Analysis Report

This report is structured as follows.

Contents

Analytical Approach

The objectives of this analysis was twofold: (1) determine what is the relationship between household size and income levels on the average water bills in Mexico municipalities and (2) determine which municipalities show average water bills that are discrepant from what would be expected based on income and household size.

For this, a regression approach was used. Since the data is nested (multilevel) with multiple observations for each municipality, a dummy-variable approach was used. The municipality variable was transformed on unique dummy variables representing each municipality and the regression equation was fit using the 'General' Escobedo' municipality as reference. Average income was regressed against average water bill with household size as controlling variable. The dummy variables were also inserted in the model to eliminate bias that might originate from the nested nature of the data.

A very large outlier was present in the data (average income of $420,258 for the 5th quintile of San Pedro Garza Garcia). This was removed before the regression took place to ensure no bias was present in the model.

Model Results

The model was significant, $F(9) = 6.289$, $p < .001$, R^2 0.556 and the coefficients are shown below. Interestingly, household size was not a significant determinant of average water bill, while income has a positive effect on the average spend ($B = 0.011$, $P < .001$). In other words, Income determines water usage, not household size. According to the model, for each additional dollar of income, an increase of $0.011 in the water bill is expected.

Coefficients[a]

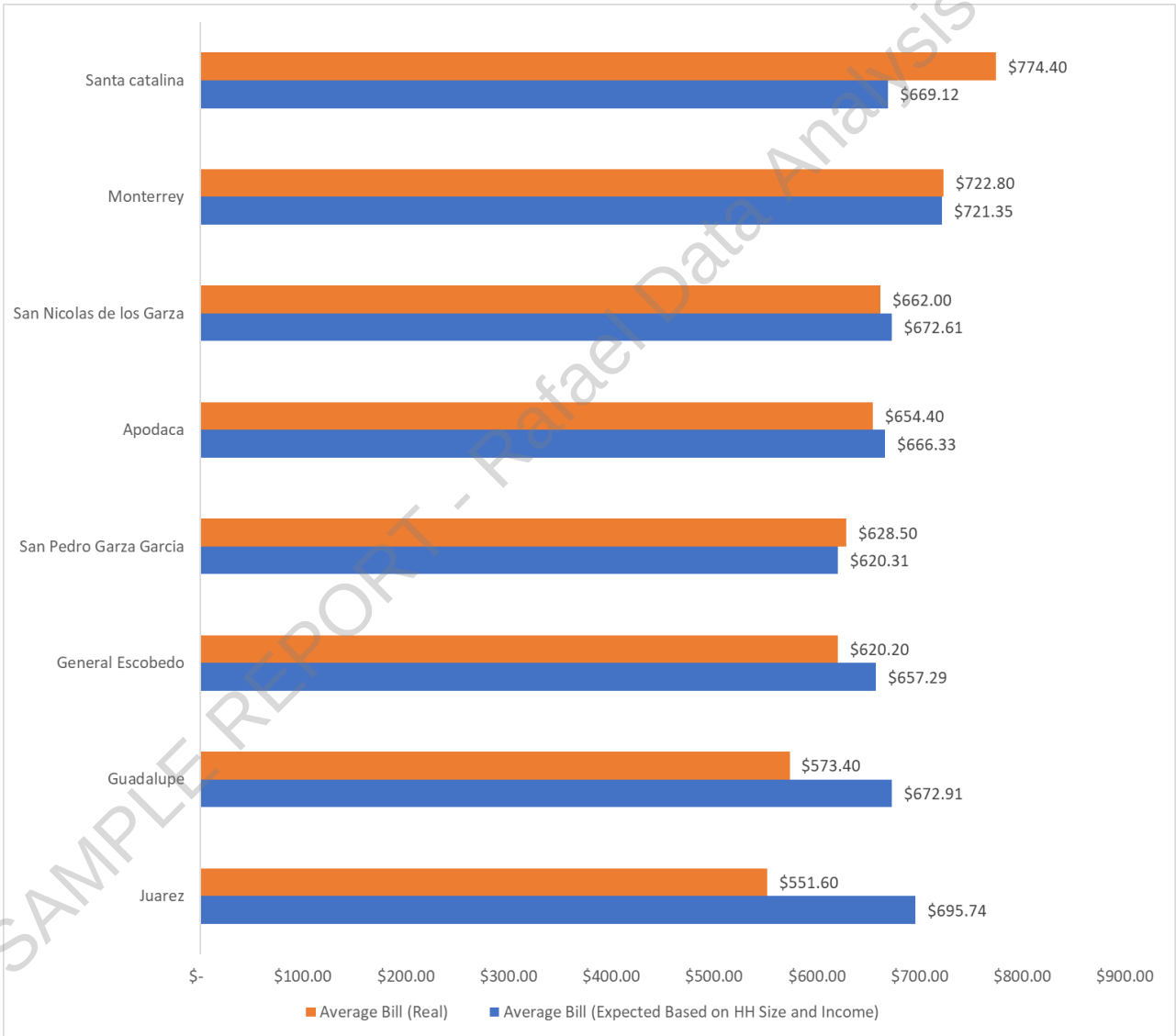
Model	Unstandardized Coefficients		Standardized Coefficients	t	p
	B	Std. Error	Beta		
(Constant)	463.595	125.684		3.689	0.001
Income monthly	0.011	0.002	0.773	5.111	0.000
Average number of people per household	-10.671	38.302	-0.045	-0.279	0.783
Municipality=Apodaca	27.806	85.050	0.047	0.327	0.746
Municipality=Guadalupe	-62.889	84.750	-0.106	-0.742	0.464
Municipality=Juarez	-100.777	86.567	-0.170	-1.164	0.254
Municipality=Monterrey	39.605	85.258	0.067	0.465	0.646
Municipality=San Nicolas de los Garza	22.556	86.192	0.038	0.262	0.795
Municipality=San Pedro Garza Garcia	39.378	91.374	0.060	0.431	0.670
Municipality=Santa catalina	142.837	84.653	0.241	1.687	0.102

