

A brief description of domestic water-use data in the city of Rabat-Salé (Morocco)

Mohamed Saffi ^{1,*}, Khalid Achoubir ^{1,*}, Hafid Bourich ¹, and Mouad Iqiqen ¹

¹Mohammed V University in Rabat, Ecole Supérieure de Technologie, Sortie des Arcs, BP 227, Salé, Morocco

Abstract. A small-scale experiment is conducted to estimate domestic water-use rate in the city of Rabat-Salé. Water consumption in 171 households has been closely monitored over a 24-month period. During this survey both the net monthly volume of tap water effectively used in each household and the exact number of people who have lived in the same household are observed. This has resulted in a series of reliable realizations of daily per-capita water rate (denoted q). A usual procedure for data summarizing is presented: It is observed that entity q , interpreted as a random variable, is satisfactorily represented by the log-normal distribution. Estimate for the expected value of q along with the corresponding 95% Cox confidence interval are given. Such estimates are (for instance) essential to future water system extension.

1 Introduction

Morocco is a semi-arid country prone to long periods of drought [1]. Its fresh water resources are scarce and exposed to high contamination risk mainly because of (i) massive and irrational use of fertilizers and pesticides, and (ii) tardy investment in wastewater treatment. For instance, the Sebou River has typified for decades the serious problem of surface water pollution in Morocco e.g. [2]. However, since 2005 a substantial effort has been made to narrow down the gap by investing more than 4 billion US dollars, as part of the national program for wastewater treatment aiming at protecting water resources against contamination and upgrading sanitary and hygiene standards for about 10 million inhabitants [3]. Yet, still large amounts of raw domestic and industrial wastewater are everyday pumped directly into fresh water reservoirs. In parallel, dams which represent one of the main Moroccan assets in terms of water resources control, are (for decades thus far) known to be effectively endangered by silting [4], and eutrophication as well [5, 6, 7, 8]. Unless some non-conventional source of energy is envisaged, the high cost of energy in the Moroccan context makes of seawater desalination not an option to come first despite of the very large national coastline. Morocco is nowadays considered one of the vertices of what is alarmingly referred to as “thirst triangle” which includes North Africa and the Middle East [9]. Under these serious and structural coercions, Moroccan water authority is now (and very likely will be) facing the formidable task of conciliating a depleting and vulnerable national water reserve on the one hand and a steadily growing and pressing demand of the population on the other hand. Consequently, among several measures to consider, accurate estimation of influential parameters is no longer a mere formality; it

has become one of the key steps to take seriously by the Moroccan decision-maker for a genuine rationalization of its water management policy.

This work starts addressing one of these influential parameters namely the rate of domestic water-use (herein denoted generically by q) for urban population.

Generally speaking, q can be defined as the daily average volume of water required by an individual to cover its basic needs. This essential quantity underlies estimation of outflows in water systems. So, it helps with shedding light on the proper size pipes should take.

In light of this, this paper attempts to go a step beyond the ‘rule of thumb’ frequently adopted to get rough-guess of the value of the parameter q . And, it suggests a small-scale experiment to explain how we can develop straightforwardly a well educated-guess of q instead. The next section in this work presents survey and methods. Section three is devoted to the processing of the gathered data following the mainstream steps. Finally, conclusions are given.

2 Survey and data

Parameter q is routinely grossed as the ratio between the billed volume of water at the scale of a city and the city nominal population P believed/assumed to have consumed this amount of water. The problem with this approach is that P is (maybe) ill-defined and oftentimes inaccurately known which results in a rough point estimate of q .

