An Analysis of the Price Elasticity of Waste

Part 1: Introduction

<u>Section A: Criterion and Predictor Variables</u>

Waste management systems have been an integral part in the evolution of human society, varying in scope and size throughout history. Regardless of industry or location, economic activity results in the production of waste. In developed economies, waste management – from sewage and water treatment to industrial and hazardous waste removal – is largely hidden from the public eye, despite individuals' daily dependence on it. An economy's waste management system is an intricately woven and broadly defined network that must have the capacity to withstand consumers' and producers' total waste output. To accurately forecast a system's capacity, we must know whether waste is a normal good, and if there is a point where it transforms to, or from, an inferior good. To study this question, I will be analyzing data provided by the Texas Commission on Environmental Quality (TCEQ). Licensed landfills are required by law to provide annual reports detailing the aggregate and diversity of accepted waste. This analysis will focus on waste reported as "municipal solid waste". To develop a price elasticity of waste, the criterion will be *the level of municipal solid waste per capita, from 2017-2021* and the predictor will be *the level of real per capita income.*

To determine if waste is a normal good, I will be analyzing how total waste production varies across regions. 190 of the 199 active reported landfills in Texas will be observed, compiled into 16 distinct regions. Waste per capita will be calculated by dividing a region's annual aggregate solid municipal waste by the region's aggregate population to model the individual consumer's annual contribution to their local landfill.

This calculation operates under the assumption that all waste produced in the region is by the local populations; contributions from outside populations are considered constant on a regional basis and should not affect correlation calculations. Though a portion of reported municipal solid waste is provided by firms, it is assumed that these contributions accurately reflect individuals' total consumption.

I anticipate a positive correlation between my criterion, waste per capita, and my predictor variable, income per capita, up until a certain income per capita. Municipal solid waste, herein regarded as simply 'waste', is a normal good until income begins to pay for services over physical goods. Regardless of how much a consumer's income increases, there is a maximum amount of physical, waste-producing goods that can be consumed on an annual basis. It is also expected that as income increases, consumers will spend a larger percentage of their income on services or more expensive, lower waste-producing goods.

Section B: Sub-Group Identification

The data will be amalgamated into five, annual observations for each of the sixteen regions outlined by the TCEQ. These observations will be divided into three sub-groups based on population density. Population density will represent how urban or rural a region's counties are. It will be calculated by dividing the region's total area by the annual population, with the bottom 33% labeled Low Density, the middle 33 – 66% labeled Mid Density, and top third labeled High Density. This is economically significant as it can be assumed that rural, compared to dense urban areas, produce different amounts and types of waste. Additionally, urban areas are more likely to utilize registered waste streams whereas some rural areas have no equivalent alternative. This means the reported average waste per capita can be expected to be lower than the actual waste per capita in less dense populations. Rural areas should potentially show more variance in waste per capita due to the unregulated nature of waste collection outside of dense urban cities.

Correlation analysis will be based on log-levels of the criterion and predictor variables to create an income elasticity of waste. I anticipate the level of income per capita will have a larger variance across sub-groups than the level of waste per capita, with high density populations presenting the largest variance. Therefore, the correlations will be stronger, and thus more elastic, the higher the population density.

Section C: Formal Hypothesis Statement

Waste is a normal good if it has a positive correlation with income per capita. I expect sub-group 3, High Density, to have the strongest correlation between the level of the waste per capita and the level of income per capita.

Part 2: Literature Review

Previous Analysis of Waste per Capita

Waste management has been largely studied in the sociology and psychology fields, yet effective models of waste streams as a market economy are limited. A central theme in existing literature is to approach waste management to create a more circular and efficient market. In a 2022 study, Romualdas Ginevičius proposed efficiency measures for municipal waste management systems in the European Union. Part of his experiments revealed that as a nation's economic development rose, aggregate waste production and management efficiency rose. Typically, countries that exhibit high economic development also have a larger proportion of the population in densely populated areas than countries with low levels of economic development. Ginevičius's findings infer that I expect the price elasticity of waste to be normal, though there is no evidence that waste may turn inferior at a certain development level. Inefficiencies could affect the reported waste per capita in a number of ways, so there will be a greater variance in the less densely populated areas.

Works Cited:

Ginevičius, Romualdas. "The Efficiency of Municipal Waste Management Systems in the Environmental Context in the European Union." Journal of International Studies (2071-8330) 15, no. 4 (October 2022): 63-79. http://search.ebscohost.com/login.aspx?direct=true&db=eohAN=EP161426864&site=ehost-live&scope=site.

