Final Report on Household Water Consumption Estimates

By Edgar Morgenroth

Economic and Social Research Institute

Produced for Irish Water

21st of July 2014

1. Introduction

Irish Water require estimates of household water consumption and water consumption differentiated by the number of persons in the household. While some data on household water consumption has been collected in Ireland by Group Water Schemes, this data is not available for closer analysis. For this reason, Irish Water collected data on water consumption for a sample of households where water meters had been fitted early (Phase 1). Data for a second sample was also collected where the number of children is identified (Phase 2). This note outlines the analysis undertaken using this data collected by Irish Water to identify per capita water consumption for different household sizes (Phase 1) and the difference between the consumption of adults and children (Phase 2)[1]. In particular it describes the data collected in the two phases and identifies the water consumption patterns observed in the samples. The analysis considers the influence of unusual data points – outliers – and also adjusts the results for differences in the distribution of household sizes in the samples compared to the CSO Census of Population 2011.

2. Per Capita Water Consumption by Household Size

The data collected by Irish water as part of Phase 1 encompasses 1650 households. Apart from household water usage which is based on meter readings for a three month period[2], the data includes details on the property type, the number of persons in the household and the number of bedrooms of the property. The largest household size includes households with six or more persons so that households with more than 5 individuals are not identified by the exact number of persons per household.

Table 2.1 shows the breakdown of the households by number of bedrooms in the property and property type. Just 5 households live in one bedroom properties and just 34 apartments are included in the sample. In the subsequent analysis no significant relationship between water consumption and the number of bedrooms and the housing type was uncovered, suggesting that for this sample at least these variables do not determine water consumption.

Table 2.1. Distribution of the Sample by Number of Bedrooms and Type of Property.

Bedrooms	Number of Households	Type	Number of Households
1	5	Apartment	34
2	150	Detached	456
3	840	Semi-detached	774
4	580	Terrace	386
5	75		
Total	1650		1650

[1] It should be noted that householders in the sample may have reduced their water consumption following the installation of meters and in anticipation of future usage related bills. It is therefore possible that the estimates already incorporate the expected reduced demand for water in response to the introduction of explicit charges.

[2] The fact that the readings only caver a three month period could omit important seasonal differences in household water consumption, for example due to watering of plants in the summer or if households leave taps running in winter to avoid frozen pipes.

The key issue of the analysis here is the level of water consumption per person. A histogram of the distribution of per capita water consumption across all households is shown in Figure 2.1 for different occupancies. The horizontal axis measures the per capita water consumption and the vertical axis measures the fraction of the observations in the sample at each level of water consumption. A plot of the normal distribution for the observed mean is also shown in the graph. The average water consumption for the 1650 households is just under 120 litres per person which corresponds to the peak of the normal distribution plot. However, just over 71% of the observations have a water consumption below the average and the median (corresponding to the level of water consumption which splits the sample into two halves) is just under 99 litres per person. Overall the plot indicates that the data is not normally distributed and that the mean is not a good indicator of the most common level of water consumption. However, in so far as this distribution is representative the mean may nevertheless be a useful magnitude for planning purposes. The same is true for the distribution of per capita water consumption for each household size which is shown in Figure 2.2. The skew in the distribution is particularly marked for single households (top left) which also shows significant heterogeneity across households. Single households could consist of single older people, single professionals or students which will all have different water consumption patterns. Larger households are more likely to contain children, which makes them demographically more similar. This highlights the need to control for confounding factors in future analysis which is not possible here given the lack of socio-demographic indicators for each household.

