SMDE Second assignment

# Executive summary

We are going to simulate the behavior of different groups of runners in Boston Marathon, based on real data taken from 2017 episode.

In particular I am going to simulate a subgroup of all the runners, men from 30 to 40 years old.

This simulation is going to be useful for planning all different resources for next version of the same event, such as Water Spots, WC, Solid meals and so on.

# System description, introduction

A marathon in a competition that consist in running 42.195 Km distance with a maximum time of 6 hours.

At the beginning line runners are divided in groups depending on their performance category, for example those runners who can run in 2:30 minutes/kilometer, 3, 3:30 and so on. The ones who perform the better start at the beginning.

On different spots alongside the marathon the organization service is going to provide different resources to runners in order to support the competition. This resources are multiple and can vary according to the organization budget and preferences. Such resources can be: Water, Solid meals like fruits, Glucose gels, WC, Medical spots, etc.

In my case I am going to do certain assumptions and reductions that can be found in the different hypotheses on the following sections.

On the other hand to have a successful simulation for a better analysis it is important to be based on previous statistical data from previous version of the same competition.

# Problem description

In our model, the main goal is to find the right amount of resources in order to avoid collisions and waiting queues in the running line.

At the beginning I have found that having a few amount of resources in Water spots, less that 50 per sport, led to issues and long queues for runners.

I have set this number to 100 resources of water per spot.

Since WC and Solid is less frequently use for runners, and our distribution function was set on that sense, I haven’t found an issue with the resources in that sense. Having for example 8 WC on the WC spots is enough. Also having 100 resources in the Solid spot and with 1 spot only is enough for the whole run.

## Systemic Structural, Systemic Data and Simplifying Hypotheses

SH\_01 Since we cannot predict the weather with exact precision we are not taking into account the weather as a factor

SH\_02 We are not going to measure or simulate injured people. We assume that anyone is injured or abandon the marathon

SH\_03 We are going to reduce the amount of water, wc and solid spots in order to simplify DOE after and not to have so many factors.

SH\_04 Where there is a 2 colliding points, for example water and solid or water and bathroom, we are going to distribute the runners between the 2 points. There is not going to be runners that go to both resources.

SS\_01 The marathon cannot last more than 6 hours. The system will be consider finished after 6 hours.

SS\_02 The total distance of the marathon is 42.195 Km

SS\_03 Time performance for each runner is going to be taken every 5 km.

SS\_04 We are going to analyze Men 30-40 years old.

SD\_01 We are going to use a linear regression model for predict the time that takes to each runner to cross each segment

SD\_02 Elite runners are going to be calculated as a percentage based on Boston Marathon 2017 on each group

SD\_03 Elite runners are going to take less to cross each performance measurement based on the same percentage defined in SD\_02

SD\_04 There is going to be 3 water points at Km 10, 21 and 30, and it can deliver 100 unit of water at time

SD\_05 There is going to be a fruit and glucose point at Half marathon and 35km and on each point there is going to be capacity of 500 units at time.