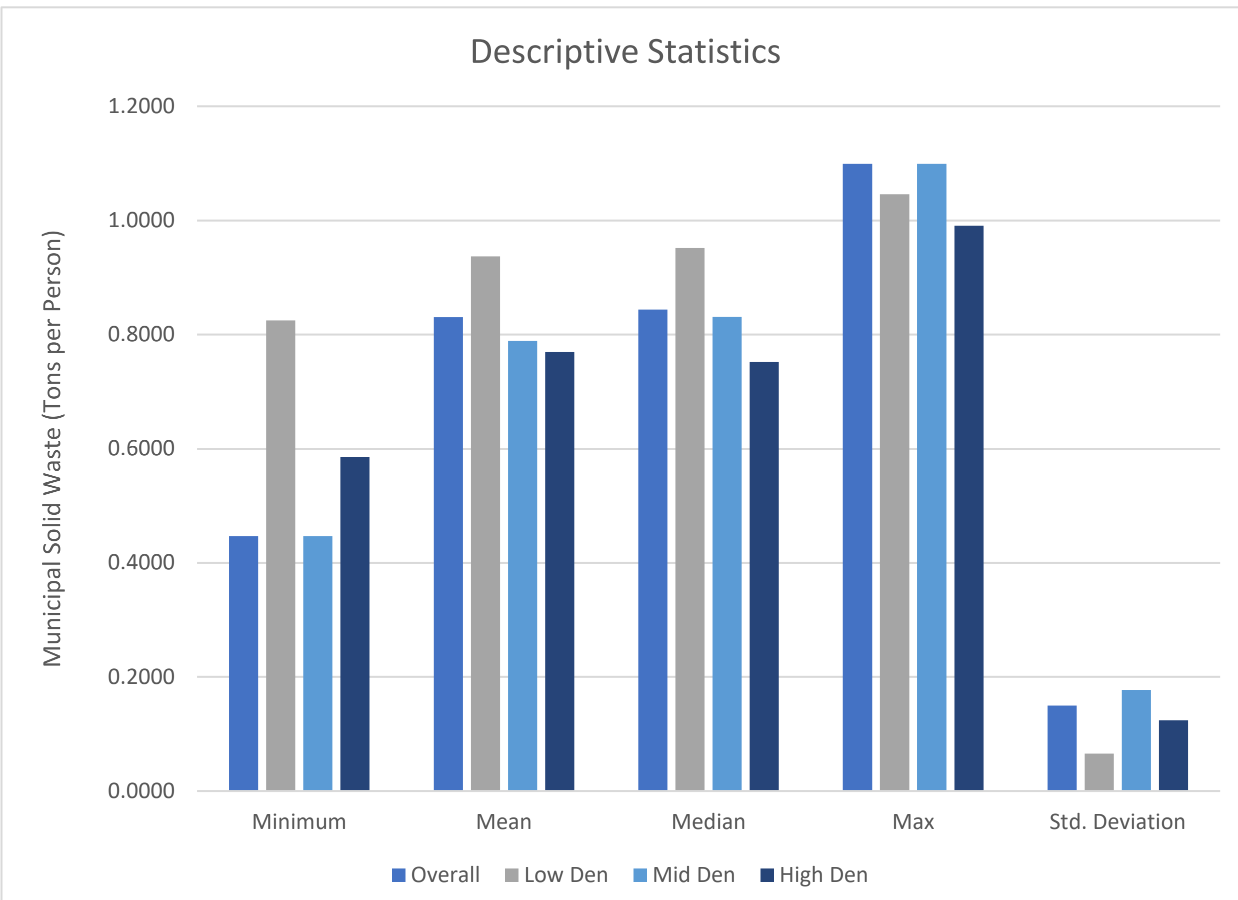
Part 3: Descriptive/Graphical Analysis

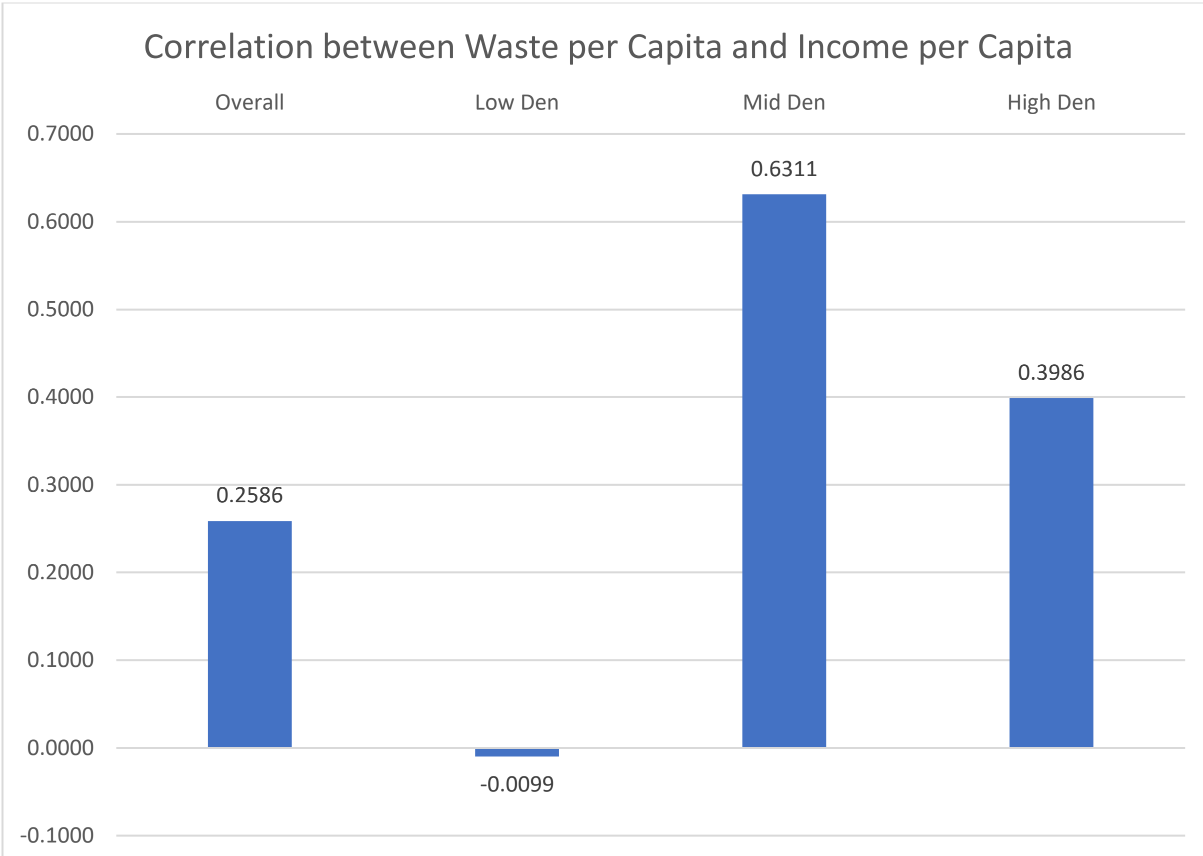
<u>Section A: Descriptive Statistics</u>