Figure 2.1. Histogram of Per Capita Water Consumption Across all Sampled Households (with normal distribution superimposed)

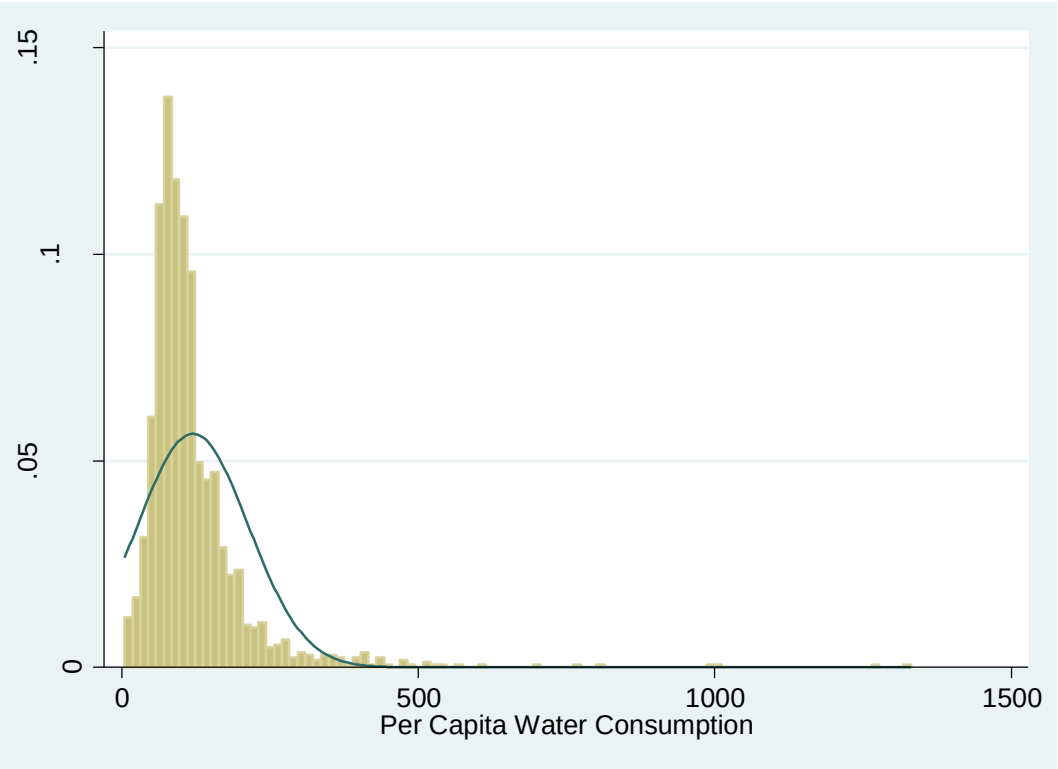
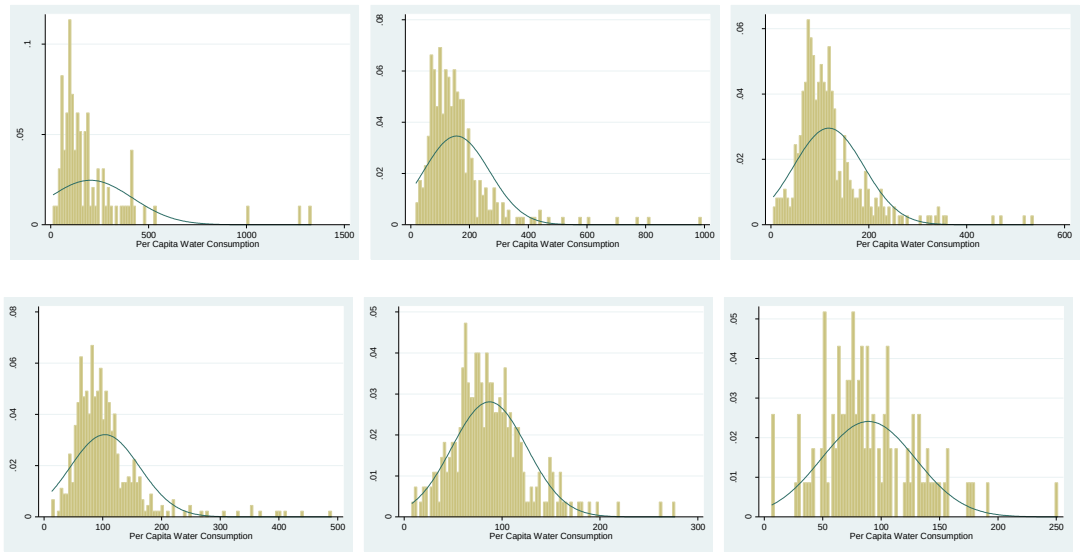