The present work proceeds within a statistical framework by interpreting q as a random variable governed by some probability density function which is dealt with a part of the question. The whole procedure is exemplified by a small-scale survey of 171 households scattered over the city of Rabat-Salé. This survey has

*Corresponding author: achoubirk@gmail.com; saffimo@hotmail.com

covered a period of 24 months and consisted in (i) monitoring the net volume V_{ij} of tap water effectively used in the household number i during month number j ; leaks in water distribution system are bypassed. V_{ij} is taken from, REDAL®, the company in charge of water and electricity distribution systems in Rabat-Salé. (ii) Monitoring the exact number P_{ij} of people who have lived in the household i during the month j . That is, the effective consumers of the billed volume of water V_{ij} : It is in this aspect that this survey differs from conventional consumer surveys. Entities P_{ij} s are reliable data delivered by the families who have accepted to take part in this project.

This approach has the advantage (i) to be based on reliable data coming from an accurately and closely monitored population sample. Of course, the sample and the observation period could be made larger in accordance with the available funds (ii) to proffer more flexibility thanks to randomness in q which helps better explain the large variations in its observed realizations.

For the time being entity q is dealt with as a global indicator characterizing the whole population abstracting from explanatory factors such as income, household size, consumer behavior and so on. These regressors are documented for instance in [10, 11, 12].

It is noteworthy that the intended meaning given herein to domestic water-use includes the following main and basic needs such as drinking; cooking; bathing; toilet flushing; washing (utensil/clothes); cleaning. However special activities, like filling a swimming pool, are not included. As a de facto practice, such water-demanding activities are usually satisfied via private and hidden (i.e. illicit) wells. A methodology for detecting massive illegal pumping is suggested in [13, 14].

The daily per-capita domestic water-use rate q , interpreted as a random variable, has a realization q_{ij} based on the household i during month j :

$$q_{ij} = \frac{V_{ij}}{P_{ij}D_j} \quad (1)$$

D_j denotes the effective number of days covered by the bill corresponding to month number j . The other entities have been defined above. Now, for a given household i the sequence of realizations $(q_{ij})_{1 \leq j \leq 24}$ is 'homogenized' as follows:

$$q_{ij} = \frac{q_{i1} + q_{i2} + \dots + q_{i,24}}{24} \quad (2)$$

Finally, the numerical value q_i is taken as an outcome of the random variable q and the observed dataset would be the sequence q_1, q_2, \dots . From the first glance, it is clear and definite that the q_i s span a substantially large range going from 28 up to 314 liters per day and per capita (l/cap d). So, proceeding with q as a random variable seems to be reasonable; and it would be rather simplistic to boil it down to a nominal constant value as it is conventionally conjectured.

3 Results and discussion

The boxplot in Figure 1a shows that the sample data are positively skewed. To alleviate the asymmetry, a new dataset x_1, x_2, \dots is formed by taking the natural logarithm of q_1, q_2, \dots . This transformation helps map the zero-lower bound random variable q into a variable $X = \log(q)$ with tails that stretch to infinity; symbol 'log' stands for the natural logarithm. It is one of the most commonly used transforms in preprocessing water resources data. It is remarkably helpful in modeling transmissivity fields in hydrogeology [15]. The famous logarithm transform has been successfully applied to other natural phenomena as discussed in [16] (to cite a few). Firstly, it appears from Figure 1b that both the box and the whiskers corresponding to the log-transformed dataset are approximately symmetrical.

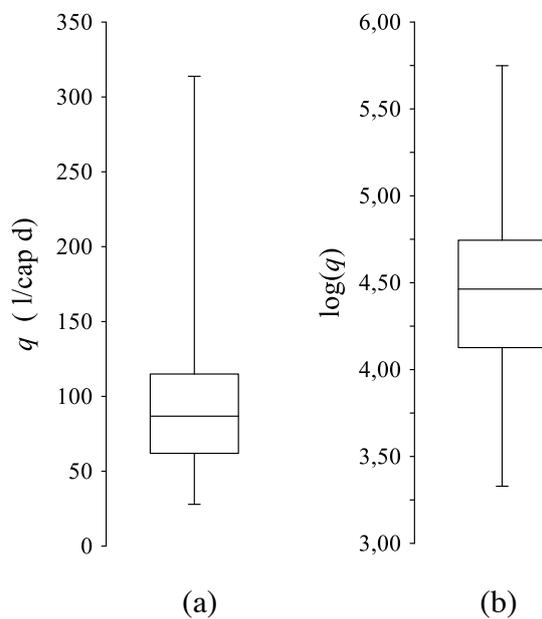


Fig. 1 .Boxplot for daily per-capita domestic water-use rate q of 171 households monitored during a two-year period in the city Rabat-Salé (Morocco) (a) untransformed data (b) log-transformed data.