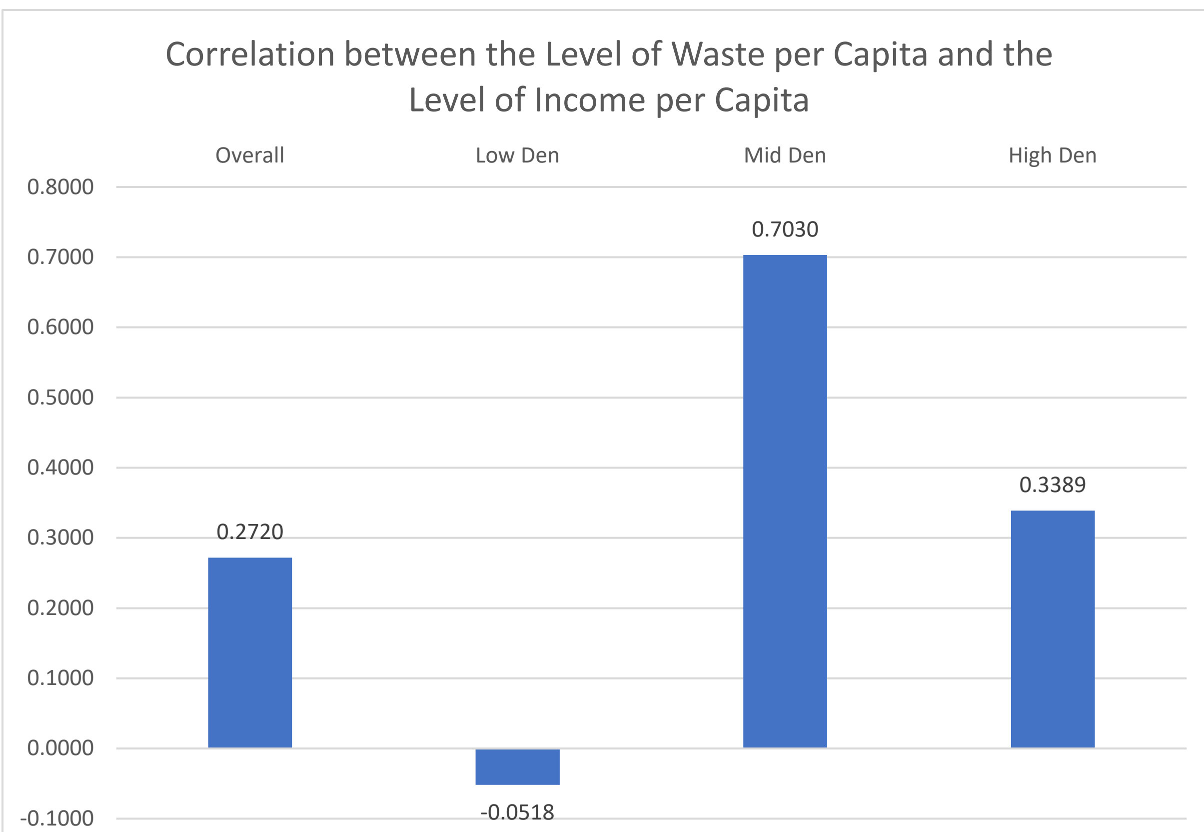
The overall correlation and overall correlation of levels are low at 0.2586 and 0.2720, respectively, showing that waste is inelastic compared to income. However, there is a large variation across sub-groups. Low density areas have the largest average waste per capita but the lowest variance; they also have a negative, near-zero correlation between waste per capita and income per capita. Alternatively, the high density areas do not have the strongest correlation as I initially hypothesized. High density populations are shown to be more inelastic than the low density areas, while mid density areas are the most elastic.

Descriptive Statistics for Waste per Capita

Group	Sample Size	Minimum	Mean	Median	Max	Std. Deviation	Correlation	Level Correlation
Overall	80	0.4464	0.8303	0.8437	1.0992	0.1496	0.2586	0.2720
Low Density	26	0.8246	0.9371	0.9515	1.0458	0.0657	-0.0099	-0.0518
Mid Density	27	0.4464	0.7889	0.8307	1.0992	0.1773	0.6311	0.7030
High Density	27	0.5857	0.7689	0.7519	0.9909	0.1238	0.3986	0.3389



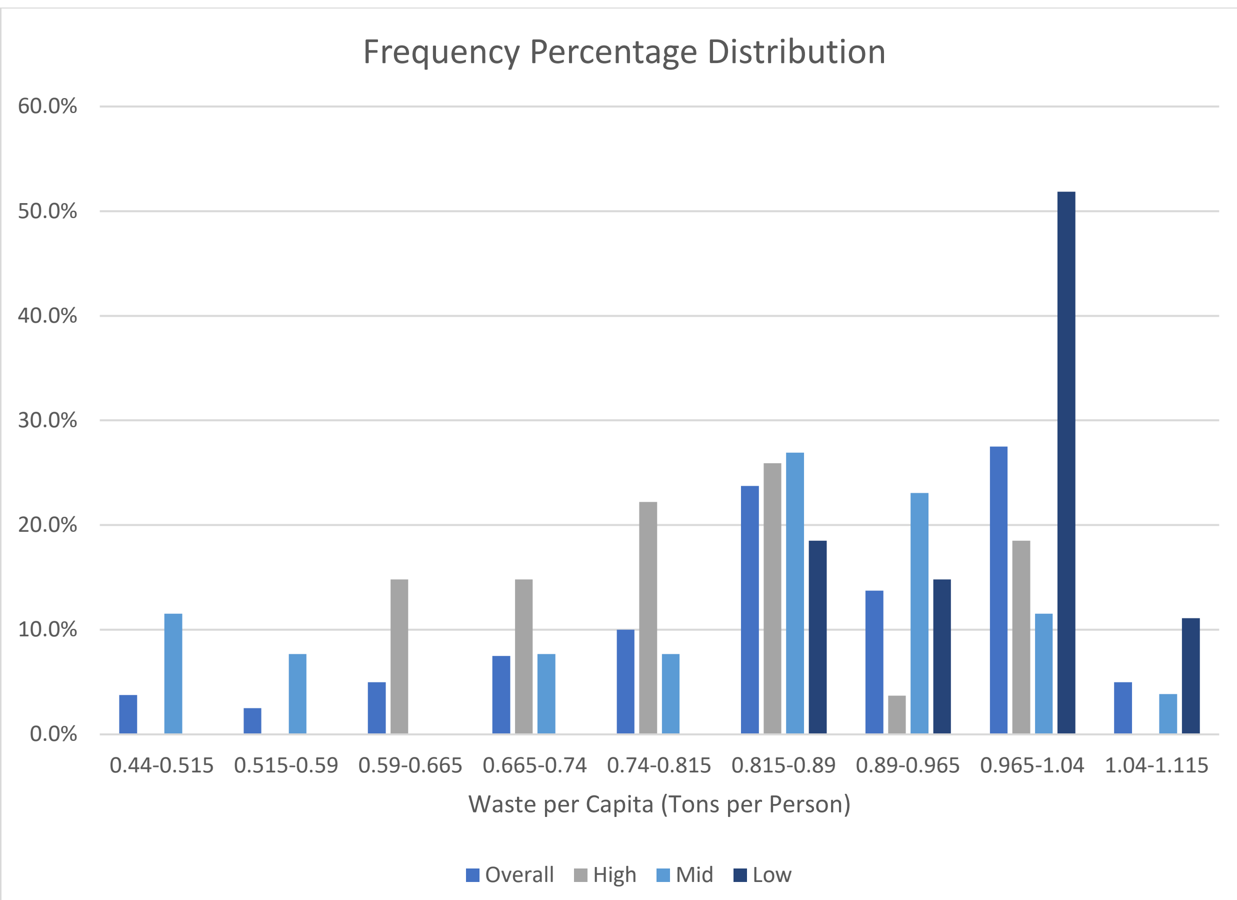

Correlation between Waste per Capita and Income per Capita
Overall
Low Den
Mid Den
High Den
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
-0.1000
0.2586
-0.0099
0.6311
0.3986