Figure 2.2. Histograms of Per Capita Water Consumptions for Households by Household size



Note: The top left corner shows the water consumption of single households corner and the bottom left corner shows the per capita water consumption for households with six or more persons.

The distribution plots indicate not only that the data observed in the sample is not normally distributed but that there are also observations that are uncharacteristic in terms of water consumption given the distribution of the sample. Such outliers can have a significant influence on the estimated average water consumption. One way to deal with outliers is to apply some (arbitrary) cut-off point beyond which all data observations are discarded for the analysis. An alternative is to carry out statistical tests for the influence on individual observations on the overall estimated mean. One measure of observations with significant influence on the overall estimates was proposed by Welsh and Kuh (1977)[3]. This measure known as DFITS is widely used and robustly identifies outliers on the basis of their influence on the overall result or the fact that the underlying model is incapable of explaining the observation (large error term). On the basis of this measure 81 observation were dropped from the analysis. These include 51 observations for which the per capita water consumption is above the average and 30 for which it is below the average.

Table 2 shows the average water consumption per capita by household size observed in the data using the full sample and the sample with outliers removed. The tables shows that the average consumption is just under 120 litres for the full sample and just under 109 litres once outliers are removed. Single households have the highest per capita consumption and this declines as the household size increases. Table 2.3 shows the increase in water consumption from adding an additional person to the household. A single household consumes the average of the single households (see table 2.2) and adding second person to a single household results in an increase of 108 litres for the whole sample. Adding a sixth or more persons to the household adds nearly 100 litres to the household water consumption.

[3] Welsh, R., and E. Kuh (1977) Linear Regression Diagnostics. *Technical Report 923-977*. Sloan School of Management, MIT. There is also an extensive discussion of the technique in Belsley, D., Kuh, E., and R. Welsh (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Table 2.2. Household Water Consumption per Person by Size of Household

