**b. Medium School (MDS)**

Half (52%) of houses have one student in medium level, and third of them with none and the rest has two.

Pie chart below demonstrates the repartition of number of habitations according to secondary level (Figure 3.16).

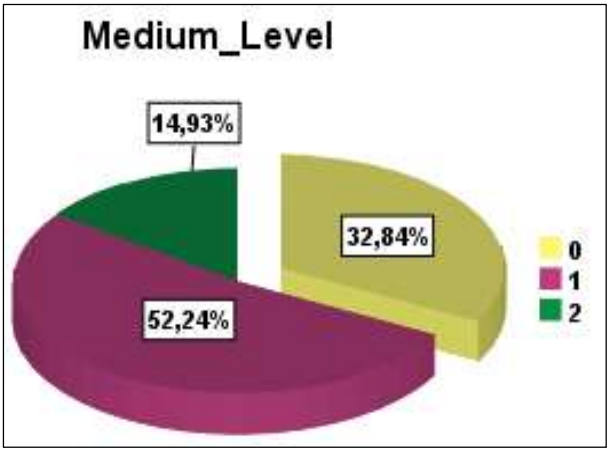


**Fig 3.16** Repartition of number of habitations according to secondary level

**c. High School (HGS)**

For high school level, the population has at least one student in this level (57,21% with one student, 24,88 % with two students and 17,91% with three in high school), the main reason behind this result is that most parents have a high school level.

Pie chart below demonstrates the repartition of number of habitations according to high school level (Figure 3.17).

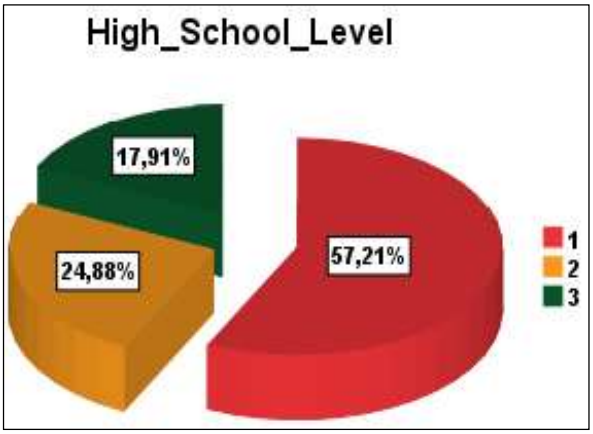


**Fig 3.17** Repartition of number of habitations according to high school level

**d. University Level (UNIV)**

Even though 24% of population are in the university but 29% of houses does not have anyone in university level. This means that there is a tendency only in someone houses to go to university.

Pie chart below demonstrates the repartition of number of habitations according to university level (Figure 3.18).

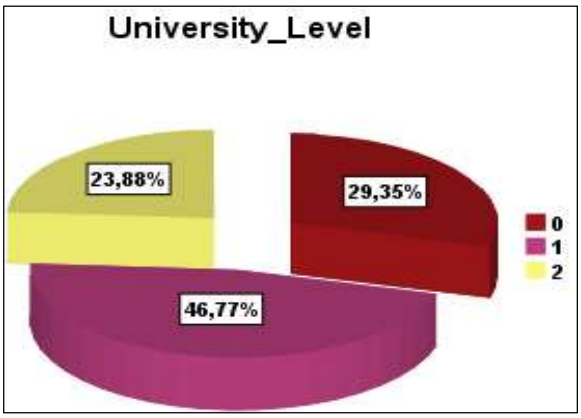


**Fig 3.18** Repartition of number of habitations according to university level

### 3.4.2.4.Household Income (INC)

Household income is a strong background variable determining water use and demand. Average monthly income in Algeria is around 39.900 DA in 2016 (ONS: National Statistics Office; http://www.ons.dz/). Higher income class constitutes people having average monthly income above 100.000 DA. Households having average monthly income less than 39.900 DA are categorized as lower income class. Generally, people with higher income are more concern about water but their lifestyle is differing from lower income group people.

Table 3.4 shows that the mean of monthly income is about 53 905 DA with minimum equal to 35 000 DA and maximum equal to 110 000 DA. Table 3.7 and figures 3.19 demonstrated that **6 %** of the respondents fall in the high-income class, while **7%** of respondents belong to class of lower income and **87%** falls in middle-income category.

**Table 3.7** Distribution of sample based on monthly income

| Income Classes | Interval | Frequency | Percentage |
|---|---|---|---|
| **Lower Income** | < 39 900 DA | 15 | 7 % |
| **Middle Income** | 40 000_990 000 DA | 174 | 87 % |
| **Higher Income** | > 100 000 DA | 12 | 6 % |
| **Total** | | **201** | **100 %** |