Correlation between the Level of Waste per Capita and the Level of Income per Capita
Overall
Low Den
Mid Den
High Den
0.8000
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
-0.1000
0.2720
-0.0518
0.7030
0.3389

Section B: Frequency Distributions

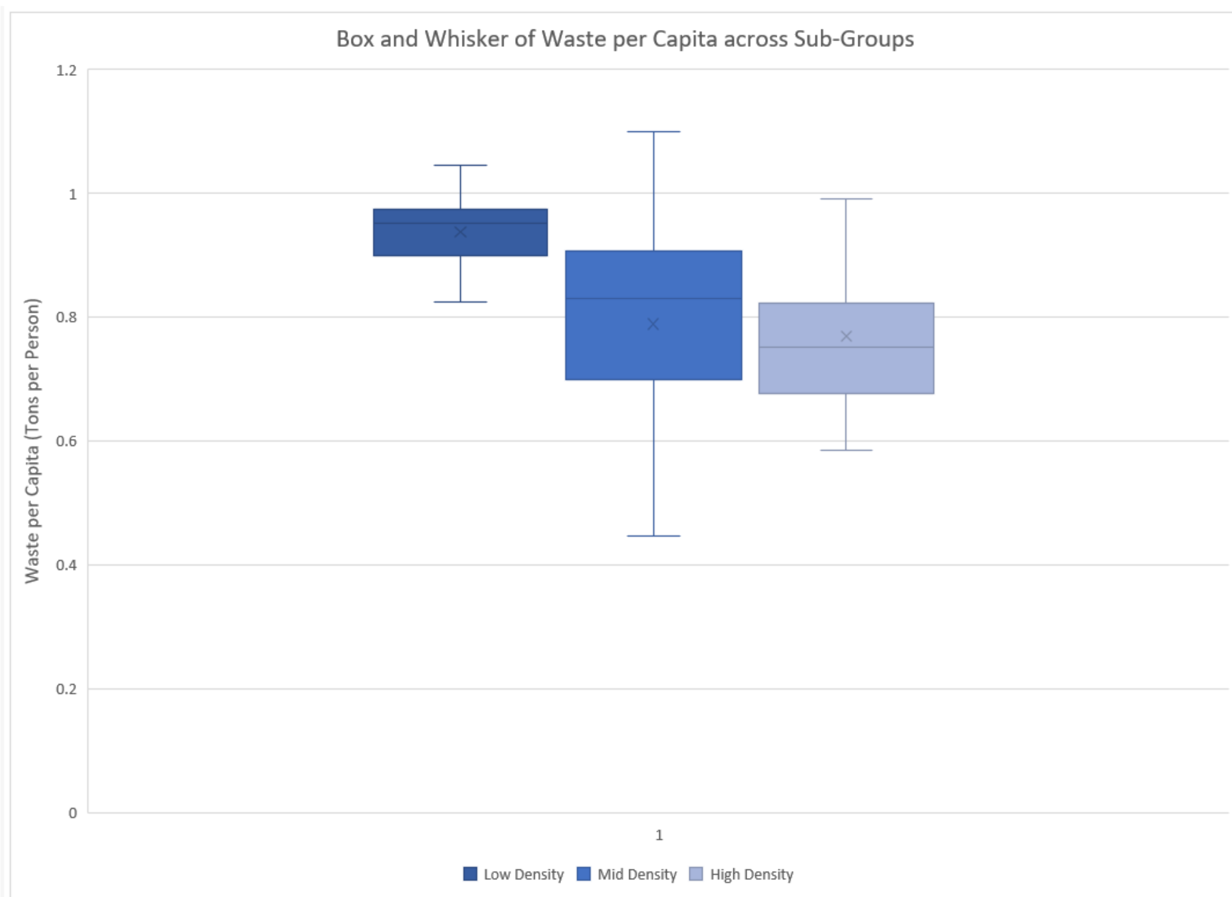
The discrepancies amongst sub-groups can be seen in the frequency distribution. Observations for low density regions are more likely to produce more waste per capita, displaying the lowest variance among sub-groups, opposite of my original assumption that rural areas would have the most variance. This means that low density areas across the state will likely produce similar waste per capita year after year.

Similarly, mid and high-density areas have similar but not exact frequency distributions at varying levels. Both sub-groups have the most observations at 0.815-0.89 waste per capita, but very different correlations. This could likely be due to the fact that there is a wider variance of income levels in urban areas than mid-density, suburban areas.



Section C: Box and Whisker Plots

Box and whisker plots show key differences in the mean, median, and variance of waste per capita. The low and mid-density areas are skewed left, where the high-density and overall groups present close to normal distributions. It also confirms that the highest variance is in the mid-density group, influencing the overall sample range, and lowest variance is in the low density sub-group. This illustrates that my assumption that low density areas would present the most variance as incorrect. Access to non-regulated alternatives may actually make the amount reported by registered landfills more consistent across low density regions.