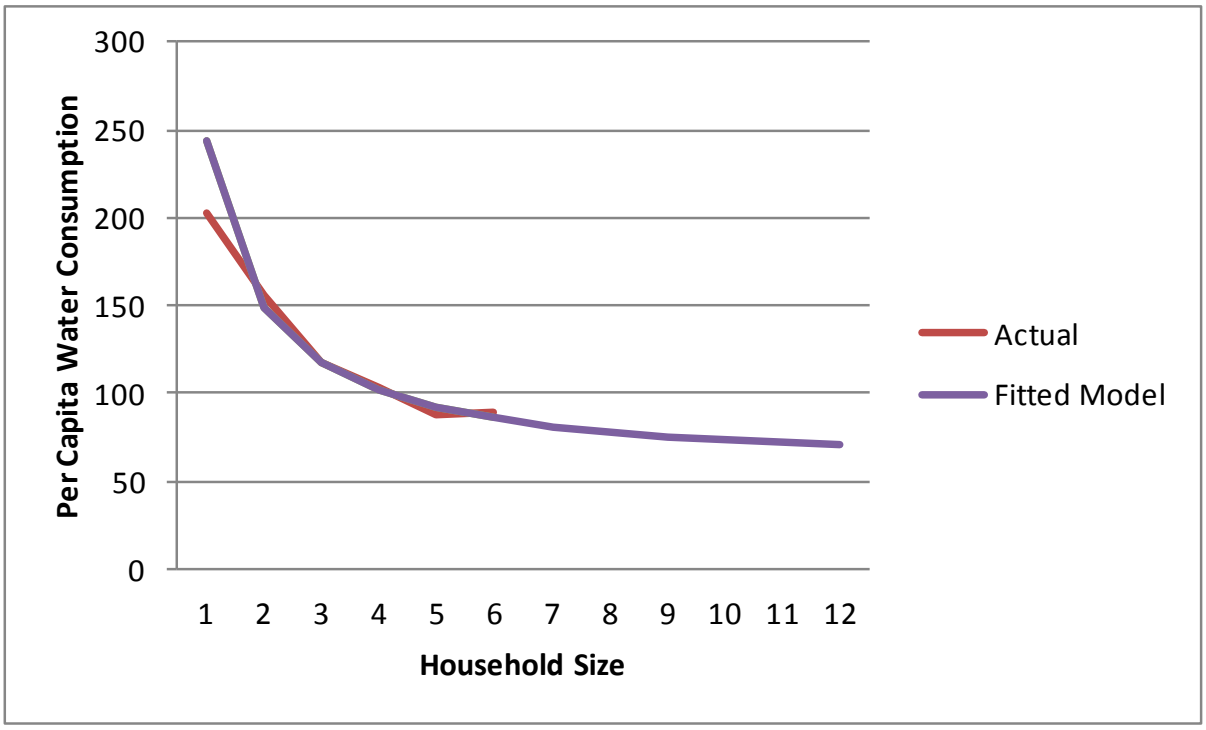
	Full Sample			Outliers Removed		
Occupancy	Number of Observations	Mean Per Capita Water Consumption	Standard Deviation	Number of Observations	Mean Per Capita Water Consumption	Standard Deviation
1	97	203.2	213.7	52	173.6	59.8
2	347	155.6	112.1	332	137.7	64.6
3	367	118.1	71.8	357	110.1	52.6
4	448	103.7	59.2	439	97.7	42.0
5	275	87.1	38.3	274	86.4	36.6
6 or more	116	89.1	40.6	115	87.7	37.8
Total	1650	119.9	93.5	1569	108.8	54.1

Table 2.3. Implied additional consumption per additional household members

Occupancy	Mean	Mean Outliers Removed
1	203.2	173.6
2	108	101.8
3	43.1	54.9
4	60.5	60.5
5	20.7	41.2
6	99.1	94.2

A limitation of the data is that households with six or more persons are grouped together. It is possible to improve on this by using a regression model of household water consumption by household size estimated only for households of 5 or less individuals and using the parameters from this predicting the water consumption for larger households. The results for this are shown in Figure 2.3. Which shows that the predicted average consumption from the fitted model is expected to decline as the household size grows to more than 6 individuals. Furthermore, the estimates predict higher consumption for single households. This result is not dependent on the sample included. The estimates suggests that each additional person consumes 55 litres per day, with a confidence interval (95%) between 49 and 61 litres (±11%).

Figure 2.3. Actual and Estimated Per Capita Water Consumption by Household Size



An important issue is the fact that the average household size in the sample of data deviates significantly from that observed in the CSO Census 2011. In particular fewer single households are observed and more large households were included in the sample. One reason for this is the fact that only household with a resident present were asked about the size of their households and the survey was carried out during the day. Thus, working single households in particular are underrepresented in the data. A comparison of the sample distribution of household sizes with that from the Census is shown in Table 2.4.

Table 2.4. Percentage of Households Classified by Household Size, CSO Census 2011 and the Sample of observations

	CSO	Irish Water
1	8.7%	2%
2	21.3%	12%
3	19.7%	19%
4	23.8%	31%
5	16.1%	24%
6+	10.5%	12%
Total	100.0%	100.0%

Assuming that the observed average household water consumption is representative even though the sample is not representative of the underlying population of households, one can reweight the observed water consumption to reflect the smaller household size found in the Census of Population. Given that smaller households have a larger per capita water consumption this increases

the average water consumption. In particular if one uses the water consumption estimates from the full sample the average consumption rises to 122 litres. Adjusting for outliers this drops to 112.2

3. Per Capita Water Consumption by Children

This section describes the data analysis for the Phase 2 sample of data which consists of 1206 observations[4]. In contrast to the Phase 1 data sample, Phase 2 includes details of household composition broken down by adults and children. Here children are defined as persons aged 18 or less. The distribution of households by household composition is shown in table 3.1. Most of the observations are of households with fewer than four adults and fewer than 4 children (90%). Table 3.2 shows the distribution of the sample across property types. A notable difference between the Phase 1 sample and the Phase 2 sample is the lack of observations for apartments.