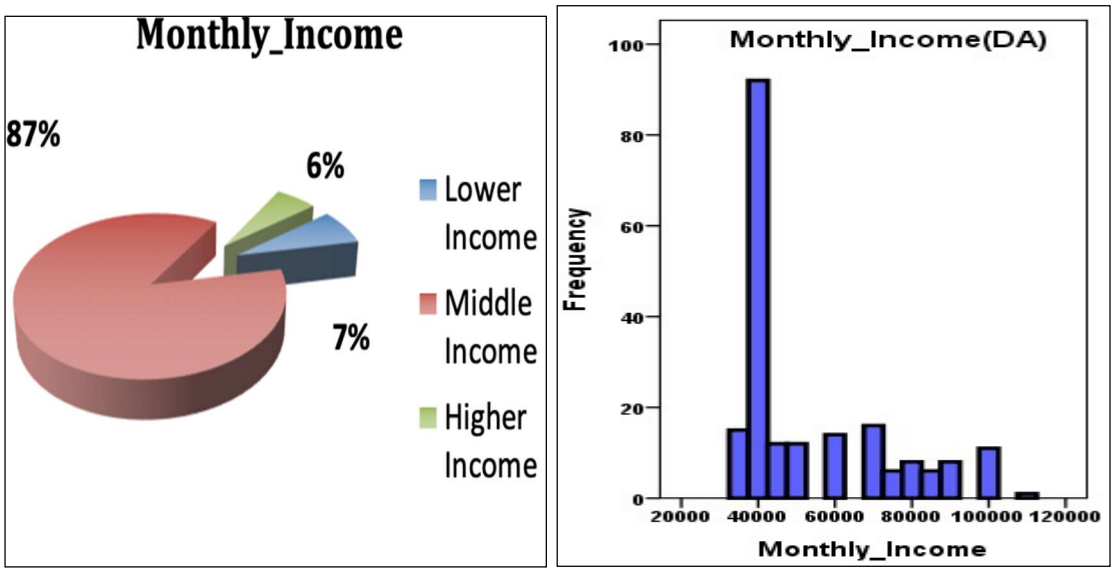


**Fig 3.19** Distribution of sample based on income categorization

### 3.4.2.5.Cars related parameters

Three parameters of cars related are: existence, number (CARN) and washing frequency. For car possession (EXC) **87%** of households have at least one car (figure 3.20).

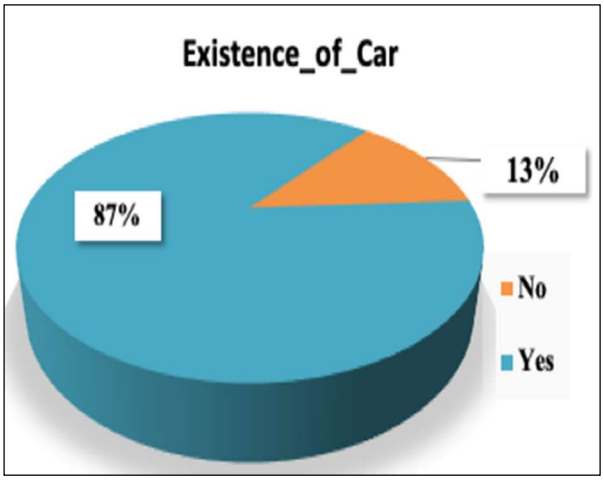


**Fig 3.20** Cars possession

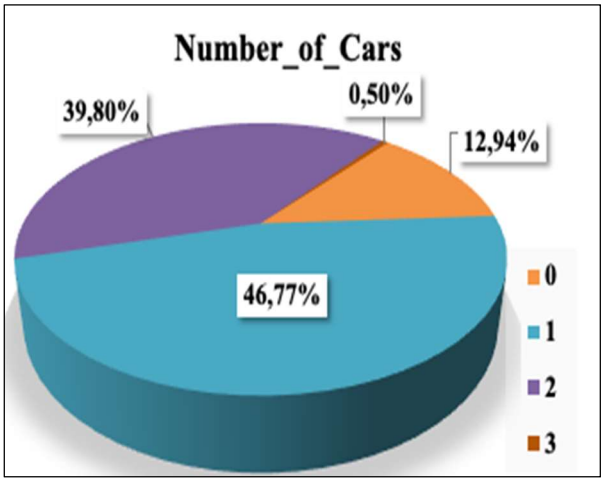


**Fig 3.21** Repartition of buildings according to cars number

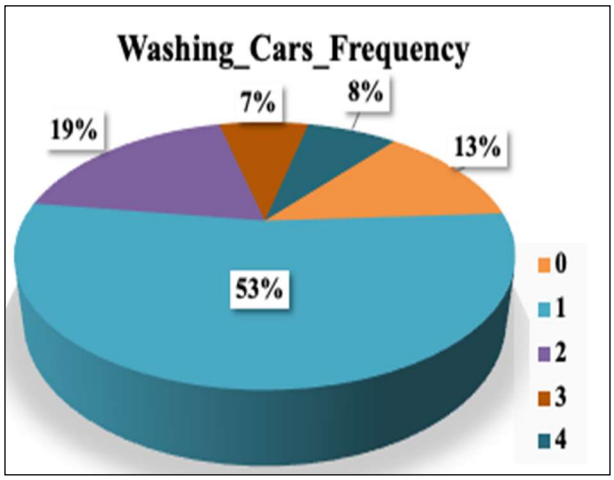


**Fig 3.22** Frequency of washing cars per month

From table 3.4 above, mean value of number of cars is equal to **two** cars, minimum and maximum number of cars are equal to zero and three cars per house, respectively. Figure 3.21 demonstrates the repartition of cars in the study area.

For cars washing behaviors (WCAR), figure 3.22 demonstrates the frequency of washing cars in houses. **53%** washing their cars one time per month, **19%** washing their cars **2** times per month, **7%** washing their cars **3** times per month and **8%** of residents washing their cars **4** times per month. Table 3.4 shows that the mean of WCAR is equal to **two**.

### 3.4.3. Indoor Habits of Residents (INH)

Indoor Habits of Residents (INH) summarize all water related practices frequencies: Washing Dishes (WDISH), Washing Clothes (WCL), Using Toilets (UTLT), Shower for Female (FSHW) and Shower for Male (MSHW). The statistical characteristics of domestic water component is presented in table 3.8. The daily and weekly water usage was determined based on residents answers, which it collected according to the questionnaire paper.