A summary of the main descriptive sample statistics is compiled in Table1. It appears that (i) the sample mean and the median take close values (ii) the interquartile range IQR is approximately equal 1.30 times sample standard deviation, and (iii) The data range covers more than five times s_x . These distinctive features are in good agreement with the conventional characteristics of a normal distribution.

Secondly, a histogram for the sample x_1, x_2, \dots , reported in Figure 2 recalls the well-known bell-shaped appearance of a normal distribution. For 171 data points, eight intervals have been used as recommended in [17]. The normal density function with parameters $\bar{x} = 4.4598$ and $s_x^2 = 0.2268$ is then plotted (dashed line) against the aforementioned histogram. It looks like a free-hand sketch of what an idealization of the probability density function of X appears to be.

Quantitatively, this intuition translates into the fact that 119 out of the 171 observations (i.e.69.6%) lie within $\bar{x} \pm s_X$, and 160 observations (i.e. 93.6%) lie within $\bar{x} \pm 2s_X$ which is again in good agreement with what is conventionally expected from a normal distribution. That is, 68% (95%) of the observations should be within one (two) standard deviation(s) off the mean, respectively. This indicates that the repartition of the log-transformed data points is not over peaked around the mean and its tails come down at a rate consistent with that of a normal distribution.

Table 1.Summary statistics of the log-transformed domestic water-use data corresponding to the 171 monitored households in the city of Rabat-Salé (Morocco).

Statistic	Observedsample value
Mean	4.4598
Median	4.4635
Standard deviation s_X	0.4762
Upper quartile	4.7444
Lower quartile	4.1265
Interquartile Range	0.6179
Min(x_1, x_2, \dots)	3.3287
Max(x_1, x_2, \dots)	5.7488
Skewness S	0.1496
Excess kurtosis E	-0.0775

Thirdly, the scatter plot in Figure 3a depicts the observed z-scores for data points x_1, x_2, \dots computed as $(x_i - \bar{x})s_X^{-1}$ against the expected z-scores found by using the table of probabilities of the standard normal distribution. In this paper, the Blom plotting position formula, widely accepted to be the best for comparing data quantiles against a normal distribution, is used. It appears that the set of paired observed/expected z-scores fall approximately along the 45-degree line passing through (0,0) with a satisfactory correlation coefficient of 0.998 which is another milestone witnessing the normality of the log-transformed data; this is to be paralleled with the banana-shaped plot of Figure 3b associated with untransformed data.

The foregoing threefold visual inspection of data provides grounds for admitting normality of the random variable X , but it remains open to some subjective interpretation. So, conducting a normality test should be the next step to take. And, the null hypothesis would be the following:

H_0 : The studied sample x_1, x_2, \dots comes from a normal distribution; against the alternative hypothesis; H_a : non-normal population distribution.

The Jarque-Bera test is applied by considering the following statistic [18]:

$$JB = n \left(\frac{s^2}{6} + \frac{E^2}{24} \right) \quad (3)$$

S is the sample skewness, $n=171$ stands for the sample size. S measures the asymmetry in the observed dataset (x_i). The excess kurtosis E assesses how thick-tailed is the observed distribution. S and E are given by:

$$S = \frac{n}{(n-1)(n-2)} \sum_{k=1}^n \left(\frac{x_k - \bar{x}}{s_X} \right)^3 \quad (4)$$

$$E = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{k=1}^n \left(\frac{x_k - \bar{x}}{s_X} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (5)$$

The foregoing expressions for S and E suit small size samples as well. For a normal distribution both S and E are equal to zero, their successive values are reported in Table1.