Table 3.1 Number of Households by Household Composition

	No Children	1 Child	2 Children	3 Children	4 Children	5 Children	6 Children	7 Children	Total
1 Adult	59	21	17	10	2	1			110
2 Adults	186	177	299	156	37	10	4	1	870
3 Adults	76	41	33	13	2	1			166
4 Adults	33	12	2	1					48
5 Adults	4	2	1	1					8
6 Adults	1	1	1	1					4
Total	359	254	353	182	41	12	4	1	1206

Table 3.2 Distribution of the Sample by Number of Bedrooms and Type of Property.

Bedrooms	Number of Households	Type	Number of Households
1	7	Apartment	0
2	189	Detached	186
3	517	Semi-detached	530
4	434	Terrace	490
5	59		
Total	1,206		1206

Before considering the data further it is useful to analyse the distribution of the per capita water consumption in the data, which is most readily done by considering a histogram of the data. The left panel of Figure 3.1 shows the histogram for the complete sample. A number of households with extremely large per capita usage distort the picture significantly. Once these are removed (using the DFFITS method) the two distributions are quite similar[5]. The mean per capita water consumption (see table 3.3) is 112 litres (125.9 including the outliers) compared to 109 litres (120 litres including outliers) in the Phase 1 data. The medians are also very similar, being 96.5 litres for the Phase 2 data

[4] One observation was deleted as it reported a household containing two children and no adults.
[5] The methodology identified just 16 outliers.

and 98.8 litres for the phase 1 data. Standard statistical tests also suggest that the two distributions appear to be drawn from the same underlying population, but given that the data in the two samples is not normally distributed, standard statistical tests need to be interpreted with caution. All analysis below omits the outliers.

Figure 3.1. Histogram of Per Capita Water Consumption Across all Sampled Households Excluding Outliers (with normal distribution superimposed)

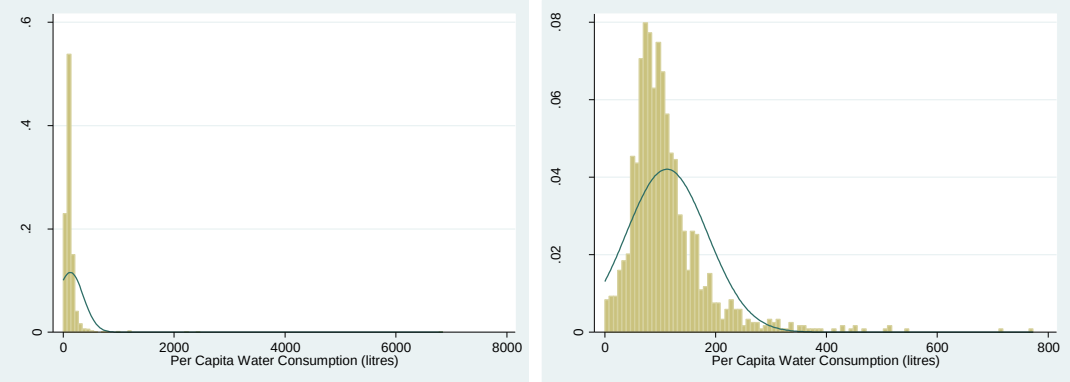


Table 3.3. Household Water Consumption per Person by Size of Household

	Full Sample			Outliers Removed		
Occupancy	Number of Observations	Mean Per Capita Water Consumption	Standard Deviation	Number of Observations	Mean Per Capita Water Consumption	Standard Deviation
1	59	214.6	320.9	58	176.2	127.7
2	207	201.6	504.1	203	149.1	89.4
3	270	116.2	85.8	269	113.7	75.4
4	383	103.9	88.6	381	98.7	51.8
5	207	89.7	48.0	206	89.3	47.7
6 or more	80	96.7	109.2	73	84.3	33.4
Total	1206	125.9	235.7	1190	112.0	73.3