**Table 3.8** Household water usage

| Statistical Parameters | WDISH (Day) | WCL (Week) | UTLT (Day) | FSHW (Week) | MSHW (Week) |
|---|---|---|---|---|---|
| **Mean** | 3,00 | 2,00 | 4,00 | 2,00 | 2,00 |
| **Median** | 3,00 | 2,00 | 4,00 | 2,00 | 2,00 |
| **Mode** | 3 | 2 | 4 | 2 | 2 |
| **Std. Deviation** | 0,61 | 0,69 | 0,95 | 0,99 | 0,95 |
| **Variance** | 0,38 | 0,48 | 0,89 | 0,99 | 0,90 |
| **Skewness** | -1,13 | 0,84 | 0,62 | 0,59 | 0,53 |
| **Std. Error of Skewness** | 0,17 | 0,17 | 0,17 | 0,17 | 0,17 |
| **Kurtosis** | 0,24 | 0,87 | 0,38 | 1,16 | -0,52 |
| **Std. Error of Kurtosis** | 0,34 | 0,34 | 0,34 | 0,34 | 0,34 |
| **Minimum** | 1 | 1 | 3 | 1 | 1 |
| **Maximum** | 3 | 4 | 7 | 7 | 5 |
| **Sum** | 517 | 342 | 871 | 480 | 413 |

#### 3.4.3.1.Washing Clothes (WCL)

Table 3.8 above shows that the mean of WCL per week is equal to **two**, the minimum and maximum frequency of washing clothes are **one** and **four** times per week. Bar chart in figure 3.23 depicted the frequency (number of time) of washing clothes per households as following:

- **41.29%** (83) wash their clothes once a week
- **49.26%** (99) wash their clothes twice a week
- **7.46%** (15) wash their clothes 3 times per week
- **1.99%** (4) wash their clothes 4 times per week.

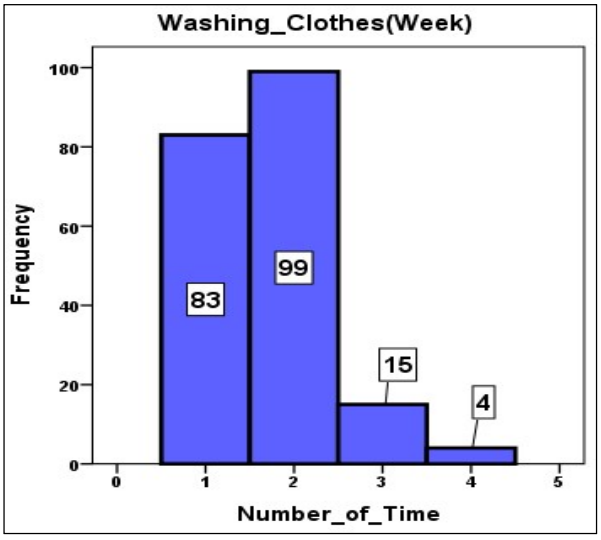


**Fig 3.23** Distribution of households according to the frequency washing clothes

The following distribution explained by Algerian behaviors in washing clothes that generally correspond to weekend only and due also to water distribution system (frequency of water supplying).

### 3.4.3.2.Washing Dishes (WDISH)

The frequency of dishwashing ranges between once and three times a day and shows that the houses could be categorized into three main groups (figure 3.24). The reason why majority of houses have a frequency of three times a day correspond to three main dishes a day (breakfast, lunch and dinner).

- **6.47% (13)** wash dishes one time per day.
- **29.85% (60)** wash dishes **2** times per day.
- **63.68% (128)** wash dishes **3** times per day.

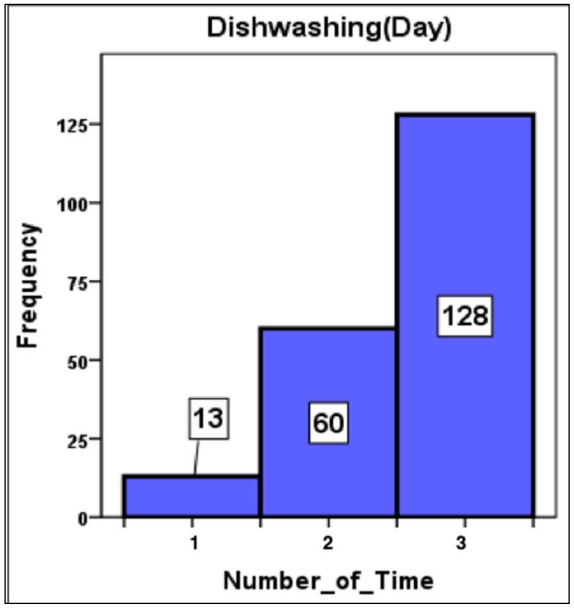


**Fig 3.24** Distribution of households according to the frequency of washing dishes

### 3.4.3.3.Using Toilets (UTLT)

Data in the table 3.8 above shows that, the mean usage of toilet is **four**. The minimum and maximum frequency of using toilets are **three** and **seven** times per day. Bar chart in Figure 3.25 depicted the frequency of using toilets per day per person as following:

- **17.91% (36)** use toilets **3** times per day.
- **43.28% (87)** use toilets **4** times per day.
- **29.35% (59)** use toilets **5** times per day.
- **6.47% (13)** use toilets **6** times per day.
- **2.99% (6)** use toilets **7** times per day.