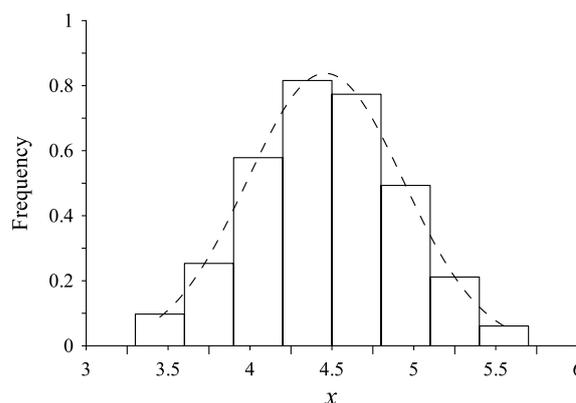


Fig. 2 A histogram (solid line) of the log-transformed dataset of the per-capita domestic water-use rate q of the city of Rabat-Salé (Morocco) versus the corresponding normal probability density function (dashed line) with parameters $\bar{x} = 4.4598$ and $s_X^2 = 0.2268$.

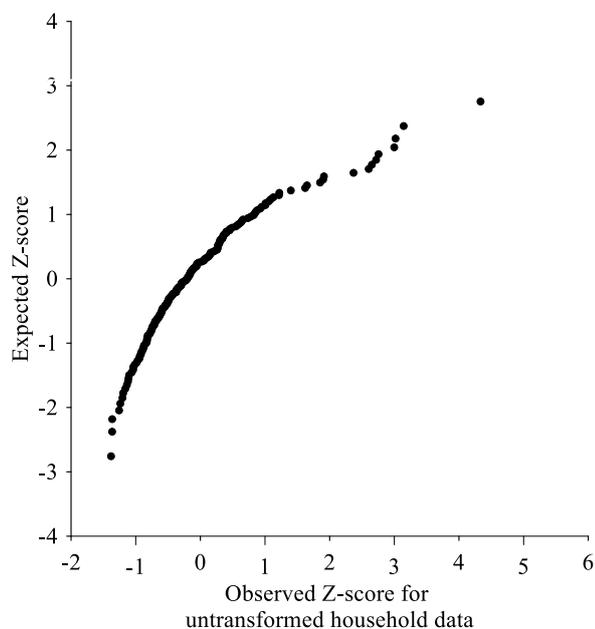
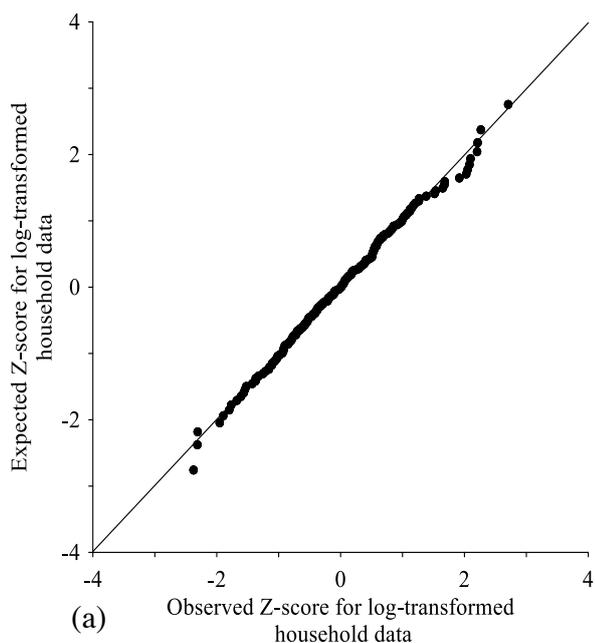
And the Jarque-Bera statistic JB condenses the effects of skewness and excess kurtosis. The test consists in comparing the value of this statistic to the Chi-square distribution with 2 degrees of freedom. For the studied sample, $JB=0.68$ is less than $\chi_{critical}^2 = 5.99$ corresponding to the preselected significance level $\alpha = 0.05$. So, the null hypothesis is consistent with the data and there is no reason (insofar) to reject H_0 . In other words, the distribution of the daily per-capita domestic water-use rate q for the city of Rabat-Salé seems to be well fitted by the log-normal probability density function.