a. Dependent Variable: Average water bill

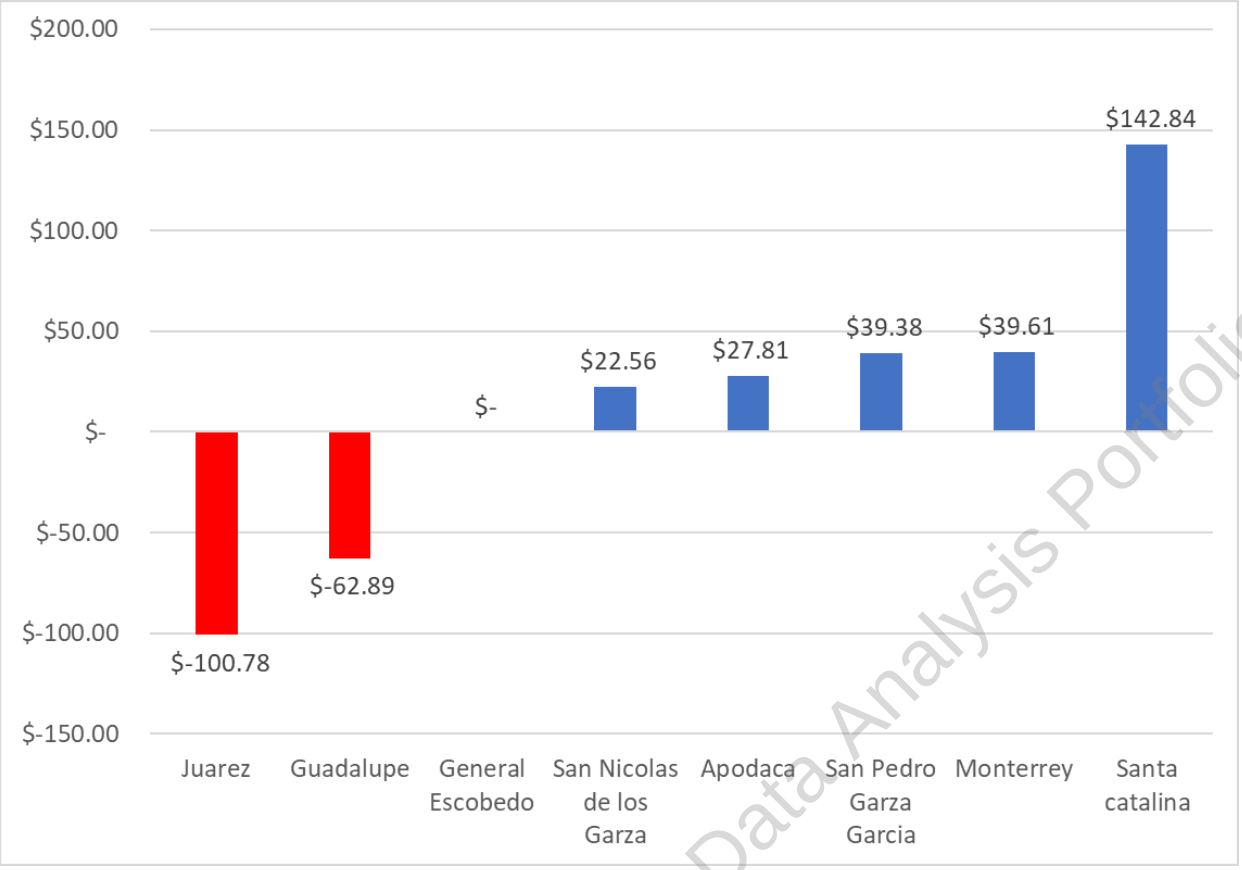
Comparison Analysis

Besides establishing the relationships described in the earlier section, the regression equation was used to calculate what would be the expected water bill based on the model, for each quintile. The quintile values were averaged for each municipality and compared to the real observed average bill. Any discrepancies between these values would mean that some additional external unobserved factor, other than income or household size, is influencing the average monthly bill for that particular municipality.

The graph below shows the comparison of expected versus real values for each municipality.



The graph below shows the differences between expected and real bill values for each municipality. It is important to notice that the values should be interpreted in a comparative fashion, with 'General Escobedo' being the reference municipality (as per the assumptions of nested data regression).



The graph shows that Santa Catalina, followed by Monterrey, San Pedro Garza Garcia, Apodaca and San Nicolas de los Garza are showing average bill values higher than would be expected based on income and household size, suggesting other factors are influencing water consumption. This indicates that these municipalities are good targets since they are using more water than would be expected.