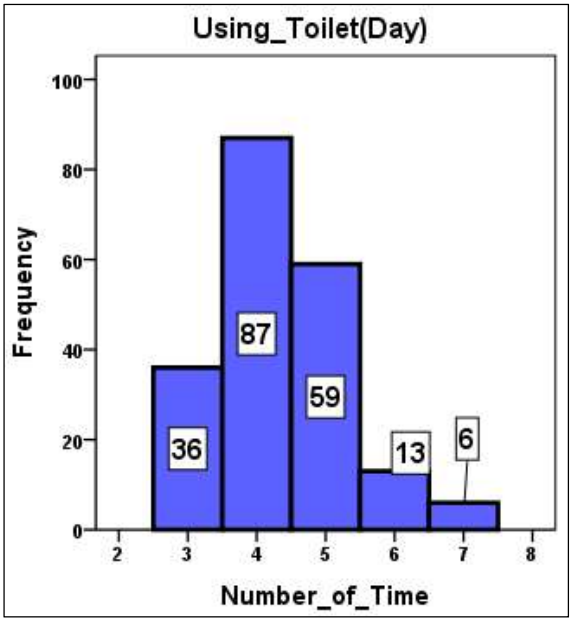


**Fig 3.25** Distribution of households according to the frequency of using toilets

### 3.4.3.4.Showering

Showering frequencies vary considerably from house to house and person to person. Showering is also related to gender. For Females (FSHW) the bath use mean is 2 (table 3.8) above).

The minimum and maximum frequency of taking shower are **1** and **7** times per week. Bar chart in Figure 3.26 showed the frequency of taking shower per week per households as following:

- **19. 90% (40)** of females take shower for once weekly.
- **35. 82% (72)** of females take shower for 2 times per week.
- **31.84% (64)** of females take shower for 3 times per week.
- **11.44% (23)** of females take shower for 4 times per week.
- **0.50% (1)** of females take shower for 5 times per week.
- **0.50% (1)** of females take shower daily.

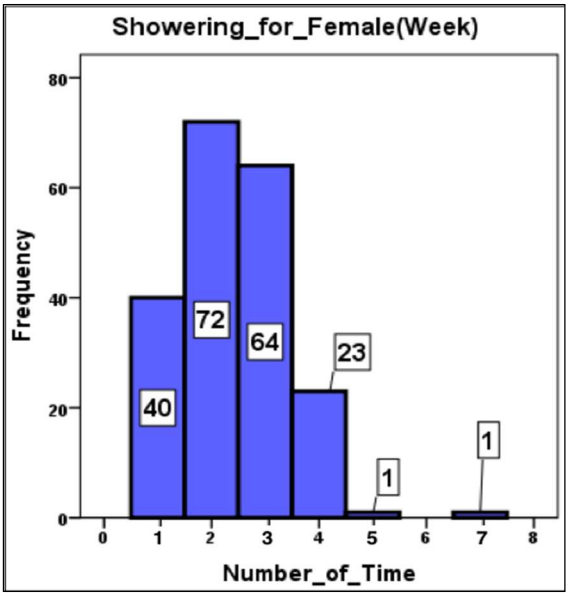


**Fig 3.26** Distribution of households according to the frequency of showering for female

Data in the table 3.8 above shows that, the mean of MSHW is equal to **2**. The minimum and maximum frequency of taking shower for males are **1** and **5** times per week. Bar chart in Figure 3.27 showed the frequency (number of time) of showering for males per week per households as following:

- **33.83% (68)** of males take shower for one time per week.
- **34.82% (70)** of males take shower for 2 times per week.
- **23.88% (48)** of males take shower for 3 times per week.
- **6.97% (14)** of males take shower for 4 times per week.
- **0.50% (1)** of males take shower for 5 times per week.

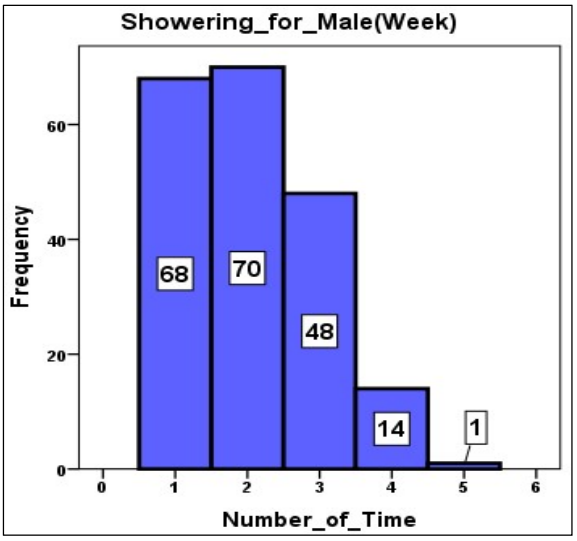


**Fig 3.27** Distribution of households according to the frequency of showering for males

### 3.4.3.5.Who use More Water Inside Every House, Male or Female?

Generally, women use more water inside every house. Highest number of liters used by women is for the household activities. Data reflect that around **47.26% (95)** of people answered that women use more water than men for domestic purpose, whereas around **44.78% (90)** of answers goes to male and the rest of percentage **7.96% (16)** is for those answering that both male and female use the same quantity of water (Figure 3.28). Women are requiring more water because they are involved in performing maximum number of domestic works.

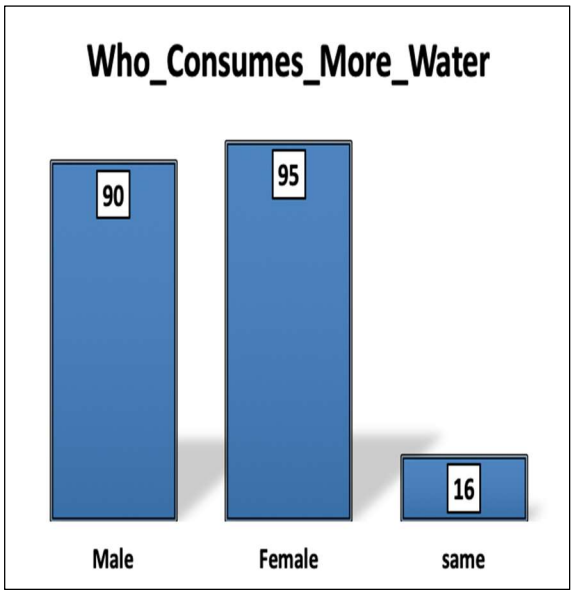


**Fig 3.28** Opinion poll about the use of water between male and female

These values of domestic water use were expected considering the different factors influencing consumption, besides the period of records was associated to nearly summer conditions, in which high temperatures occurred.