The key question of this analysis is whether adults and children have different level of water consumption. Table 3.4 shows the average per capita water consumption by household composition. This shows that the per capita water consumption declines with increasing number of adults or increasing number of children. In considering these averages it is important to keep in mind that for some household types the number of observations is very small (e.g. there is just one household consisting of three adults and five children in the sample). Only for households with two adults are there sufficient observations to consider the effect of an additional child in the household. Using this table it is possible to calculate the additional water consumption of each child and adult. However a more convenient method is to carry out the calculation using regression analysis, which yields estimates of water consumption per additional child of 53.2 litres and additional adult of 76.1 litres[6].

[6] The regression is estimated over all households without children and households with two adults and children, covering a sample of 1032 (11 outliers are omitted).

The estimated standard errors for these values imply a range between 54 and 98 litres (±29%) for the consumption of an additional adult and between 44 and 63 litres (±18%) for the consumption of an additional child. A formal test reveals that the water consumption of children is only statistically different from that of adults at the 94% confidence level. However, given the non-normality of the data one needs to treat this test with caution.

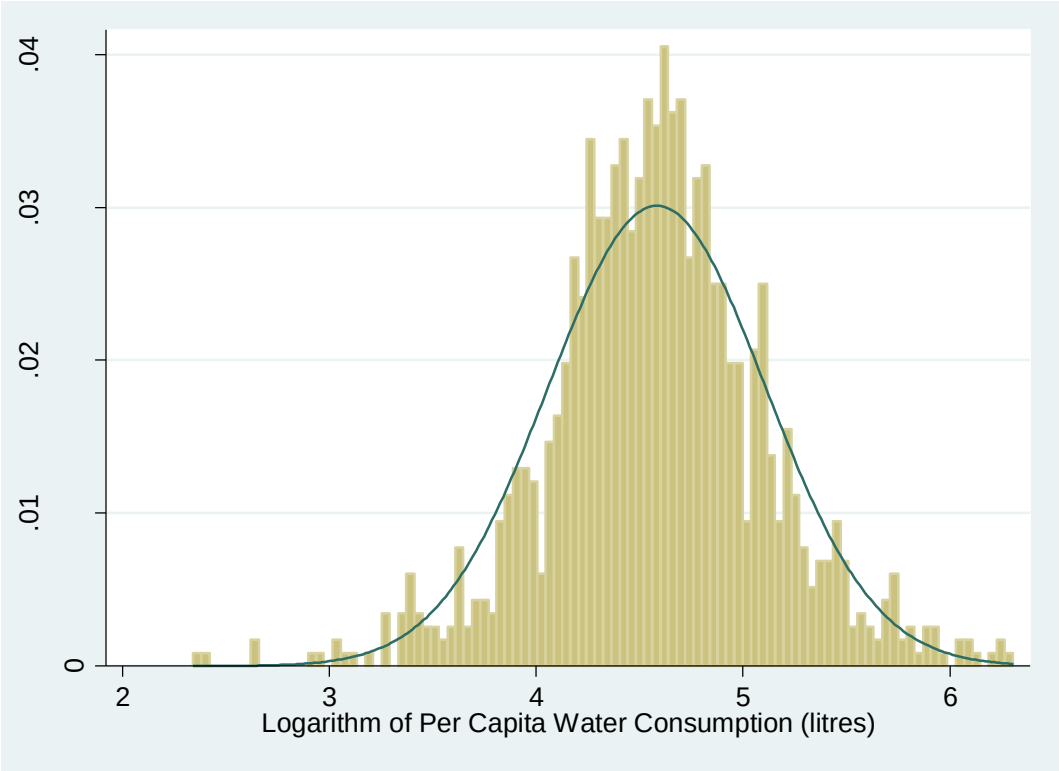
Table 3.4. Average Water Consumption by Household Composition

