Final Project in the Statistical Learning Course

Topic： Are There Statistically Significant Differences in the Amount of Plastic Waste Per Capita Between Countries, Based on Their Income Level ？

המרכז האקדמי רופין

Niv Alex

Yuval Geyari

June 12, 2025

Abstract

The plastic waste crisis represents a global environmental and social challenge with severe implications for ecosystems, public health, and sustainable economies. This study examines the relationship between national income levels and the amount of mismanaged plastic waste generated per capita. To this end, two data sources were integrated: country income classifications from the World Bank and per capita mismanaged plastic waste data from Our World in Data. The datasets were merged into a unified dataset comprising over 150 countries, categorized into four income groups: low, lower-middle, upper-middle, and high income.

Initial normality testing using the Shapiro–Wilk test indicated that the data were not normally distributed. Accordingly, the Kruskal–Wallis test was employed to conduct a non-parametric comparison across the four groups, revealing statistically significant differences in plastic waste generation. Subsequently, a post hoc analysis using the Mann–Whitney U test identified which income groups differed significantly from one another. The results indicated that high-income countries generate significantly less mismanaged plastic waste per capita compared to all three other income groups.

These findings highlight the critical link between economic prosperity and environmentally responsible behavior. They suggest that the international community should prioritize support for low-income countries by investing in infrastructure, environmental technologies, and public awareness to mitigate plastic pollution. The study also underscores the disparity between global environmental responsibility and the technical and economic capacities of different countries to address the plastic waste problem, particularly in developing nations. In light of this, the promotion of international cooperation is recommended to strengthen the ability of low-income and lower-middle-income countries to address this pressing global issue.

תוכן העניינים

1 Introduction

1.1 Background: The Global Plastic Problem

Since the invention of synthetic plastic in the early 20th century, this material has become an integral part of nearly every sector of modern industry and daily life. Today, plastic is found in food packaging, medical equipment, electronic components, synthetic clothing, vehicles, and more. Its key properties - low production cost, flexibility, light weight, and durability - have made plastic the dominant choice over many natural materials.

According to an OECD report from 2022, global plastic production reached approximately 460 million tons in 2019 and is expected to double by 2060 if no substantial changes are made to plastic recycling policies.
Of all the plastic ever produced, around 79% has accumulated as non-degradable waste, 12% has been incinerated, and only about 9% has been effectively recycled.

One of the central issues with plastic is its extreme durability: most types of plastic do not decompose for hundreds of years. Plastic waste spreads across the air, land, and oceans, breaking down into microscopic particles (microplastics) that can even enter the human body through food and drink. In addition to affecting humans, plastic waste causes severe harm to marine life. According to United Nations data, over one million seabirds and approximately $100,000$ marine mammals die each year as a result of ingesting plastic or becoming entangled in it.
Entire marine ecosystems are being damaged by plastic bags, fishing nets, and plastic bottles that either sink to the ocean floor or drift across vast distances—forming massive "plastic islands" such as the Great Pacific Garbage Patch, which is estimated to be twice the size of the state of Texas.

The situation in Israel is also alarming. According to a 2022 report by the Ministry of Environmental Protection, Israel generates approximately 1.5 million tons of plastic waste annually, of which about 70% consists of packaging waste. Israel ranks among the highest OECD countries in per capita plastic consumption.

1.2 Research Objective and Research Question

sThe objective of our study is to examine whether there are statistically significant differences in the amount of plastic waste generated per capita among countries, based on their national income level as classified by the World Bank (2023).

The research focuses on a comparative analysis across four income categories: low, lower-middle, upper-middle, and high income, using data from open global sources.

Our central research question is: Does a country's income level constitute a statistically significant factor influencing the amount of plastic waste generated per capita, and what are the differences between income groups in this context?

The overarching goal of this final project is to contribute to a better understanding of the relationships between economic status, environmental consumption, and waste management infrastructure.
This is achieved through the use of advanced statistical methods suitable for non-normally distributed data, as taught in the course throughout the semester.

1.3 Previous Research on the Topic

In recent years, several key studies have examined the relationships between plastic waste generation, national income levels, and the associated environmental impacts.

One of the most influential studies in this field is by Jambeck et al. (2015), published in Science, which analyzed sources of land-based plastic waste entering the oceans across 192 countries. The findings revealed that while low and middle-income countries tend to produce less plastic waste per capita, they contribute disproportionately to marine pollution due to the lack of adequate waste management infrastructure.