### 3.4.4. Physical Characteristics of Housing Units (PHC)

Beside of human activities, the physical characteristics of building units could affect WCP. Five main characteristics are covered: Total Area (TAR), Building Area (BAR), Number of Rooms (ROMN), Garden Area (GAR) and Garden Watering program (GWAT).

**Table 3.9** Summary of physical characteristics of houses

| Statistical Parameters | TAR (m²) | BAR (m²) | ROMN | GAR (m²) | GWAT |
|---|---|---|---|---|---|
| **Mean** | 186,67 | 164,97 | 6,00 | 21,70 | 2,00 |
| **Median** | 200,00 | 164,00 | 5,00 | 16,00 | 2,00 |
| **Mode** | 80 | 90 | 4 | 20 | 2 |
| **Std. Deviation** | 77,29 | 75,81 | 3,09 | 18,14 | 0,79 |
| **Variance** | 5973,33 | 5746,91 | 9,54 | 328,98 | 0,62 |
| **Skewness** | 0,15 | 0,19 | 0,92 | 2,16 | 0,56 |
| **Std. Error of Skewness** | 0,17 | 0,17 | 0,17 | 0,17 | 0,17 |
| **Kurtosis** | -1,27 | -1,19 | -0,20 | 4,12 | -0,22 |
| **Std. Error of Kurtosis** | 0,34 | 0,34 | 0,34 | 0,34 | 0,34 |
| **Minimum** | 80 | 40 | 2 | 2 | 1 |
| **Maximum** | 320 | 302 | 13 | 80 | 4 |
| **Sum** | 37520 | 33158 | 1193 | 4362 | 381 |

#### 3.4.4.1. Buildings

Building unit is considering as significant indicator of water use. The type of houses influences the water use. Moreover, to consider all variables such as garden parameters, “single houses” are selected for the study. Figure 3.29 shows the repartition of houses according to their total area (TAR). Table 3.9 demonstrates that the minimum size of houses is **80 m²** and the maximum size is **320 m²**. While, the mean is equal to **186.67 m²**. The samples show big variety with sixteen categories.

Table 3.9 demonstrates also the minimum size of building is **40 m²** and the maximum size is **302 m²**. Figure 3.30 shows the repartition of houses according to their building area (BAR). The biggest three categories according to BAT are: 100 m², 160 m² and 280 m² with 17,65%, 14,71% and 17,65% respectively.

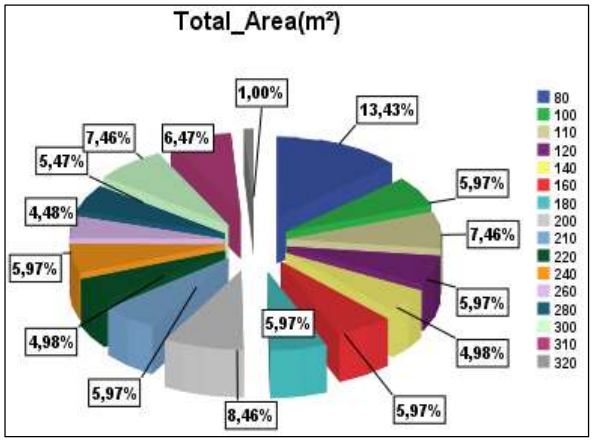


**Fig 3.29** Distribution of houses according to their total area

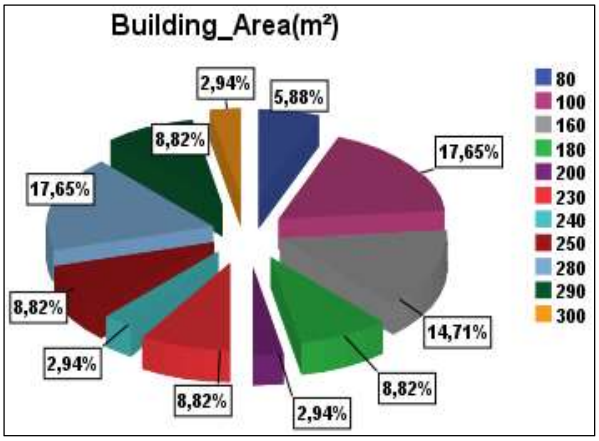


**Fig 3.30** Distribution of houses according to their building size

Number of rooms per household (ROMN) is also key factor in physical characteristics, it is related directly life quality. Table 3.9 demonstrates that ROMN ranges between **2** and **13** rooms. The mean is **6** per house. Pie chart in Figure 3.31 shows the repartition of houses according to their number of rooms. The biggest category is 4 rooms and represents 22,39%.

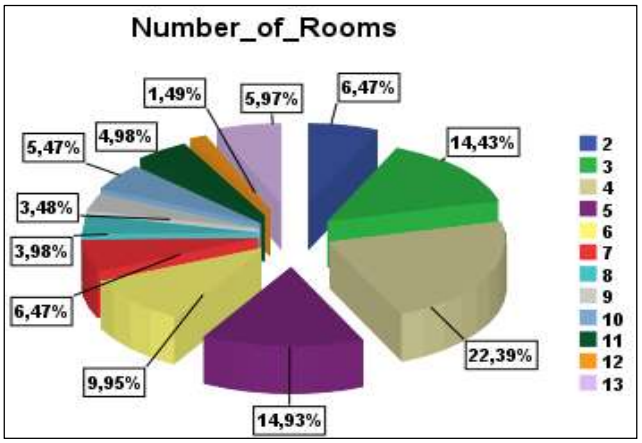


**Fig 3.31** Distribution of houses according to their number of rooms

### 3.4.4.2. Garden Factors

Like building area, garden is an important independent indicator influencing water use. A larger quantity of water is used for gardening and lawn watering. Generally, houses built with garden are higher consumers of water compared to houses without any type of garden. In the present work, all households have a garden.