Part 4: Single Sample Confidence Intervals and Hypothesis Tests

Section A: Confidence Intervals of Sample Means and Sample Variances

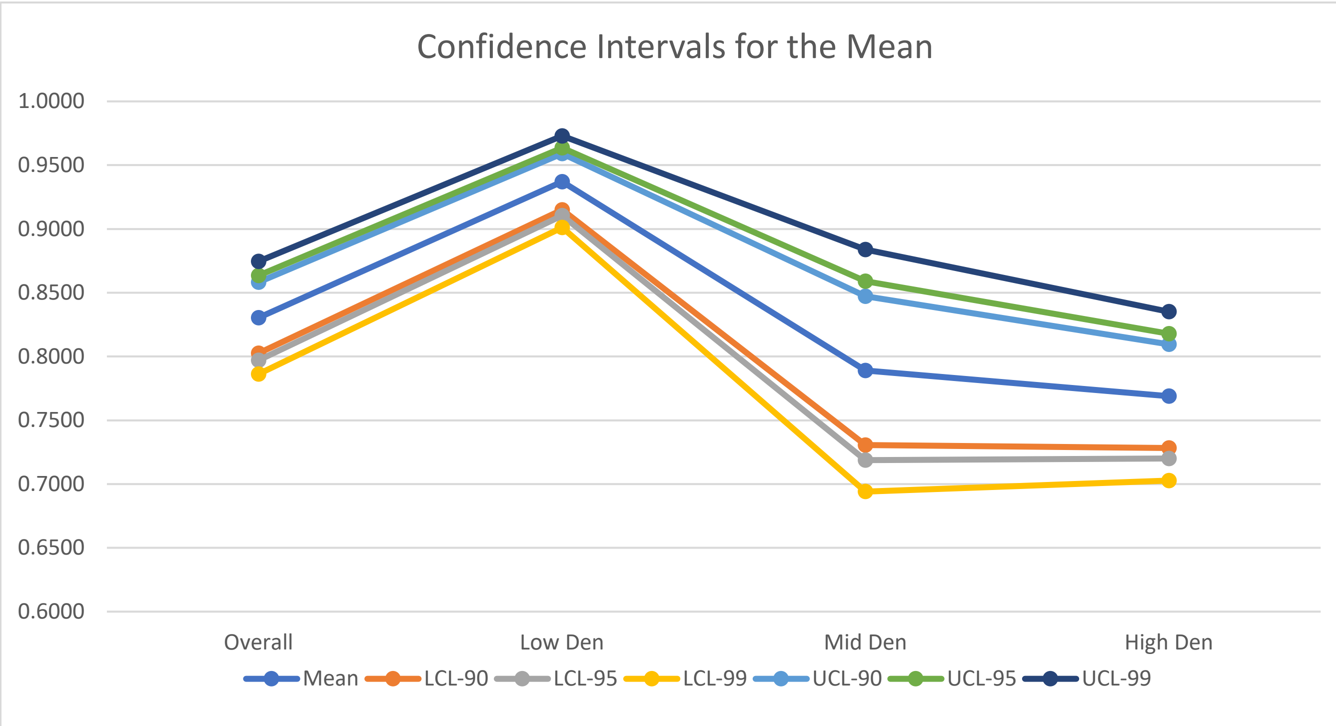
Means

For the overall and sub-group samples, confidence intervals of 90%, 95%, and 99% were calculated for the sample means of my criterion variable. For the overall sample means, 90% of observations fell between 0.8025 – 0.8582 waste per capita; 95% fall between 0.7970 – 0.8636, and 99% between 0.7862 – 0.8745.

Among my sub-groups, I expect 90% of the low density sub-group's observations to range from 0.9151 – 0.9591, 95% between 0.9150 – 0.9636, and 99% ranging from 0.9011 – 0.9730. The mid density observations are expected to range from 0.7307 – 0.8471 at 90% confidence, 0.7188 – 0.8591 at 95% confidence, and 0.6941 – 0.8838 at 99% confidence. For the high density sub-group, I expect 90% of the observations to range 0.7283 – 0.8095, 95% to range 0.7199 – 0.8179, and 99% to range from 0.7027 – 0.8351.

Confidence Intervals for the Mean

Group	Mean	LCL-90	LCL-95	LCL-99	UCL-90	UCL-95	UCL-99
Overall	0.830	0.802	0.797	0.786	0.858	0.864	0.874
Low Den	0.937	0.915	0.911	0.901	0.959	0.964	0.973
Mid Den	0.789	0.731	0.719	0.694	0.847	0.859	0.884
High Den	0.769	0.728	0.720	0.703	0.810	0.818	0.835



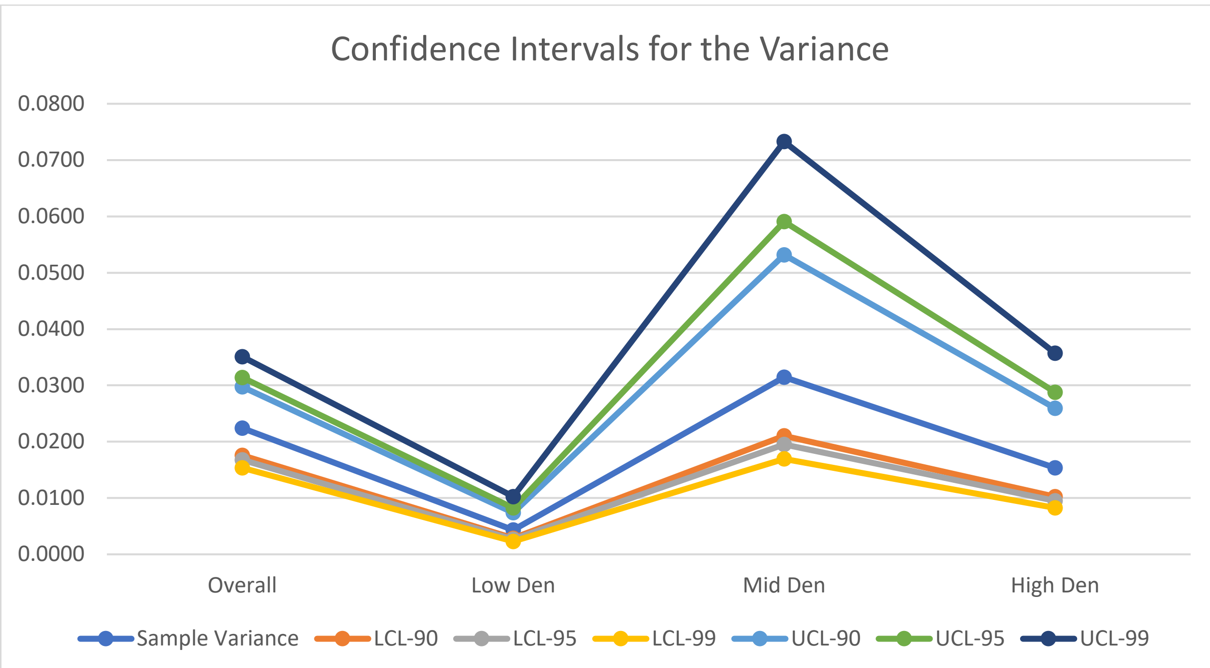
This shows a larger difference in 90% to 95% confidence for the high and mid density sub-group, whereas the low density sub-group confidence intervals don't change as much. This means the low density observations cluster closer to their mean than the other groups or the overall sample.

Variance

For the sample variances, the overall data can be expected to have 90% of the observations' variances ranging 0.0176 - 0.0297, 95% of the data 0.0168 - 0.0314, and 99% of the data 0.0154 - 0.0351. Among the sub-groups, the low density sub-group shows 90% of the variances ranging 0.0029 - 0.0074, 95% ranging 0.0027 - 0.0082, and 99% ranging 0.0023 - 0.0103**.** The mid density sub-group shows 90% of the variances ranging 0.0195 - 0.0591, 95% ranging 0.0169 - 0.0733, and 99% ranging

0.0023 - 0.0103. In the high density sub-group shows 90% of the variances ranging 0.0102 - 0.0259, 95% ranging 0.0095- 0.0288, and 99% ranging 0.0083 - 0.0357.