Another significant report by the World Bank (2018), conducted as part of the What a Waste initiative, presented data on waste generation by geographic region and income level. It showed that high-income countries produce over $2kg$ of waste per person per day, compared to only $0.6kg$ in low-income countries. However, recycling in wealthier countries tends to be more efficient and widespread.

The OECD's 2022 Global Plastics Outlook report concluded that although OECD countries account for less than 20% of the world's population, they are responsible for nearly 50% of global plastic consumption. Nonetheless, due to advanced regulations and well-developed infrastructure, recycling rates in these countries are significantly higher than the global average.

These findings support the hypothesis that the relationship between income level and environmental harm caused by plastic is complex and influenced by additional factors such as infrastructure, regulatory frameworks, and consumption culture.

2 Data Description and Sources

To conduct this research, we utilized two primary datasets. These were merged based on the English name of each country in order to create a unified dataset that enables comparative analysis by income level and per capita plastic waste.

2.1 Description of the First Dataset : world-bank-income-groups.csv

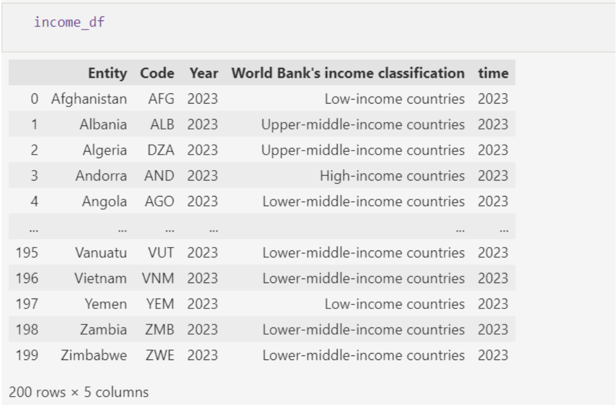
The first dataset, titled world-bank-income-groups.csv, is based on the World Bank's 2023 income classification, which categorizes countries into four income groups:

- Low income
- Lower-middle income
- Upper-middle income
- High income

This classification is based on **Gross National Income (GNI) per capita**, and it serves as a standard metric for global comparisons in both economic and environmental indicators.
The dataset includes approximately 200 rows, with each row representing a single country. The main columns are:

- Entity : The name of the country
- Code : A unique country code
- World Bank income classification : The income group assigned to that country

איור 1: world-bank-income-groups.csv



```
income_df
```

	Entity	Code	Year	World Bank's income classification	time
0	Afghanistan	AFG	2023	Low-income countries	2023
1	Albania	ALB	2023	Upper-middle-income countries	2023
2	Algeria	DZA	2023	Upper-middle-income countries	2023
3	Andorra	AND	2023	High-income countries	2023
4	Angola	AGO	2023	Lower-middle-income countries	2023
...	...	...	...	...	...
195	Vanuatu	VUT	2023	Lower-middle-income countries	2023
196	Vietnam	VNM	2023	Lower-middle-income countries	2023
197	Yemen	YEM	2023	Low-income countries	2023
198	Zambia	ZMB	2023	Lower-middle-income countries	2023
199	Zimbabwe	ZWE	2023	Lower-middle-income countries	2023

200 rows × 5 columns

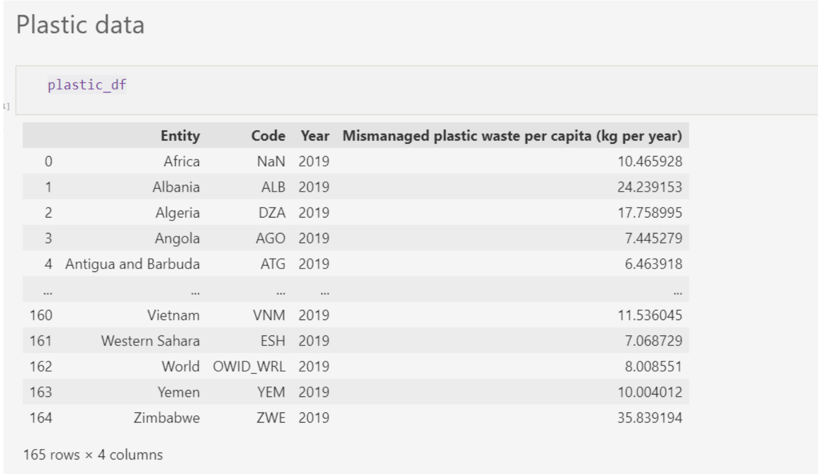
2.2 Description of the Second Dataset : mismanaged-plastic-waste-per-capita.csv

The second dataset, mismanaged-plastic-waste-per-capita.csv, is sourced from the environmental data repository of the Our World in Data platform.
This dataset provides a quantitative measure of mismanaged plastic waste per capita, representing the amount of plastic waste generated per person per year that is not properly managed.