To measure the influence of garden on WCP, the area of garden (GAR) and watering frequency (GWAT) are used.

The mean of GAR is **21.70 m²** with a maximum of **80m²** and 13 groups. Figure 3.32 shows the repartition of houses according to their garden size. Three main size groups exist are 12, 16 and 20 square meters.

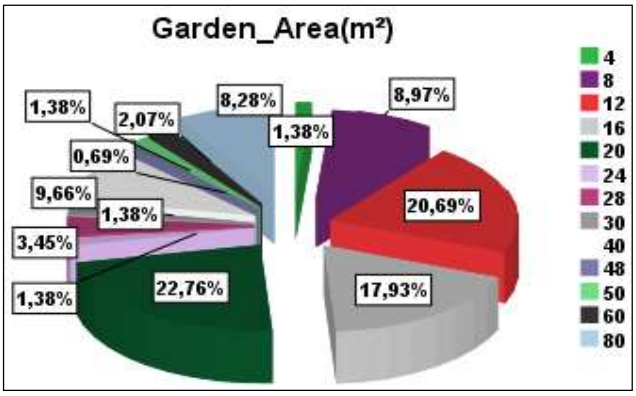


**Fig 3.32** Distribution of houses according to their surface area

Residents of study area are categorized into four groups. 33,83% tend to water their garden once a month, almost a half with 45,77% twice a month, 16,41 % with three times and the smallest portion with 3% four times a month (figure 3,33). Garden watering is also related to weather and other climatic factors.

The general mean of watering frequency is 2 (twice a month).

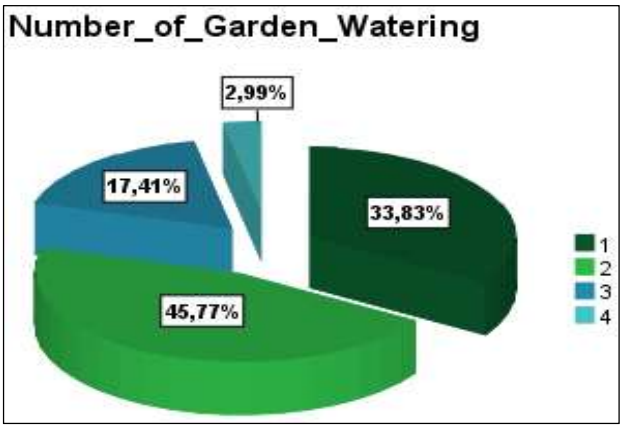


**Fig 3.33** Distribution of houses according to their number of time watering garden

### 3.4.5. Climatic Factors (CLF)

Climatic factors dictate the WCP by influencing the human behaviours in form of changing hygienic and outdoor practices (people tend to bath more in warm season, water gardening, etc.).

The climate in Sedrata is warm and temperate in general. The rain falls mostly in the winter, with relatively little rain in worm seasons. Sedrata climate is classified as Csa by the Köppen-Geiger system. The average annual temperature is 14.2 °C | 57.7 °F. The annual rainfall is 523 mm | 20.6 inches (Website of climate data).

The impact of climate on domestic WCP was assessed through the monthly mean precipitation 'PRE' and mean temperature 'TEM'. The values of mean CLF are illustrated in figure 3.34. The link between temperature and precipitation is clear where hot seasons are characterized with low precipitation and vice versa.

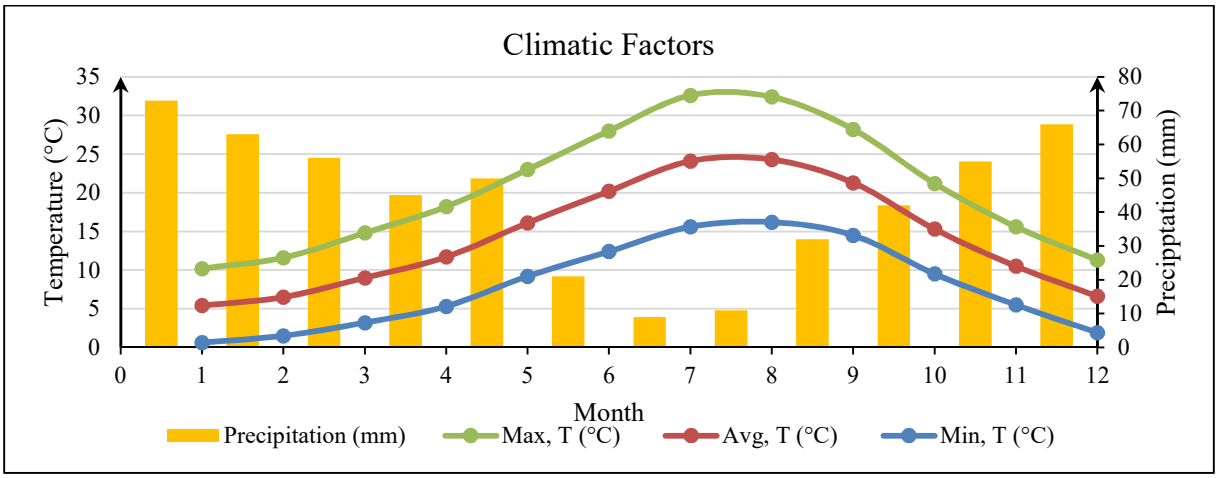


**Fig 3.34** Monthly precipitation (mm) and temperature (°C)

The highest rainfall **PRE** (January **(73mm)** and December **(66mm))** correspond to coldest months, and same remark for lowest PRE and TEM (figure 3.34). This variation is typical in north Algeria and covers all the region of south Mediterranean.