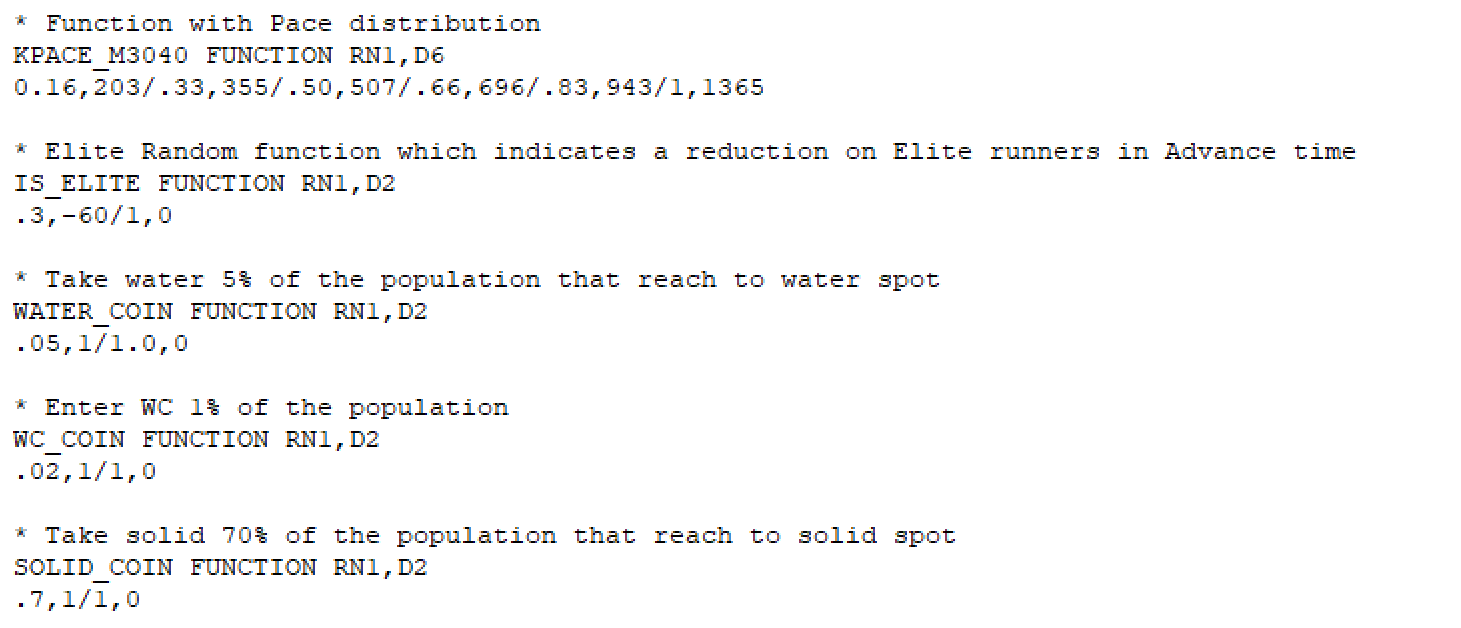
SD\_06 There is going to be 2 bathroom points at Km 25 and 30.

# Model specification

We are going to specify the model use for the simulation.

## Functions

We have defined several functions to be use both in the linear model, as well as in some random behavior such as going or not to WC or taking or not Water.

As we can see *KPACE\_M3040* is the distribution for the Pace in the km 30 to 40 which is the factor that determine in this case on the linear model.

The rest of the function are just flipping coin function with some probability distribution to decide for example if the runner is going to the WC or is going to take Water.

## Storages

Storages are defined in Hypotheses section and basically is the amount of resources for each type in each spot of the competition.

## 

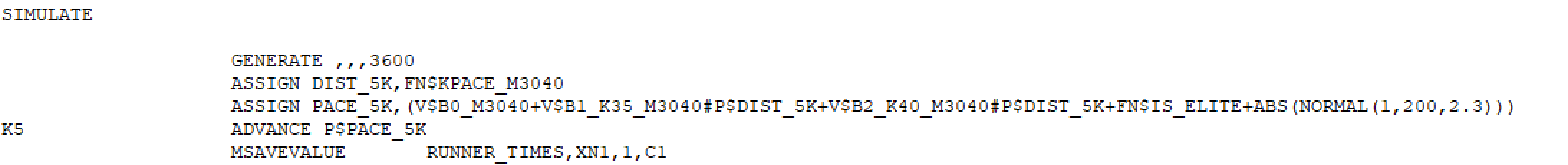
## Variables

Variables are to set some constant alongside the execution as well as a matrix to store some intermediate values for each runner.