Mismanaged plastic waste refers to plastic waste that is not recycled, not incinerated under controlled conditions, and not disposed of in sanitary landfills. This definition also includes materials that are openly burned, dumped into water bodies, or handled at non-hygienic and non-regulated waste disposal sites.

This metric serves as a significant indicator of the environmental risk posed by plastic waste in a given country.
The dataset contains 165 rows, each representing a country. (A slight filtering step is necessary, as the dataset also includes rows for continents- such as Europe, Africa, and Asia—as well as a global average listed in row 162.)
For each country, the dataset includes a column titled Mismanaged plastic waste per capita, which indicates the average annual amount of mismanaged plastic waste (in kilograms) produced per person in that country.

Plastic-waste-per-capita.csv : 2 Figure



Plastic data

```
plastic_df
```

	Entity	Code	Year	Mismanaged plastic waste per capita (kg per year)
0	Africa	NaN	2019	10.465928
1	Albania	ALB	2019	24.239153
2	Algeria	DZA	2019	17.758995
3	Angola	AGO	2019	7.445279
4	Antigua and Barbuda	ATG	2019	6.463918
...	...	...	...	...
160	Vietnam	VNM	2019	11.536045
161	Western Sahara	ESH	2019	7.068729
162	World	OWID_WRL	2019	8.008551
163	Yemen	YEM	2019	10.004012
164	Zimbabwe	ZWE	2019	35.839194

165 rows × 4 columns

2.3 Description of the Final Dataset: final_df

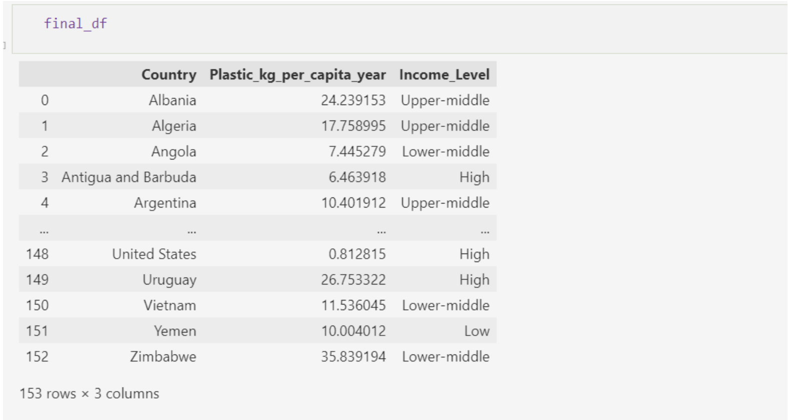
During the data processing stage, the plastic waste data were filtered to include a specific year (e.g., 2019) to ensure consistency.
Subsequently, the two original datasets were merged based on the Entity column (country name). After merging, only countries with complete data for both income level and plastic waste per capita were retained for analysis.

The final dataset includes the following columns:

- **Country** – The name of the country
- **Income_Level** – The income classification (Low, Lower-middle, Upper-middle, High)
- **Plastic_keg_per_capita_year** – The amount of mismanaged plastic waste per capita per year, measured in kilograms

The resulting dataset, named final_df, contains information on **153 countries**, and serves as the foundation for the statistical analysis conducted in this study.

איור 3 : final_df



```
final_df
```

	Country	Plastic_kg_per_capita_year	Income_Level
0	Albania	24.239153	Upper-middle
1	Algeria	17.758995	Upper-middle
2	Angola	7.445279	Lower-middle
3	Antigua and Barbuda	6.463918	High
4	Argentina	10.401912	Upper-middle
...	...	...	...
148	United States	0.812815	High
149	Uruguay	26.753322	High
150	Vietnam	11.536045	Lower-middle
151	Yemen	10.004012	Low
152	Zimbabwe	35.839194	Lower-middle

153 rows × 3 columns

3 Statistical and Computational Methods

The statistical analysis conducted in this study aimed to determine whether there are statistically significant differences between groups of countries classified by income level - with respect to the amount of plastic waste generated per capita.
Since the data are not normally distributed (as shown in the next section), non-parametric methods were applied.

