We created a data set to simulate reality due to the lack of real data sets and the need to collect readings for a very long time.

In the beginning, we simulated the personal information of 10,000 people from national numbers and quadruple names in addition to the device ID, as we will detail below.

Other values are individuals' consumption rates and we have relied on several factors affecting the amount of consumption by estimating the variance of the effect on one individual and the individual effect depending on the number of individuals (unconscious numerical bias)

PERSONAL INFORMATION:

1 - THE DEVICE IDENTIFIER.

Consists of 5 numbers a random function was used and called recursively. If the number was previously created, else it does not exist, it is added to a list and written in the excel file.

2- THE NATIONAL NUMBER OF THE BREADWINNER.

Numbers were formed starting from 9411 to 9972, which means who was born from 1941 to 1997, with ages ranging from 24 to 80 years old.

3- FULL NAME.

A file containing approximately 1000 quadruple names was used, where the names were separated and re-grouped randomly, and with this, we obtained 10,000 different quadrant names.

Now let's understand what the factors are affecting water consumption and how they are simulated.

THE FACTORS ARE AFFECTING.

1. RATE OF INDIVIDUAL CONSUMPTION

The main random factor is the rate of individual consumption, according to the consumption rates of individuals in Jordan, which ranges from 90 to 120 litters / day per person.

2. NUMBER OF FAMILY MEMBERS

A factor with them and influencing the total consumption is the number of individuals in a single family: we created a variable with a random number of members per family, where the range was from 2 to 9 individuals per family, to a maximum.

3. SEASONS CHANGE

The third factor affecting is the consumption season. Of course the individual consumes more water in the summer than in the winter, and with this we have created a random value that bears one of the four seasons for each family.

Where the average impact of the seasons was multiplying the consumption value by the value of the season and the range of the impact was from 0.6 to 1.4 per capita consumption according to the season, so that the distribution of the impact of the seasons was as follows.

Summer = consumption * [1.2 - 1.4)

Spring = consumption * [1.0 - 1.2)

Autumn = consumption * [0.8 - 1.0)

Winter = consumption * [0.6 - 0.8)

4. UNCONSCIOUS NUMERICAL BIAS

The last parameter, which is the effort to try to simulate consumption more accurately, as in most cases the consumption of a person in a family of 3 members is relatively less if the same individual is in a family of 9 members. We called this effect (unconscious numerical bias), because by increasing the number of individuals they go without realizing to the increase in water consumption and the equation for this bias is the number of individuals plus 10 divided by 100, that is, the parameter according to the number of individuals is from 0.12 to 0.19 of the total value, with the rate of bias added as in the following equation.

- **Consumption value + (consumption value x (number of people + 10/200))**
- **Consumption value + (consumption value * (0.060 - 0.095))**

Thus, it affects a small percentage of the consumption of the small family, while its influence is relatively large in the large family.

Using the previous factors, we constructed random consumption values (weekly - monthly - periodic "three months") for each family.

✓ As a result, the equation will be as follows:

```
family_member   = random(2,10)
consumption     = random (90,125) Litter #diffrance random for each equation
Seasons_factor = random (0.6,1.5)
numerical_bias =(family_member+10)/200)*family_member -
                                        - * consumption*Seasons_factor*90/1000)
```

Equation :

weekly = family_member*consumption*Seasons_factor*7/1000)+numerical_bias

monthly = family_member*consumption*Seasons_factor*30/1000)+numerical_bias

periodic = family_member*consumption*Seasons_factor*90/1000)+numerical_bias

❖ We repeated the process of fixing the personal information to obtain 10 records for each family, with a total of 100,000 records, in preparation for data analysis.