Fractional factorial design

- In statistics, **fractional factorial designs** are experimental designs consisting of a carefully chosen subset (fraction) of the experimental runs of a full factorial design . The subset is chosen so as to exploit the sparsity-of-effects principle to expose information about the most important features of the problem studied, while using a fraction of the effort of a full factorial design in terms of experimental runs and resources. In other words, it makes use of the fact that many experiments in full factorial design are often redundant, giving little or no new information about the system.

½ fractional design from FFD 2^4
I = ABC

A	B	C	D	Output function
-1	-1	-1	-1	45
1	-1	-1	1	100
-1	1	-1	1	45
1	1	-1	-1	65
-1	-1	1	1	75
1	-1	1	-1	60
-1	1	1	-1	80
1	1	1	1	96

The simple algorithm

- Choice of specific “contrasts” or aliases indicating the replacement of real factors affecting the experiment with their alternatives (aliases);
- This choice is, usually, based on specific preliminary information about the system studied;
- In fractional design shown before, each main effect is aliased with a 3-factor interaction (e.g., A = BCD), and every 2-factor interaction is aliased with another 2-factor interaction (e.g., AB = CD). The aliasing relationships are shown in the table below. This is a resolution IV design, meaning that main effects are aliased with 3-way interactions, and 2-way interactions are aliased with 2-way interactions.

Aliases

Single effect is aliased with 3-way interactions and 2-way interactions with another 2-way interaction

A = BCD

B = ACD

C = ABD

D = ABC

AB = CD

AC = BD

BC = AD

Differences to the FFD

- Reduction of number of real experiments (e.g. instead of 16 only 8 if 4 factors are involved)

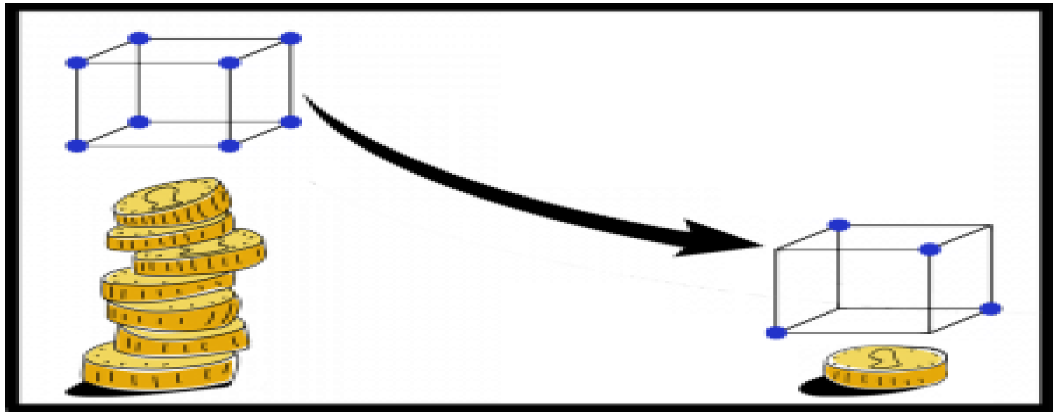
- But reduction also of the number of regression coefficients in the final model (loss of information?)

The new model

- The new model uses for interpretation not 8 but 4 regression coefficients:
- $$Y = a'_0 + a'_1 x_1 + a'_2 x_2 + a'_{12} x_1 x_2$$
- The new coefficients are combinations of the coefficients from the full factorial design
- a'_0 includes a_0 as well as the 3-way interaction impact $a_1 a_2 a_3$

- $a'_1 - a_1$, also the 2-way interaction $a_2 a_3$
- $a'_2 - a_2$, also the 2-way interaction $a_1 a_3$
- a'_{12} - a_{12}, also the single coefficient a_3 in case the aliases is $x_3 = x_1 x_2$

Advantages and disadvantages of fractioning



What is up?

- Less real experiments
- Quick final results
- Options for validation by using different aliases
- Incomplete model
- Satisfactory preliminary information needed
- Options for further reduction of the number of experiments by applying ¼ fractional design

Random balance method for factor ranking

- This is an experimental design method whose goal is to rank (put into a hierarchical order) big number of input factors using a very small number of real experiments, e.g. 10 factors are ranked by 12 experiments;
- Important: the input factors should take two levels of variation as in case of full factorial design;
- The design itself is constructed by a random (lottery) principle and the major point is separation of the input factors into groups of two as for each group a full factorial design 2^2 is made;
- Specific requirement is the introduction of dummy factors being important for assessment of the experimental error.

What is next for the method application?