3.1 Normality Testing Using the Shapiro–Wilk Test and Visual Tools

- Shapiro–Wilk Test – This is a statistical test used to assess whether a dataset follows a normal distribution.

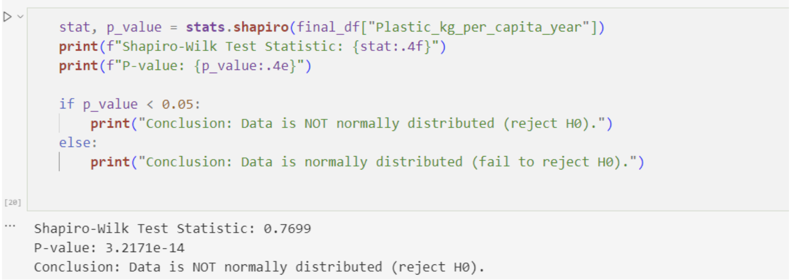
$$H_0 : data\ has\ a\ normal\ distribution$$

$$H_1 : data\ has\ not\ normal\ distribution$$

A p-value less than 0.05 indicates rejection of the null hypothesis H_0 , suggesting that the data do not follow a normal distribution.

The following results were obtained:

Shapiro–Wilk Test :4 Figure



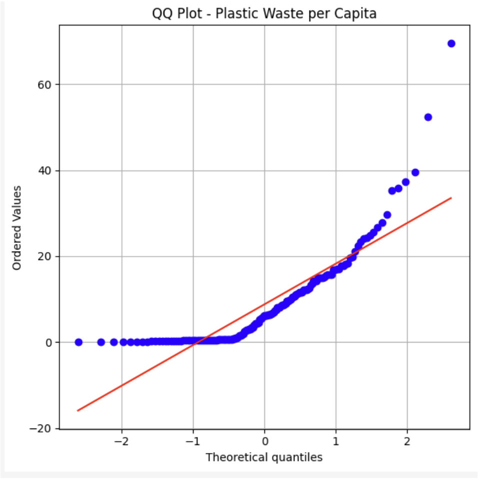
```
stat, p_value = stats.shapiro(final_df["Plastic_kg_per_capita_year"])
print(f"Shapiro-Wilk Test Statistic: {stat:.4f}")
print(f"P-value: {p_value:.4e}")

if p_value < 0.05:
    print("Conclusion: Data is NOT normally distributed (reject H0).")
else:
    print("Conclusion: Data is normally distributed (fail to reject H0).")
```

```
Shapiro-Wilk Test Statistic: 0.7699
P-value: 3.2171e-14
Conclusion: Data is NOT normally distributed (reject H0).
```

As shown, the result was p_{value} = $3.217 \cdot e^{-14}$, which approaches zero. Therefore, we reject H_0 and conclude that the data do not follow a normal distribution.

- Visual Inspection – Using Python, we generated histograms, QQ plots, and boxplots to gain an initial impression of the data distribution.
  The QQ plot compares the distribution of our data to a perfectly normal distribution (represented by the red line).
  As shown, the left tail (lower values) falls below the red line, while the right tail (higher values) rises sharply above it.

QQ-Plot : 5 Figure



Therefore, we conclude that the data do not follow a normal distribution.

3.2 Kruskal–Wallis Test

Since the assumption of normality was not met, we applied the non-parametric alternative to ANOVA, known as the Kruskal–Wallis test, which is used to compare k independent (unpaired) populations.
Given k groups of sizes $n_1, n_2 \ldots, n_k$ the total number of observations is:

$$N = \sum_{i=1}^{k} n_i$$

After sorting all the data in ascending order and assigning ranks, the Kruskal–Wallis test statistic Q is calculated as:

$$Q = \left[\frac{12}{N(N+1)} \cdot \sum_{i=1}^{k} \frac{T_i^2}{s_i}\right] - 3 \cdot (N+1)$$

Where:
N = total number of observations
K = number of groups
T_i= sum of ranks for group i
s_i= size of group i