About 88% of total annual rainfall occurs during the period from September to May and the rest 12% in the remaining period. The rainfall in the region shows spatial disparity and due to very rough and uneven terrain, droughts in some parts of the region and floods in some others.

On yearly scale, precipitation is highly non predictable for example the difference between 2015 and 2016 is 480 mm that represents 44,7% decrease.

## 3.5. Data Preparation

In any statistical analysis, the collected data must be prepared. The figure 3.35 below shows the adopted methodology of data preparation:

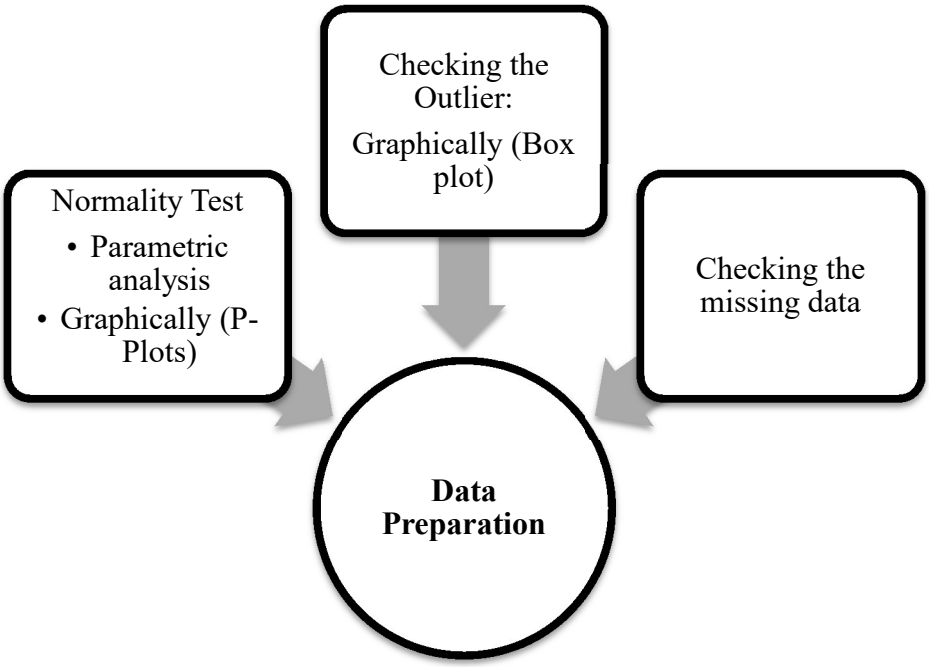


**Fig 3.35** Data preparation methodology

### 3.5.1. Checking the Normality of Parameters

In most traditional statistical models, the data must follow a normal distribution before the model coefficients can be estimated efficiently. If this is not the case, the distribution of data should be applied. The normal distribution is one of the most important and the most widely used example of a continuous random variable. Normal distribution has a bell-shaped curve with the center of the bell located at the arithmetic mean **(μ)**.

The standard deviation **(σ)** controls the depth of this bell. It expressed by **N (μ, σ²)** and defined by two arithmetic parameters; **mean (μ)** and **variance (σ²)**.

To confirm the normality distribution, **"Skewness"** and **"Kurtosis"** are calculated using Statistical Package for the Social Sciences (SPSS) Software.

**Skewness** characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending towards more positive values. Negative skewness indicates a distribution with an asymmetric tail extending towards more negative values (Microsoft, 1996). Skewness value indicates the distribution side and closeness of it to zero means symmetric distribution of parameter. In normal distribution, the **mean** and **median** values should be **very close** to each other and produce a skewness statistic of about Zero.

**Kurtosis** characterizes the relative peakedness or flatness of a distribution compared to the normal distribution. Positive kurtosis indicates a relatively peaked (leptokurtic, too tall) distribution and negative kurtosis indicates a relatively flat (platykurtic) distribution or even concave if the value is large enough (Microsoft, 1996). Normal distributions produce a kurtosis statistic of about **zero**.

For a random variable *x*, when *log (x)* probability distribution is normal, probability distribution if *x* is defined as logarithmic normal distribution (log-normal distribution). In this case, the distribution of parameter deviates from the symmetry. Also, checking the normality can be done using *Z* score of *Kurtosis* and *Skewness* of each variable. It calculated by the following formula:

$$Z_S = \frac{Skewness}{\sqrt{\frac{6}{N}}} \qquad (3.1)$$

$$Z_K = \frac{Kurtosis}{\sqrt{\frac{24}{N}}} \qquad (3.2)$$

$Z_k$ and $Z_S$ values should be within ±2.58 for $p$=0.01 or ±1.96 for $p$=0.05 for considering data as normal distributed. In equation below, Z normal distribution is calculated by using a table. $Z_{95}$ value means the Z value corresponding to 0.95 probability value (=1.65).

$$Z = \frac{X-\mu}{\sigma} \qquad (3.3)$$

With: $x$ is $x_{95}$; $\mu$ is mean value; $\sigma$ is standard deviation and Z is $Z_{95}$ value.

According to results in tables 3.2; 3.3; 3.4; 3.8 and 3.9 the distribution of each parameter is given in table 3.10 below. Table 3.10 shows that WCP 2014, WCP 2017 and WCP of first trimester have small and positive skewness and kurtosis, which indicate a normal distribution of data. The rest of water consumption variables have big and positive skewness and kurtosis, which indicate reasonably, have non-normal distribution.