- The calculation algorithm is very simple – it determines the difference (span) between the output function values for the “high” and the “low” level of variation for each input factor;
- The bigger the difference the higher the rank of the respective factor; the positive difference (higher output function values for the “high” level of factor change) stands for positive factor impact on the output and, on contrary, the negative difference indicates negative impact;
- The method is not suitable for modeling and optimization, just for ranking.

Organization of the random design

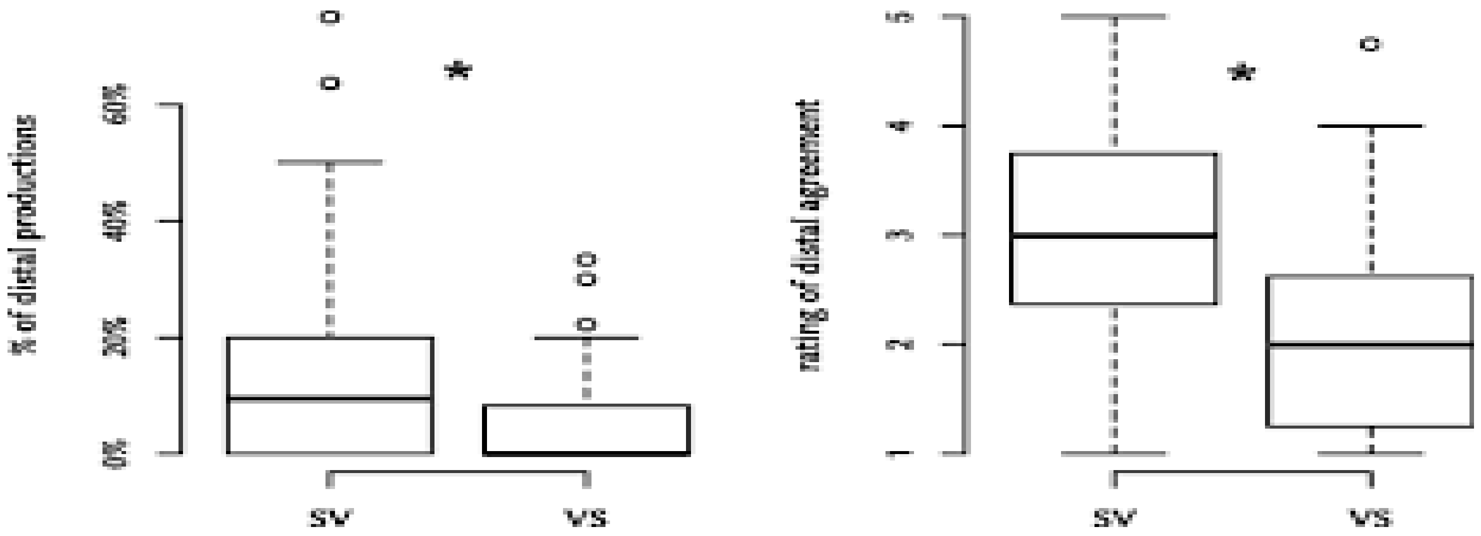
- Separation of the inputs into groups of two;
- Addition of dummies to make the number of inputs an even figure;
- Full factorial design for each couple of factors – table with four rows and (short version only for linear impacts) two columns;
- Organization of the “random” construction of the general design in such a way that for each real experiment (row in the general table of the random design) respective “+” and “-” for each factor are put (according to the outcome of the “lottery”);

The design table (simplified version)

Factor 1 (real)	Factor 2 (real)	Factor 3 (dummy)	Factor 4 (real)	. . . .	Factor N (real)
+	-	+	+	…	+
+					
-					
+					
…					

Data Treatment

- Graphical output



Interpretation

- As already mentioned – the dimension of the box (area) formed for each factor (and the sign with respect to the median value) are measure for the impact of the factor;
- Dummies should be theoretically not boxes but lines – they are imaginary factors (like the distance between the feet of the scientist and the experimental plot as "–" level and between the head and the plot – "+" level); so their area is a measure for the noise or the experimental error;
- It is also possible to calculate the rank as difference between the output function values for both levels of variation of the factor value.

Disadvantages of the strategy for ranking

- Some scientists believe that the random balance method is “too random” and not enough “balanced”. The choice of experiments is arbitrary and this is a reason for different arguments;
- There are even jokes with the methods: “It is the same if you take one million monkeys, make them hit the keyboard, and expect a play by Shakespeare to appear on the paper” says one of the fathers of the experimental design George Box. He continues “ but it look nice and keep the soul fresh”.
- Nevertheless, the method is widely applied in practice.