## Simple Advance Block

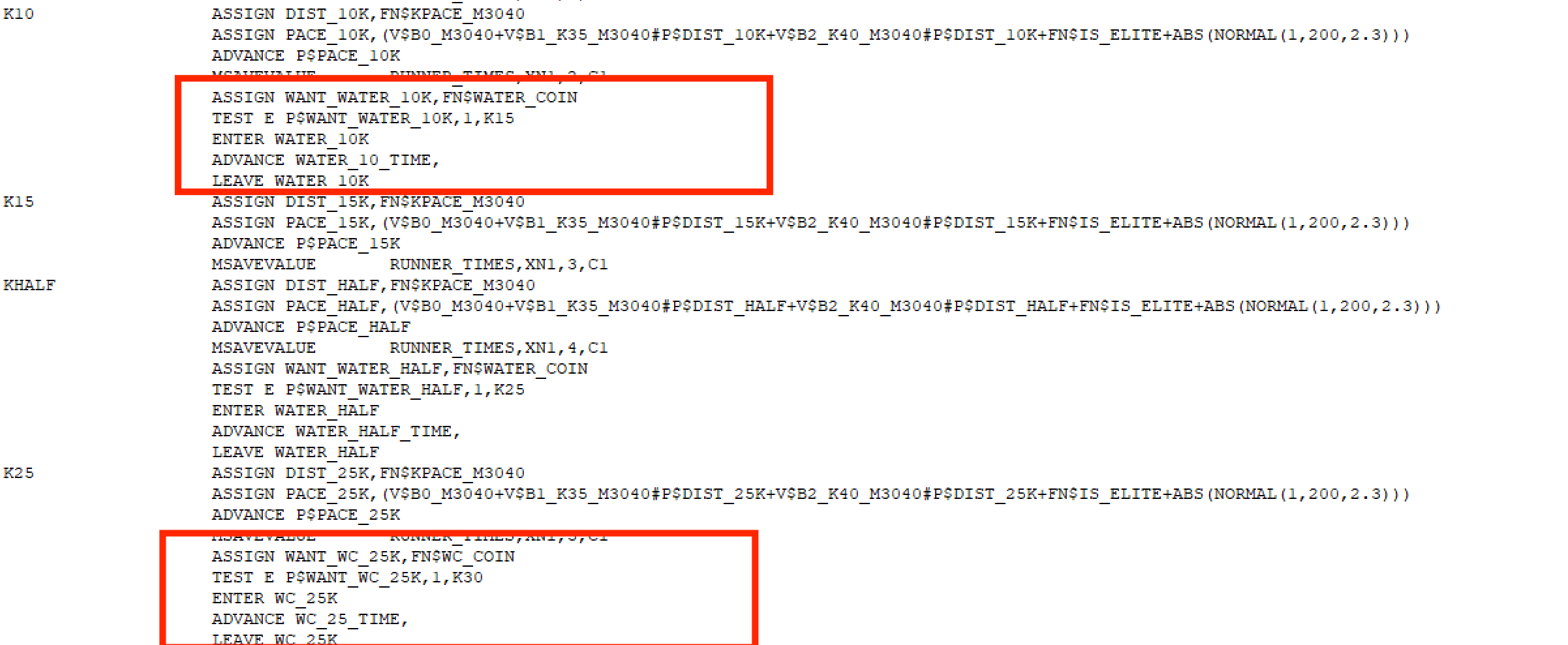
We can start explaining a simple advance block which take place at the beginning of the competition.

Basically this block is doing is to calculate based on the Linear Regression model, the amount of time the runner is going to take to cross that mark. There is a slightly additional amount to simulate elite runners and a normal distribution with some noise in case of error.