**For socio-economic parameters**, all the parameters have small and negative kurtosis. Only household size, the fourth age category and number of cars have small and negative skewness. Whilst, the rest of socio-economic parameters have positive skewness. As a result, all these parameters have **normal** data distribution.

Total area, building area, number of rooms and frequency of garden watering **in physical characteristics** of building units have small positive skewness and small negative kurtosis. These parameters indicate **normal** data distribution. Only garden area has big positive kurtosis and skewness which reasonably has **non-normal** distribution. All **indoor habits** variables have normal distribution with small positive skewness and kurtosis for washing clothes, using toilets and shower for female frequencies. While, the frequency of dishwashing has small negative skewness with small positive kurtosis. Also, shower for male has small skewness and small negative kurtosis.

**Table 3.10:** Statistical distribution of parameters

| Statistical Values | Skewness | Kurtosis | Distribution |
|---|---|---|---|
| **Water Consumption** | | | |
| **WCP_2012** | 1,21 | 2,68 | **Non-Normal** |
| **WCP_2013** | 2,33 | 12,67 | **Non -Normal** |
| **WCP_2014** | 0,65 | 0,25 | Normal |
| **WCP_2015** | 3,23 | 20,23 | **Non -Normal** |
| **WCP_2016** | 1,39 | 2,72 | **Non -Normal** |
| **WCP_2017** | 0,98 | 0,84 | Normal |
| **Mean_WCP_1_Trimester From 31 Jan to 31 Mar** | 0,89 | 1,01 | Normal |
| **Mean_WCP_2_Trimester From 31 Mar to 31 Jun** | 1,02 | 2,01 | **Non -Normal** |
| **Mean_WCP_3_Trimester From 31 Jun to 31 Sep** | 1,33 | 3,37 | **Non -Normal** |
| **Mean_WCP_4_Trimester From 31 Sep to 31 Dec** | 1,95 | 7,91 | **Non -Normal** |
| **Socio-economic parameters, Physical characteristics of buildings & Indoor habits** | | | |
| **FEM** | 0,24 | -0,51 | Normal |
| **MAL** | 0,38 | -0,16 | Normal |
| **HOUS** | -0,37 | -0,1 | Normal |
| **AG1** | 0,05 | -0,47 | Normal |
| **AG2** | 0,51 | -0,73 | Normal |
| **AG3** | 0,46 | -1,21 | Normal |
| **AG4** | -0,32 | -0,98 | Normal |
| **PRS** | 0,73 | -0,44 | Normal |
| **MDS** | 0,22 | -0,78 | Normal |
| **HGS** | 0,81 | -0,6 | Normal |
| **UNIV** | 0,08 | -1,10 | Normal |
| **INC** | 1,09 | -1,13 | Normal |
| **CARN** | -0,33 | -0,72 | Normal |
| **WCAR** | 0,85 | -0,05 | Normal |
| **WDISH** | -1,13 | 0,24 | Normal |
| **WCL** | 0,84 | 0,87 | Normal |
| **UTLT** | 0,62 | 0,38 | Normal |
| **FSHW** | 0,59 | 1,16 | Normal |
| **MSHW** | 0,53 | -0,52 | Normal |
| **TAR** | 0,15 | -1,27 | Normal |
| **BAR** | 0,19 | -1,19 | Normal |
| **ROMN** | 0,92 | -0,20 | Normal |
| **GAR** | 2,16 | 4,12 | **Non -Normal** |
| **GWAT** | 0,56 | -0,22 | Normal |

To confirm the results from table 3.10 above, graphs like Histograms and probability plots are obtained for this purpose.

Probability plots are for variables cumulative proportions against the cumulative proportions of any of a number of test distributions. They are generally used to determine whether the distribution of a variable matches a given distribution. The points cluster around a straight line if variable matches the test distribution. P-P graphs plot the cumulative probabilities (values range from 0 to 1), with observed probabilities (cumulative proportion of cases). In normal P-P plots, normal distribution of data set is on y-axis (expected cumulative probabilities), while in log-normal P-P plots, log-normal distribution of data set is on y-axis (expected cumulative probabilities).

Basing on these normal and log-normal P-P plots, distribution of data set is confirmed graphically by related their distribution type in figures 3.36 below.

**Fig 3.36 P-P plots for variables (ANNEX 03)**

### 3.5.2. Checking the outliers

The best technique to identify the outliers is the **"box plot"**, as illustrated in Figures 3.37 below.

**Fig 3.37 Box plot for variables (ANNEX 04)**

The box plots of variables are examined as part of the diagnostic phase of data preparation, to conduct the statistical and numerical techniques. The figure 3.37 shows the existence of outliers in some variables. This last should be removed from data sets.

## 3.6. Conclusion

Water Scarcity in semi-arid areas is a major problem encountered by city planning. Moreover, providing enough and drinkable water all day still challenging and to better rationalize water usage, all parameters governing water consumption should be considered in detail.

Water Consumption is clearly a multi-variable function where each variable has a different weight and pattern. For measuring reliably their impact on WCP, all related parameters are collected directly from authorities or from questionnaire for indoor/outdoor consumers habits.

The pre-processing task ensured that the obtained data is representative and the subsequent analyses are all valid. In fact, after removing outliers 201 household remained with dataset of 4824 valid water consumption values.

Parameters affecting WCP are categorized into three: indoor habits, socio-economic parameters and physical characteristics of buildings. This distinction between inputs helps to asses separately the significance of every parameter.

One of the most important statistical features is the normality where tests like Kurtosis and Skewness are employed. Finally, when data does not follow a normal distribution, it must be normalized because the AI modelling requires a normality distribution.