

SIMULATE YOUR DATA.

The structure of the generated dataset must follow the structure presented on the next table.

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | .. | Factor 10 | Answer |
|-----------------|----------|----------|----------|----------|----------|----|-----------|--------|
| Individual 1 | | | | | | | | |
| Individual 2 | | | | | | | | |
| | | | | | | | | |
| Individual 2000 | | | | | | | | |

1. Define, for each factor (from 1 to 5) a distribution (the RVGs that you prefer, uniform, normal, exponential, etc.). For the factors 6 to 10 define a function that uses the previous variables, as an example $F_6 = F_1 + 2F_3$.
2. Define an answer variable that will be composed by a function that combines a subset of the previous factors plus a normal distribution you know (to add some random noise).

Note: to do some semantics to this exercise, the Answer will be the time needed to do an specific task that is composed by a set of sub-tasks (represented by the factors columns).

OBTAIN AN EXPRESSION TO GENERATE NEW DATA.

Imagine that you don't know nothing regarding this dataset. You need to explore it because you want to define a model to obtain new data for your DOE (you want to detect the possible relations and the interactions between the factors, or maybe you want to test alternatives or predict future scenarios).

1. Explore the possible relations of all the factors and the answer variable, you can use any technique developed during the course.
2. Describe what you find on this analysis and, explain if it is coherent with the knowledge you have from the data.
3. Propose an expression (as an example using a LRM) to understand the relations between the data. **What are the factors that affects the answer?**

SIMULATE NEW DATA

The simulation model will be a quite simple model composed by one server by each one of the factors you use on the answer. If the answer is

$$answer = F_3 + 5 F_6$$

Use any of the simulation algorithms presented on theory to do the codification of the model.

Perform some basic validations on your model.

DOE

Now you have a model (LRM or simulation) to generate new data. This model can be used to generate data for the different scenarios that must be considered.

1. Define a DOE to explore with what parametrization of the 10 factors the answer obtains the best value (define what means best, i.e. maximize or minimize the value).
2. Detect and analyze the interactions.

Remember

- *Set the objectives.*
- *Select the process variables. Hypotheses to be tested, etc.*
- *Define an experimental design.*
- *Execute the design.*
- *Check that the data are consistent with the experimental assumptions.*
- *Analyze and interpret the results, detect effects of main factors and interactions.*