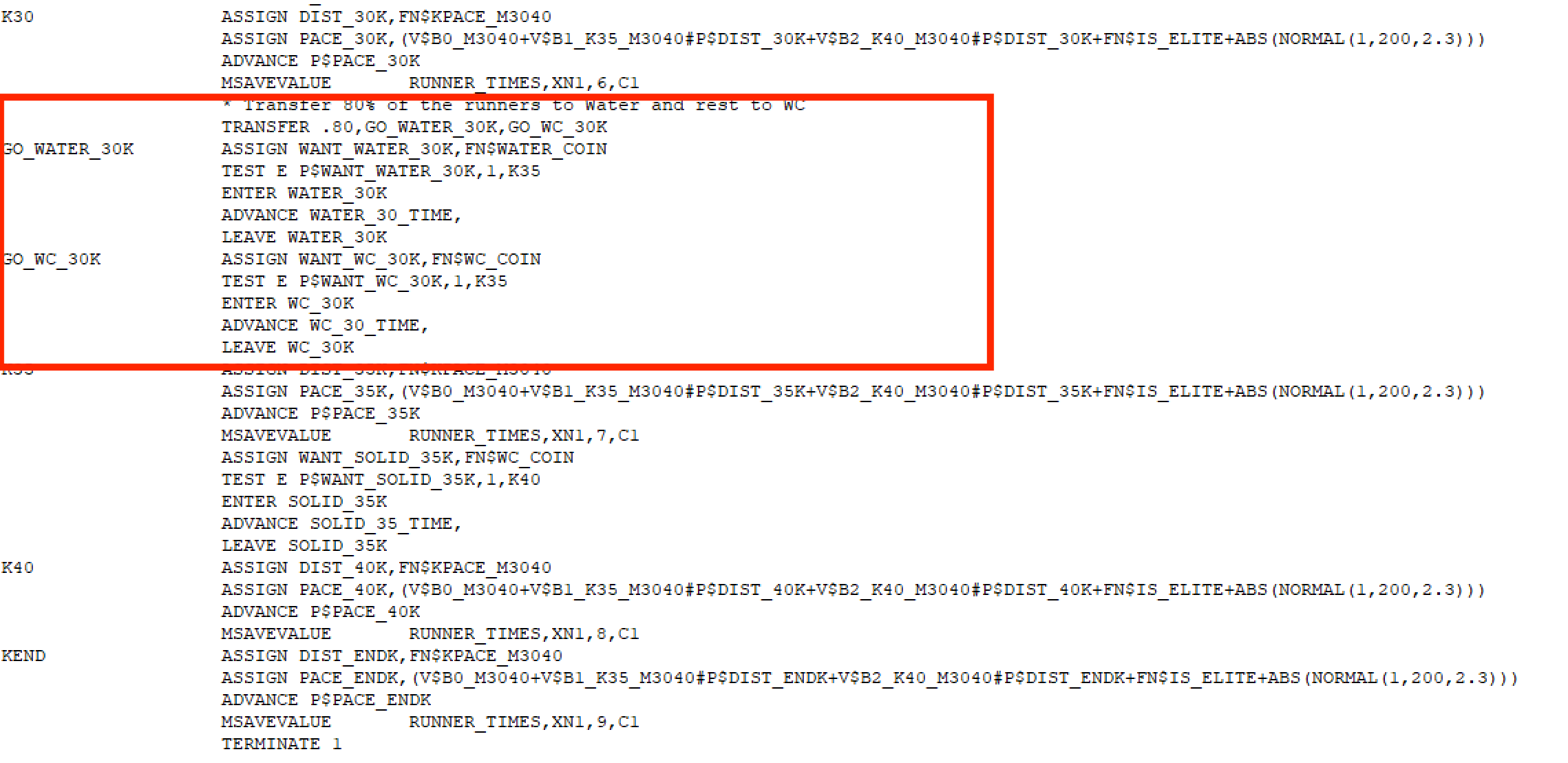
## Water and WC Resources

We can appreciate in the screenshot that we are flipping a coin to see if the runner is going to take water or not. The same behavior we are doing for WC.



## More than one resource

In the case of K30 we have 3 resources at the same spot, Water and WC. In order to represent that some runners are going to take water and some others go to the WC, we are using a TRANSFER with a probability distribution to split the transactions in both resources.



# Codification

Check file ***marathon\_boston.gps***

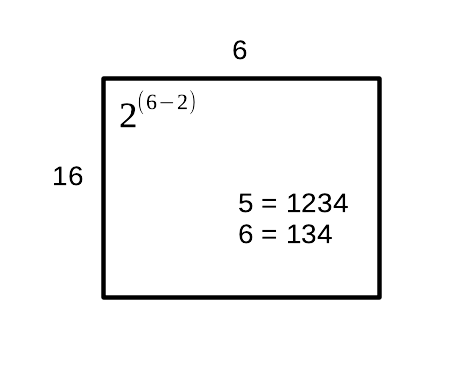
## Data.

The data used along the model was:

* ***boston\_marathon\_men\_3040.csv*** and ***boston\_marathon\_men\_3040.xlsx*** which is the dataset with the marathon of 2017 for that specific group
* ***InputMenExperiment.txt*** which is the input file with the different experiments for the Factorial analysis

# Definition of the experimental framework

Check ***doe\_men\_30\_40.xlsx***



In the simulation we have defined 6 factors, between Water, WC and Solid distributed along the race.

From those 6 factors as we can see we are reducing to 4, having on the main factors all the different types and leaving outside the factors that are repeated like Water or WC.

# Model validation

I am going to do the following validations:

* Black Box validation: Comparing the data gather in the simulation against the historical data
* Fixed Values with Traces.

# Results /Conclusions

According to the simulation run and based on an overall validation with the real data, it seems that the simulation is behaving very close to the real marathon. We are obtaining in average a running time for each runner of 12800 seconds which is 3 and half hours for a marathon of men from 30 to 40 years old.

On the other hand, I have simplified several hypotheses which in fact in a real marathon affects and I am not counting into account, like weather conditions for example.