Confidence Intervals for the Variance							
Group	Sample Variance	LCL-90	LCL-95	LCL-99	UCL-90	UCL-95	UCL-99
Overall	0.022	0.018	0.017	0.015	0.030	0.031	0.035
Low Density	0.004	0.003	0.003	0.002	0.007	0.008	0.010
Mid Density	0.031	0.021	0.020	0.017	0.053	0.059	0.073
High Density	0.015	0.010	0.010	0.008	0.026	0.029	0.036



The confidence intervals for the sample variances show the largest variance disparity between confidence intervals is for the mid density sub-group and the lowest for the low density sub-group. This further illustrates that the low density observations are the most clustered, while the mid density observations are the most dispersed, opposite of my original assumption.

Section C: Single Sample Hypothesis Test

Means

Single sample hypothesis tests were conducted with the null hypothesis that sub-group means are equal to the overall mean. For the low density sub-group, the null hypothesis was rejected for all degrees of alpha, showing that the low density sub-group's means are not equal to the overall means. I reached the same conclusion for the high density sub-group at 90% and 95%. All other sub-group intervals fail to reject the null hypothesis that means equal to the overall. This means that in the high density sub group, 99% of the data displays means equal to the overall, but not at lower intervals. Alternatively, the mid density sub-group models the overall sample in terms of equal means at all confidence levels, indicating that the overall sample means are highly influence by the mid density sub-group as well as a few outliers in the high density sub-group.

Single Sample Hypothesis Tests for Means: Two Tailed Tests

Group	Mean	t-stat	t-Crit 90	Concl 90	t-crit 95	Concl 95	t-Crit 99	Concl 99
Low	0.937	8.284	1.708	Reject	2.060	Reject	2.787	Reject
Mid	0.789	-1.213	1.706	F.T.R.	2.056	F.T.R.	2.779	F.T.R.
High	0.769	-2.577	1.706	Reject	2.056	Reject	2.779	F.T.R.

Variance

Single sample hypothesis tests were also conducted for sample variances under the null hypothesis that variances are equal to the overall variance. For the low density sub-group, all levels of alpha reject the null hypothesis, so variances are not equal. For the mid and high density sub-groups, the null hypothesis fails to be rejected, indicating that both sub-groups have equal variance to the overall variance.

Single Sample Hypothesis Tests for Variances, Two Tailed Tests

Group	Variance	Chi-Stat	Chi-L 90	Chi-U 90	Concl 90	Chi-L 95	Chi-U 95	Concl 95	Chi-L 99	Chi-U 99	Concl 99
Low	0.004	4.821	14.611	37.652	Reject	13.120	40.646	Reject	10.520	46.928	Reject
Mid	0.031	36.528	15.379	38.885	F.T.R.	13.844	41.923	F.T.R.	11.160	48.290	F.T.R
High	0.015	17.802	15.379	38.885	F.T.R.	13.844	41.923	F.T.R.	11.160	48.290	F.T.R

Part 5: Two Sample Confidence Intervals and Hypothesis Tests

Section A: Pair-wise Hypothesis Tests of Equal Variances

Three pair-wise hypothesis tests were conducted for equal variance between the sub-groups: low-mid, low-high, and mid-high. All three tests result in a P-Value less than 0.05, rejecting the null hypothesis that the variances are equal between any two sub-groups.

Variance Test F-Stat

	Low	Mid
Mid	0.1373	
High	0.2816	2.0519

P-Values

	Low	Mid
Mid	2.073E-06	
High	1.118E-03	9.638E-01

Conclusion: Reject if P-value is greater than 0.95 or less than 0.05

	Low	Mid
Mid	Reject	
High	Reject	Reject

The pair-wise tests are used to determine which, if any, of the groups have similar patterns of sample variance. Each of the P-values are well below the calculated F-statistics, indicating that there is close to no likelihood that any sub-group would produce equal variation of their production of waste-per-capita. Each of the sub-groups has a unique pattern of variance which will influence the pair-wise hypothesis tests for differences in means.

Section B: Pair-wise Confidence Intervals and Hypothesis Tests of Equal Means

The same pairs are used to calculate the confidence intervals for the difference in means assuming unequal variances, then used to test the null hypothesis that the difference in means equals zero.

Pair-Wise Confidences Intervals for Differnces in Means
95% Confidence Intervals

		Low	Mid
Mid	UCL	0.222	
	LCL	0.074	
High	UCL	0.223	0.104
	LCL	0.113	-0.064

Pair-Wise Hypothesis Tests for Differences in Means
t-statistics

	Low	Mid
Mid	4.0607	
High	6.2080	0.4810

P-Values

	Low	Mid
Mid	2.83E-04	
High	2.41E-07	0.6328

Conclusions: Reject if P-Value is Less that 5%

	Low	Mid
Mid	Reject	
High	Reject	F.T.R.

The previous test for equal means determined that each group-pair would need to be individually compared to test for equality across means. Means are considered to be equal if the null hypothesis fails to be rejected, indicating that there is little difference in the sample means for the group-pair. Calculated P-values indicate that I reject the null for the low-mid and low-high groups. I fail to reject the null for the mid-high group. This indicates that the low density sub-group's average production of waste-per-capita does not equal the production in mid and high density locations. However, the mid and high density sub-groups display a similar mean waste per capita. This information supports at least part of my initial hypothesis, in that waste-per-capita does not change much between urban to suburban areas but both display different means than rural areas.