An estimate (denoted $\hat{\mu}_q$) of the expected value of q is not recovered simply by taking the antilog of the sample mean \bar{x} , the sample variance s_X^2 should also be accounted for according to the following well-established relationship [19, 20]:

$$\hat{\mu}_q = \text{Exp} \left(\bar{x} + \frac{s_X^2}{2} \right) \quad (6)$$

Equation (6) combined with data from Table1 yields for the city of Rabat-Salé:

$$\hat{\mu}_q = 96.9 \text{ l/capd} \quad (7)$$



(b)

Fig. 3 Normal probability plot for the studied dataset of the city of Rabat-Salé (Morocco): (a) log-transformed data (b) untransformed data.

Various methods dedicated to computing confidence intervals for the mean of a log-normal distribution are compared in the literature e.g. [21, 22, 19]. In this study, we use the standard Cox’s method which performance remains satisfactory even when the sample size is not sufficiently large. The procedure is carried out in two steps:

First, the lower and the upper bounds, L_{cox} and U_{cox} , of the confidence interval for $\log(\mu_q)$ are evaluated:

$$L_{cox} = \bar{x} + \frac{s_x^2}{2} - z \sqrt{\frac{s_x^2}{n} + \frac{s_x^4}{2(n+1)}} \quad (8)$$

$$U_{cox} = \bar{x} + \frac{s_x^2}{2} + z \sqrt{\frac{s_x^2}{n} + \frac{s_x^4}{2(n+1)}} \quad (9)$$

z is the percentage point of the standard normal distribution corresponding to the desired level of confidence; $z=1.96$ for a 95% confidence level. In [22] use of t - statistic is proposed instead of z . The term under the square-root symbol in equations (8-9) represents an estimate of the variance of $\hat{\mu}_q$ (denoted by $\hat{var}(\hat{\mu}_q)$); [19] handled it with $(n - 1)$ in the denominator instead of $(n + 1)$. The amendment reported in equations (8-9) was made by Chami et al [21], in their paper’s subsection ‘Cox method’, in order to render $\hat{var}(\hat{\mu}_q)$ unbiased. For large samples this would make no difference. According to Table1, $L_{cox}=4.4971$ and $U_{cox}=4.6485$.

Second, the anti-logarithm of L_{cox} and U_{cox} are taken in order to recover the confidence interval for the mean value of the domestic water-use rate q . That is, one can be 95%-confident that the expected value of q lies between $Exp(L_{cox})$ and $Exp(U_{cox})$ i.e. between 89.8 and 104.4 liters per capita per day (l/cap d).

3 Conclusions

This work outlines a preliminary investigation of the daily per-capita domestic water-use rate q in the city of Rabat-Salé (Morocco). A follow-up is to come in a near future. The dataset, upon which this study is based, is obtained by surveying 171 households over a 24-month period. Conventionally, q is simply approached as a ratio between the billed volume of water and some nominal population. The latter is ill-defined and entails good deal of arbitrariness. In contrast, this study suggests interpreting q as a random variable which realizations are based on a well-defined and closely monitored population sample.

The assumption that q is a random variable is well supported by the data coming from the survey. That is, the observed values of q range from 27.9 up to 313.8 liters per capita per day. And a statistical approach seems to be the adequate framework to assimilate naturally a dataset of this kind. It is observed that the data are well fitted by the log-normal distribution. Therefore, rationally defined estimate (i.e. beyond the so-called rule of thumb) of the mean value of q (equation (7)) along with the associated confidence interval (equations (8-9)) are given in the case of the city of Rabat-Salé. Such statistics are essential for planning future water mains for Moroccan cities.

