variability not explained by the model. This variability is zero for non-replicated experiments.
Influence on variance	Separated average and variability for low (-) and high (+) levels of factors.
Source	
Average(-), (+)	Average response for low (-) and high (+) levels of factors.
Variance(-), (+)	Response variance for low (-) and high (+) levels of factors.
Ratio(+/-)	Ratio of variances at high and low level of the factors. Too high

	or too low value of the ratio may indicate significant influence of the given factor on response variability which can be interpreted as decrease or increase of quality if *Y* is the quality parameter or stability of the response variable.
Residuals and prediction	Table of predicted response and residuals. This table is applicable only for repeated experiments, otherwise responses are the same as measured responses and residuals are zero.
Response	Measured response *Y*.
Prediction	Predicted response Y_{pred} from the computed model.
Residual	Residuals $Y - Y_{pred}$.

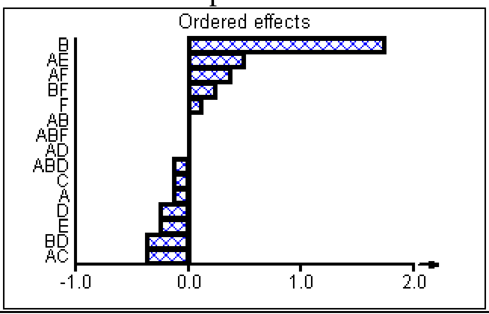
Graphs

Effects plot



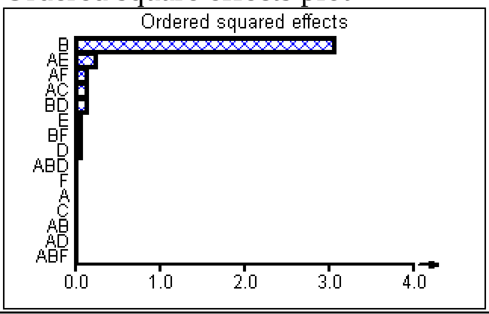
Plot of the computed effects sorted alphabetically and by the interaction order. Greatest values (regardless of the sign) may suggest significant influence of the respective factor or interaction. This plot should be compared with the Effects QQ-plot.

Ordered effects plot



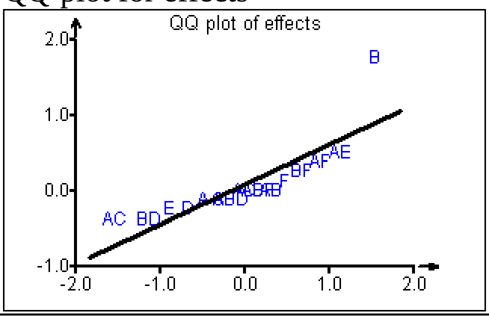
The same as the previous plot, the values are sorted decreasingly.

Ordered square effects plot



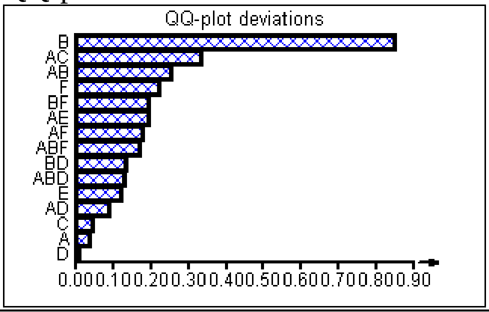
Plot of the squared computed effects sorted decreasingly. Greatest values may suggest significant influence of the respective factor or interaction. This plot should be compared with the Effects QQ-plot.

QQ-plot for effects



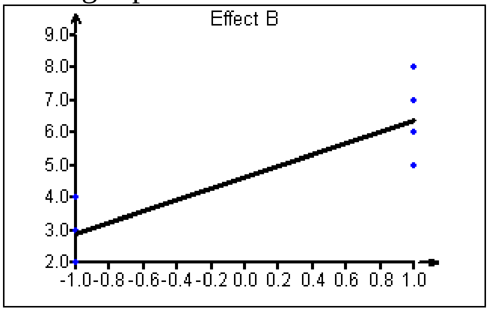
If all effects and interactions are zero, the effects distribution follow the normal distribution. In QQ-plot we can see deviations from normal distribution for individual factors. Such deviations (like factor B on the picture) can be interpreted as significant effect of the factor.

QQ-plot deviations

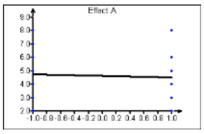


Absolute deviations from the line in the QQ-plot. High values suggest significance of factors.

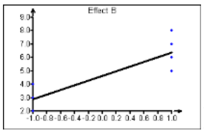
Averages plot



Averages plot gives average response for low and high level of each factor. The scale on all plots are the same so the plots can be compared.

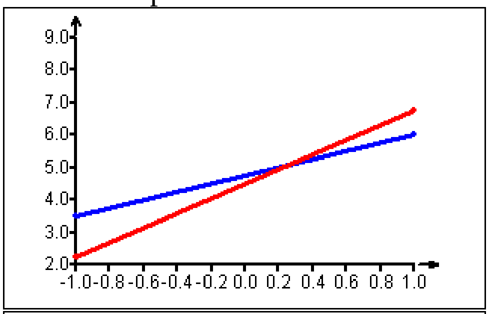


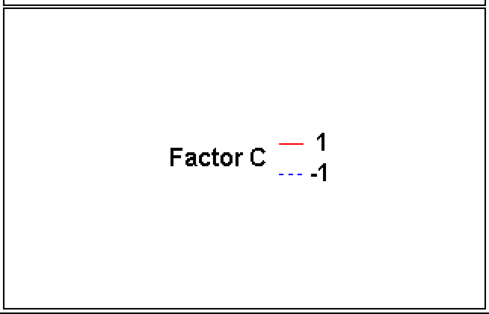
Example of small effect



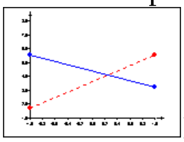
Example of high effect.

Interaction plots

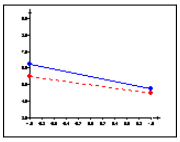




Interactions plot can reveal possible significant interactions of the first order between factors. Interaction will be diagnosed if the slopes of the blue and red line differ significantly. The scale on all plots are the same so the plots can be compared.



Example of a significant interaction



Example of an insignificant or no interaction

Interaction of two factors, say A and B mean that a factor A influences the response differently in dependence on the level of factor B.