	No Children	1 Child	2 Children	3 Children	4 Children	5 Children	6 Children	Total
1 Adult	176.2	147.9	96.4	115.7	93.9	126.4		150.8
2 Adults	149.2	114.6	99.0	88.3	83.9	82.1	54.8	109.9
3 Adults	115.6	86.0	96.1	82.1	73.1	79.3		101.1
4 Adults	106.7	79.8	108.7	170.5				101.3
5 Adults	101.3	117.7	77.4	53.5				92.1
6 Adults	176.7		92.5					92.5
Total	142.2	111.0	98.6	89.7	83.9	85.6	54.8	112.0

One way to deal with the non-normality of the data is to use an appropriate transformation. A simple transformation which results in a normal distribution of the water consumption data is to take logs. The average of the logged per capita water consumption data is 4.58 which taking the exponential is 97.8 litres, which is identical to the median of the transformed data. The fact that the logarithmic transformation results in a distribution that is approximately normal can be seen in Figure 3.2 which shows a histogram of the logarithm of per capita household water consumption again excluding outliers[7]. Conducting the same regression analysis yields a set of parameters that can be tested using conventional tests. The hypothesis that the consumption of children is equal to that of adults in the sample data used here can be rejected with 99.4% confidence. Estimating over all households but excluding outliers yields estimates for additional water usage of children and adults that are statistically significantly different at the 90% confidence level (92.6%). It is thus highly likely that the water consumption of children differs from that of adults. However, given the confidence intervals of the estimates this difference may not be very large.

[7] Testing for outliers using the log data revealed additional

Figure 3.2 Histogram of the Logarithm, of Per Capita Water Consumption Across all Sampled Households Excluding Outliers (with normal distribution superimposed)



Summary

Based on the analysis of a sample of household data on water consumption that was collected by Irish Water the per capita water consumption was estimated. Assuming the data is representative of households of each household size and adjusting for the deviation of the distribution of household sizes between the sample and the CSO Census of Population, the average per capita water consumption is estimated to be 122 litres. Removing households with unusually large and small water consumption reduces the estimate to just over 112 litres per capita.

The second part of the analysis concerned the potential difference between the water consumption of children and adults and the level of the consumption of children. The initial data analysis revealed some differences, however, using the raw data which is not normally distributed, conventional statistical test are not valid. Therefore to apply formal statistical tests it was necessary to transform the data. Using the transformed data, the consumption was found to be statistically different between adults and children. The estimated difference is approximately 23 litres per day but it is recommended that further analysis with a sample that includes more households with a larger number of children be conducted. It should also be pointed out that the age of the children will be an important determinant of water usage. The definition of a child in the data used here is any person up to the age of 18. Clearly the water consumption of an 18 year old person is likely to be very similar to that of an adult, while that of a younger child is likely to be lower.

A number of caveats should be noted. It could not be established whether the data, which was collected by Irish Water, is a representative sample of the underlying population of households.

Furthermore, the households in the sample may have reduced their water consumption due to the installation of meters and the fact that the data only cover a three month period, which omits potential seasonal peaks in water consumption, could downward bias the results. Further research using a representative sample should be carried out. Such future research should also account for other confounding variables (e.g. income, appliance ownership, social class, employment status, age etc.) which have been found in the international literature to impact on water consumption. As it is not possible to assess whether the sample is representative of the population with respect to these variables the estimated water consumption may be biased. However, given the relatively small effect from reweighting the sample it is unlikely that accounting for these factors will increase the average consumption to the level that it had previously been assumed to be at (145 litres).