Part 6: ANOVA Tests

Section A: Single-factor ANOVA

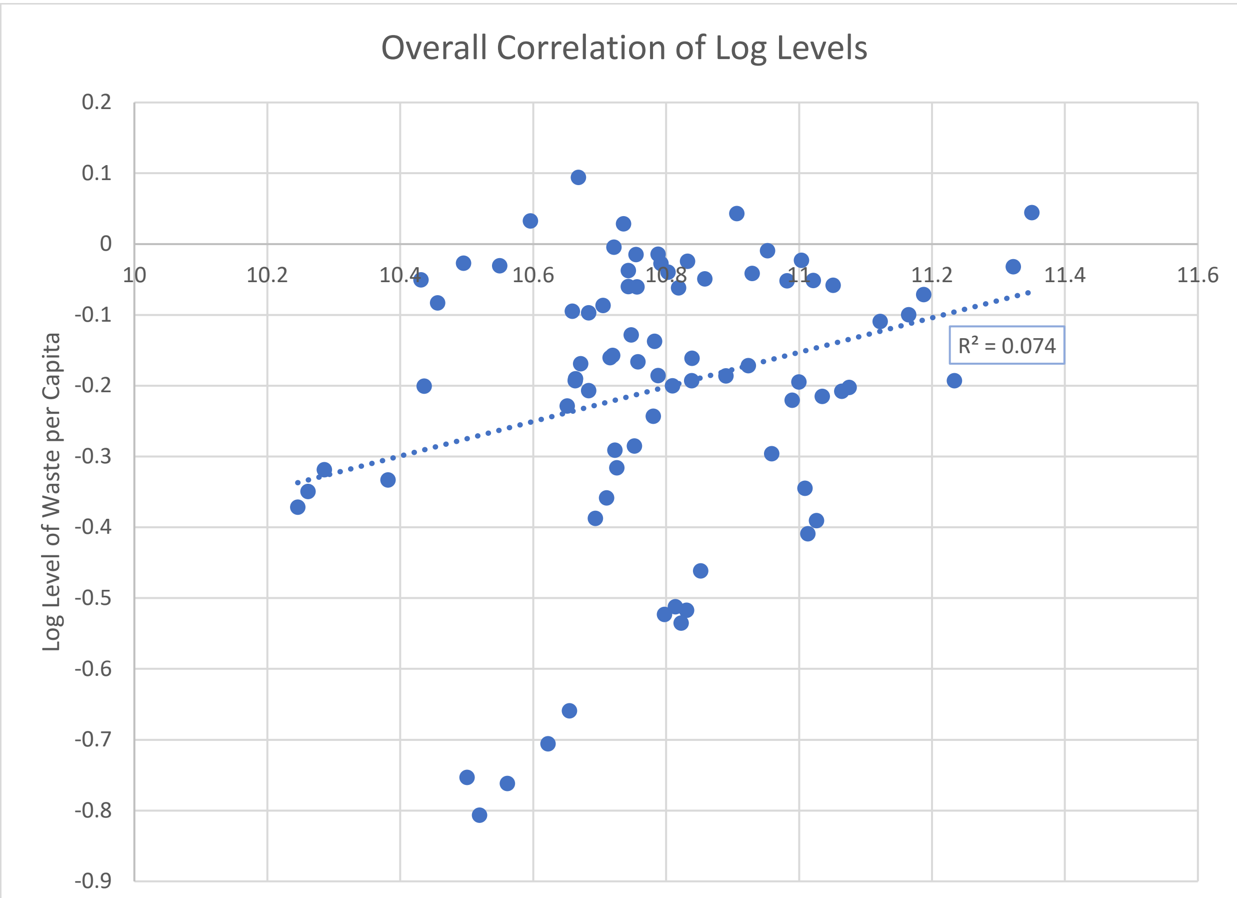
Running a single-factor ANOVA test across the three sub-groups produces a P-value of 0.0000145, far less than the cutoff point of 0.05. There is very little confidence in the null hypothesis that the means across groups are equal. This indicates that there is a significant amount of variation across sub-groups, as initially hypothesized.

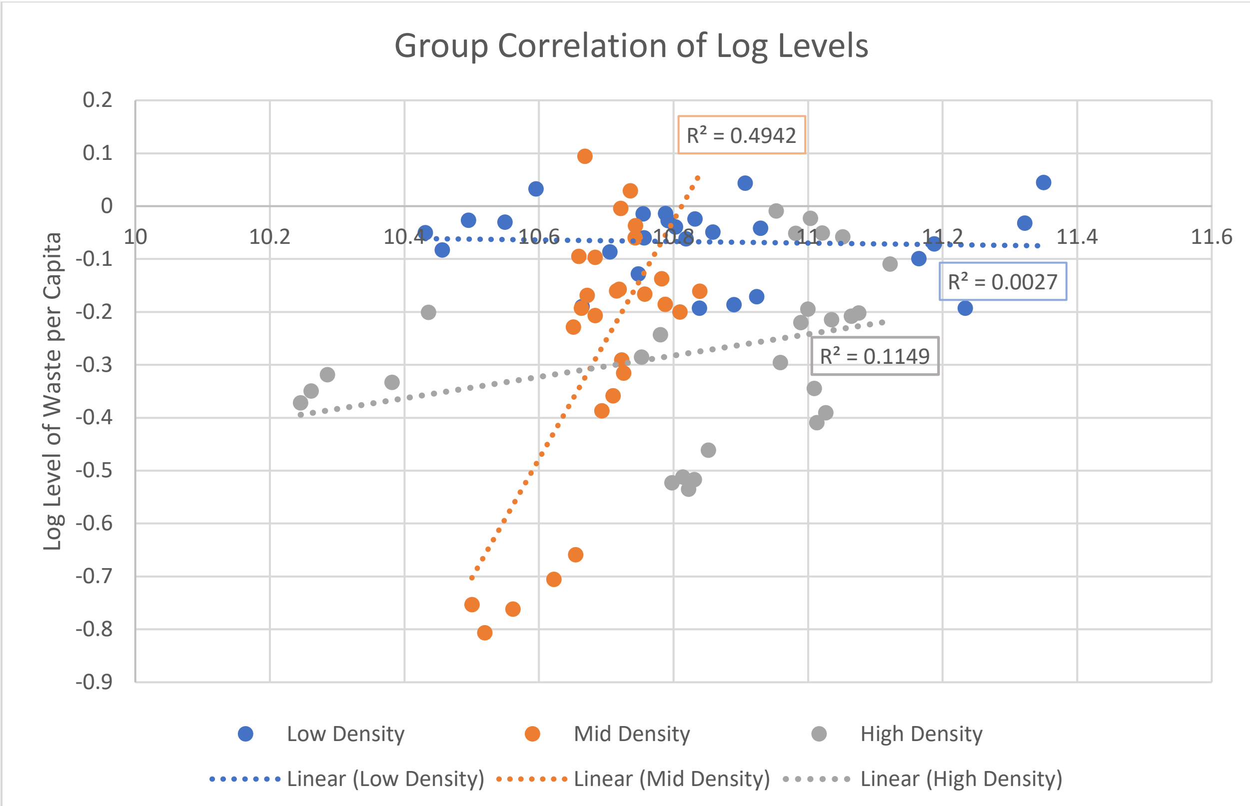
This seems contrary to the earlier test result that argued the high and mid density groups display similar means; this similarity was determined by testing for a difference amongst means. ANOVA's null hypothesis instead tests for joint equality across sub-group means. The different results could indicate that the difference in means of the low versus mid and high density subgroups outweighs the similarity of the means between the mid and high sub-groups, causing greater influence on the overall sample.

Part 7: Correlation Analysis

The criterion and predictor variables are converted to log levels for correlation analysis.

<u>Section A: Scatter Plots and Trend Lines</u>





<u>Section B: Tests of Significant Correlation</u>

For the overall and sub-group samples, the null hypothesis that there is zero correlation between the predictor and criterion variable, $H_0 : \rho_{XY} = 0$, was tested with an alpha of 0.05 against the student's t-distribution.