Acknowledgement. The authors would like to thank families who have participated in the survey and have offered their help by supplying him with their personal information. He also thanks Mr I. Muad, Mr D. Ayub and Mr L. Alae for their help while gathering data.

The authors are grateful to the two anonymous referees for helping with improving the manuscript form.

References

1. L. Stour, A. Agoumi, *Hydroécologie Appliquée*, **16**, 215–232 (2008)
2. M. Saffi, D. Ouazar, *Sag curve and river parameters determination: The Sebou River*, in Proceedings of Water Industry Systems, Volume I, pp 241-254, Exeter, UK (1999).
3. Ministry of Energy, Water and Environment. *Water pollution sources in Morocc., Department of Water Pollution Control*. Agdal, Rabat. (2014)
4. A. Lahlou, *Envasement des barrages au Maroc* (Wallada, Casablanca, 1986)
5. A. Bouloud, A. Foutlane, L. Revue H.T.E (Maroc), **119**, 21–24 (2001)
6. Y. Elghachtoul, M. M. Alaoui, H. Gabi, *Rev. Des Sci. De L'Eau*, **18**, (numéro hors série) 75–89 (2005)
7. M. Saffi, D. Ouazar, *Apport de l'approche compartimentale pour l'analyse de l'aération des lacs réservoirs*, LASH, LASH/RR-301299, EMI, Rabat, Décembre (1999).
8. M. Saffi, D. Ouazar, *Apport de l'approche compartimentale pour l'analyse de l'aération des lacs réservoirs: Application et validation méthodologiques*, LASH, LASH/RR-280200, EMI, Rabat, Février (2000).
9. S. Dionet-Grivet, *Géopolitique de l'eau [Water geopolitics]*. (Ellipses, Paris, 2011)
10. J. F. Kenny, K. E. Juracek, *Description of 2005-10 domestic water use for selected U.S. cities and guidance for estimating domestic water use*, USGS Scientific Investigations Report 2012–5163, Reston, Verginia (2012).
11. J. Schleich, T. Hillenbrand, *Determinants of residential water demand in Germany*, Working Paper Sustainability and Innovation, No. S 3/2007, Institute Systems and Innovation Research, Fraunhofer (2007)
12. Y. Shan, L. Yang, K. Perren, Y. Zhang, *Household water consumption: Insight from a survey Greece and Poland*, in the Proceedings of the 13th Computer Control Industry Conference, CCWI, De Montfort University. (2015)
13. M. Saffi, A. Cheddadi, *Hydrol. Sci. J.*, **52**(4), pp 763-776 (2007)
14. M. Saffi, A. Cheddadi, *Hydrol. Sci. J.*, **55**(8), pp 1348-1356 (2010)
15. R. J. Hoeksema, P. K. Kitanidis, *Water Resour. Res.*, **21**(4), 563-572 (1985)
16. E. Limpert, W. A. Stahel, M. Abbat, *Bioscience*, **51**(5), 341–352 (2001)
17. R. L. Iman, W. J. Conover, *A modern approach to statistics*. (John Wiley and Sons, New York, 1983)
18. C. M. Jarque, A. K. Bera, *Int. Stat. Rev.* **55**(2), 163–172 (1987)
19. X. H. Zhou, S. Gao, *Statistics in Medicine*, **16**, 783–790 (1997)
20. C. Walck, *Statistical distributions for experimentalists*, Internal Report SUF–PFY/96–01. University of Stockholm. (2007)
21. P. Chami, A. Robin, S. Ashok, *J. Appl. Sci.*, **7**(13), 1790–1794 (2007)
22. U. Olsson, *J. Educ. Stat.*, **13**(1) (2005)