DESIGN OF EXPERIMENTS (DOE)

wikipedia.com -

> design of experiments, or experimental design, is the design of all information-gathering exercises where variation is present, whether under the full control of the experimenter or not

this theory was the first serious attempt to systematise the application of statistics for industrial purposes [maximising information, that is, proposing reliable enough forecasts on the basis of historical data understanding]

two main parts

a. design creation – minimum number of measurements containing the maximum amount of information possible
b. resulting information analysis – 2 values [measurements] per variable + linear models

main features deriving from the aforementioned bipartition

a. rigid input system - specific values [experimental measurements conditioned by the method]
b. indirect assessments [ANOVA] and, in general, not too reliable predictions

a.- design

there are multiple choices...

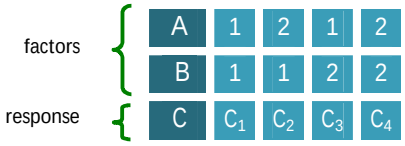
- randomised block design
- full factorial designs
- fractional factorial designs
- plackett-burman designs
- response surface designs

NOTE – full factorial is the most detailed design - the rest of them are just simplifications over it

...although all of them have in common that the measurements-collection process is conditioned by the chosen method

- example 1 -

2 factors / 1 response [2 independent vs. 1 dependent] → A-B / C
2 levels [number of values for each factor] → 1, 2
number of combinations [full factorial design] → $levels^{factors} = 2^2 = 4$



unlikely the other theories analysed so far, DOE can not deal with any random dataset, a specific (& determined-by-the-method) set of measurements has to be carried out in order to fulfil the corresponding requirements

b.- information analysis

two main types of methods can be distinguished here

1. regression methods – all the different methodologies [depending upon the chosen design] are based on the same idea: simple linear regressions [1 independent vs. 1 dependent]

- example 1 -

A1=Ac2-Ac1 / B2=Bc2-Bc1 / C1=C2-C1 ➡ regression1

A2=Ac3-Ac1 / B2=Bc3-Bc1 / C2=C3-C1 ➡ regression2

A3=Ac4-Ac1 ; B3=Bc4-Bc1 / C3=C4-C1 ➡ regression3

A4=Ac3-Ac2 ; B4=Bc3-Bc2 / C4=C3-C2 ➡ regression4

A5=Ac4-Ac2 ; B5=Bc4-Bc2 / C5=C4-C2 ➡ regression5

A6=Ac4-Ac3 ; B6=Bc4-Bc3 / C6=C4-C3 ➡ regression6

2. analysis of variance (ANOVA) – this theory was created by one of the precursors of modern DOE [Sir Ronald Fisher] - indirect analysis [see 1. statistics / b-3. PLS]

trendingBot point of view

one of the main statistical drawbacks [simplifications => user-defined parameters] has been removed from the regression part [not completely from ANOVA analysis] by applying detailed combinatorics

although the previous statement might indicate close similarities between this theory and trendingBot's algorithm, a more in-depth analysis will remove any possible doubt

a. combinatorics at different stages –
   DOE combines the measurements [impose how they have to be taken]
   trendingBot combines everything given input [freedom to measure]
b. multi-variable regressions –
   DOE simple linear regressions over just 1 independent variable at each step
   trendingBot polynomial fits over ALL the given independent variables at each step
c. rigid vs. flexible –
   DOE relies on linear calculations and renounce to take any "exponential help"
   trendingBot accounts for as many exponential variations as possible
d. overview –
   DOE aims to maximise the data collection process [low experiments & high information]
   trendingBot "squeezes" the given data to get as much information as possible