

## 1 DEFINE YOUR OWN RNG

Propose a RNG for use in your analysis.

1. Implement this RNG.
2. Test the correctness of your RNG using at least **two** tests. You can implement your tests or can use the RNG tests implemented on R. Explain the main idea (is not needed to enter on the details) of the test selected.

*You can implement this RNG on R on in the language you prefer. Remember that you can define your own functions in R like this:*

```
My_RNG <- function (x, seed=7) {
  #Do something
  return(x)
}
```

Some useful code:

```
library(randtests)

example = My_RNG(1000)

bartels.rank.test(example)
cox.stuart.test(example)
difference.sign.test(example)
rank.test(example)
turning.point.test(example)
```

## 2 SIMULATE YOUR DATA.

The structure of the generated dataset must follow the structure presented on the next table. Here is not needed to use your RNG.

	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  $F6 = F1 + 2F3$ .
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).

### 3 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 (using a LRM) to generate new data. This is the method that you are going to use to generate new values using a subset of the factors, for a more complex dataset one can use other approaches like Simulation (see next).
  - a. The simulation model will be a very simple model composed by one server by each one of the factors you use on the answer. If the answer is

$$answer = F3 + 5F6 + Random\ noise$$

- b. Then the model will be like the one represented below, where the arrivals will be represented by a constant distribution with 1 of value. The answer will be the overall service time.



### 4 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.

### 5 VALIDATION

Describe the validation process and methods you will use to verify the model.

Perform at least one validation method by each group member, on the model i.e if your group is composed by three members apply three validation methods.