Group	Correlation	R-Squared	Sample Size	zr
Overall	0.272	0.074	80	
Low	-0.052	0.003	26	-0.052
Mid	0.703	0.494	27	0.873
High	0.339	0.115	27	0.353

Tests of Individual Correlation Significance H0: p=0
Two-Tailed Tests,
alpha = 0.05

Group	t-calc	t-Critical	Conclusion
Overall	2.496	1.991	Reject
Low	-0.254	2.064	F.T.R
Mid	4.942	2.060	Reject
High	1.801	2.060	F.T.R

For the low and high sub-groups, I fail to reject the null hypothesis, indicating that there is not a significant level of correlation between the two variables. On the other hand, the overall and mid sub-groups reject the null, showing there is a level of correlation for them. This shows that the correlation between the level of waste per capita and the level of income per capita in the mid density sub-group could be strong enough that it offsets the low correlations presented in the low and mid subgroups. This is the opposite of my original theory, that the high density sub-group would display the largest correlation between the two variables.

The mid density sub-group's strong correlation is likey due to the larger variance in the log of waste per capita; though the high density sub-group has a similar spread, the broad variance of the log of the income per capita weakens the overall correlation. This confirms and denies parts of my original theory as to whether the criterion or predictor would influence elasticity the most. In the mid sub-group, there is less difference in income but a large difference in waste production, showing that waste is highly elastic. In the high and low sub-groups, there is a larger difference in income than waste production, showing that waste is inelastic. The difference in waste production in high density sub-group is slightly greater than the low density sub-group, making the correlations *slightly* stronger.

Another of my original assumptions is also disproved, that all sub-groups will have a positive correlation between the criterion and predictor variable. The low density group has a correlation that is negative; though it is near zero, there is still a -5% correlation which could offer insight into a potential tipping point of waste from an inferior to normal good.

Section C: Pair-wise Correlations Tests

The previously derived pairs, low-mid, low-high, and mid-high, are used to test the null hypothesis that the correlation for each group is equal.

Pair-Wise Correlation Test H0: pi = pj			
Z-calc			
	Low	Mid	
Mid	-3.170		
High	-1.387	1.802	
Two-Tailed test, alpha			0.05
Z-critical	1.96		
Conclusions			
	Low	Mid	
Mid	Reject		
High	F.T.R	F.T.R	

In comparison to the previous tests, part of the results seems contradictory. First, I reject the idea that the low and mid sub-groups have similar correlations, and fail to reject that the low and high sub-groups are different. However, calculations also show that I should fail to reject that the mid and high sub-groups have similar correlations. This is likely in part due to the drastic variances of the criterion and predictor across sub-groups. The low and high sub-group have similar distribution of income but not waste, while the mid sub-group is more similar in waste than income to the high sub-group.

Section D: Joint Equality Test

Additionally, the sub-groups were used to calculate a test of joint equality of correlations. The Chi-Sq Calculation was far larger than the Chi-Sq Critical value, leading to rejecting the null hypothesis that the correlations are equal across sub-groups.

Joint Multi-Group Correlation Test, alpha =				0.05		
Group	Sample Size	Correlation	zr trans.	nj-3	(nj-3)*zr^2	(nj-3)*zr
Low	26	-0.052	-0.052	23	0.062	-0.003
Mid	27	0.703	0.873	24	18.297	15.976
High	27	0.339	0.353	24	2.988	1.054
Sums				71	21.347	17.027
						289.920
	Chi-Sq Clac	17.26				
	Chi-Sq Crit	5.99				
	Conclusion	Reject				

This supports my assumptions following the pair-wise correlation test, that the difference in correlation of the low and mid sub-groups is great enough of offset the overall sample. This further disproves my original hypothesis that the correlation will be the largest in the high density sub-group and that there will be a strong correlation in all sub-groups. The large difference in correlations ultimately leads to the rejection of joint equality.

Part 8: Conclusions, Discussions, and Limitations

Before final conclusions can be drawn, the efficacy of the data must be addressed. Raw data was manually collected from various sources, and complied into a data set for the express purpose of this project. Human error will always influence analytics results but does not have near the influence as the largest limitation of the data - it does not accurately reflect the variable it aims to represent. Upon initial collection, the facility waste data was to be sorted into unique regions depending on areas of service overlap. These regions would accurately reflect waste per capita, as the counties served generally have a similar population density.

However, the scope of the project restricted time necessary to develop a complex sorting algorithm. Ultimately, the data was divided into the TCEQ's existing 16 regions, with five years of data for each region, creating 80 overall observations. The data does not accurately reflect waste per capita based on population density, as some regions contained a mix of low, mid, and high density populations. Additionally, this number of observations barely meets the threshold to apply the central limit theorem, so even if the data was effective, conclusions drawn on these samples may not reflect the actual metrics of the population data.

Despite these drawbacks, the data does pose room for interesting discussion. For the overall sample, there is very little correlation between waste production and income levels, showing that waste is inelastic to income. The variations between sub-groups paint a very different story.

All of my original hypotheses ultimately fell through. The only group to exhibit a strong positive correlation is the mid density sub-group, which also had the largest variation in waste per capita. Though the high density sub-group displayed high variance in the income per capita, the low density sub-group did as well, while the mid density sub-group did not. These unique differences in the mid density sub-group lead to it having the strongest correlation between waste production and income. These analytics illustrate that waste is a normal, highly elastic good in mid density populations, a normal but more inelastic good in high density populations, and has no correlation, but borders on potentially being an inferior good, in low density populations.

The difference in correlations, particularly for the mid density group, are likely due to the flawed grouping of regions. The regions grouped as mid density may have counties that are all truly suburban type density, but it should be considered that these regions may be a combination of low density rural counties and high density urban counties, evening out to a mid-level population density. Arguably, the mid density sub-group may more accurately reflect the income elasticity of waste for the population than my overall sample because of this.

Though the data and grouping may not provide an accurate model of the income elasticity of waste, the basis of this project could be expanded and integrated into an algorithm to create a model that updates with each new report. As previously mentioned, the data set was compiled by manually entering observations from individual facility reports across multiple years. With direct database access, data could be collected over a wider range of years and worked through a sorting algorithm to create aggregate waste streams that accurately reflect their relative population density.

Though the results from analysis are unreliable due to the grouping method, they still provide valuable insight into the variation of income elasticity of waste over the state of Texas. Distribution analysis could easily be replicated with more effective data to truly answer the question of whether waste is a normal good.