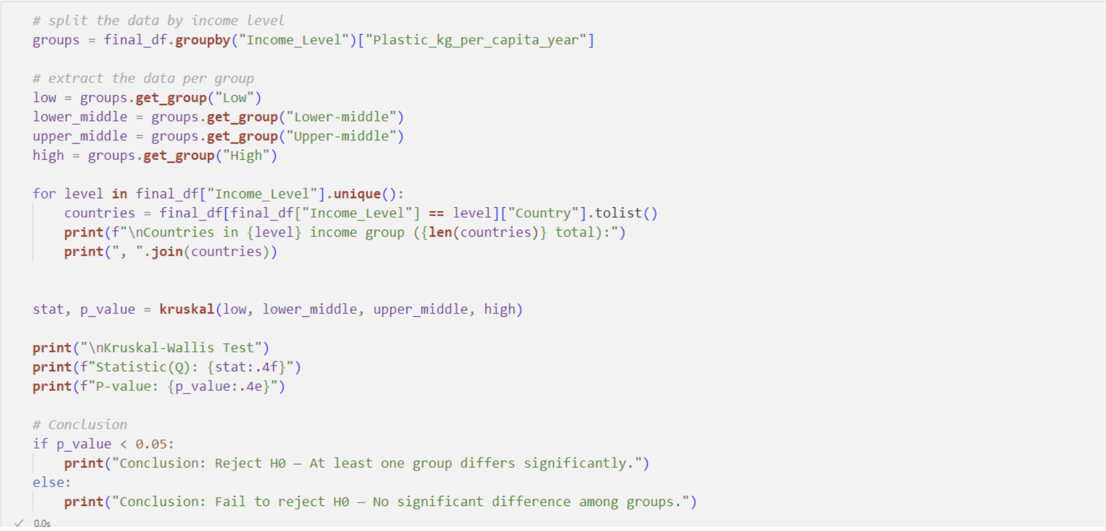
Hypotheses:

$$H_0 : \text{ no diffrence between groups}$$

$$H_1 : \text{at least one group is diffrent}$$

The test statistic Q approximately follows a chi-squared distribution with $df = k - 1$ degrees of freedom.
In our case, we applied the Kruskal–Wallis test to four independent groups, representing the income levels: low, lower-middle, upper-middle, and high.

Kruskal–Wallis in python : 6 Figure



```
# split the data by income level
groups = final_df.groupby("Income_Level")["Plastic_kg_per_capita_year"]

# extract the data per group
low = groups.get_group("Low")
lower_middle = groups.get_group("Lower-middle")
upper_middle = groups.get_group("Upper-middle")
high = groups.get_group("High")

for level in final_df["Income_Level"].unique():
    countries = final_df[final_df["Income_Level"] == level]["Country"].tolist()
    print(f"\nCountries in {level} income group ({len(countries)} total):")
    print(", ".join(countries))


stat, p_value = kruskal(low, lower_middle, upper_middle, high)

print("\nKruskal-Wallis Test")
print(f"Statistic(Q): {stat:.4f}")
print(f"P-value: {p_value:.4e}")

# Conclusion
if p_value < 0.05:
    print("Conclusion: Reject H0 – At least one group differs significantly.")
else:
    print("Conclusion: Fail to reject H0 – No significant difference among groups.")
```

✓ 0.0s

The results obtained were: $Q = 52.1901$ and p_{value} = $2.7283 \cdot e^{-11}$.
Since the p-value approaches zero, we reject the null hypothesis H_0.
This means that, based on the Kruskal–Wallis test, we can conclude that there is a statistically significant difference between at least one pair of income groups (low, lower-middle, upper-middle, and high income).

3.3 Post-Hoc Analysis Using the Mann–Whitney U Test

3.3.1 Mann–Whitney U Test

After the Kruskal–Wallis test indicated statistically significant differences between income groups, a post-hoc analysis was conducted to determine which pairs of groups differ significantly.
Since the data are not normally distributed, we applied a non-parametric approach based on the **Mann–Whitney U test**, which evaluates whether there is a statistically significant difference between two independent populations.
Each Mann–Whitney U test in this analysis assessed whether two income groups differ in terms of per capita plastic waste.

However, conducting multiple comparisons on the same dataset increases the risk of Type I errors (false positives), i.e., incorrectly concluding that a difference exists when it does not.
As the number of pairwise comparisons increases, the overall probability of such an error also rises.

To address this issue, we applied the **Bonferroni correction** – a relatively simple method to adjust the significance threshold when multiple comparisons are made (in our case, 6 comparisons).
The logic behind this correction is to divide the original significance level ($\alpha = 0.05$) by the number of comparisons k, thereby reducing the overall probability of a Type I error.
Alternatively, one can multiply each individual p-value by the number of comparisons k, which is the method implemented in our Python analysis.

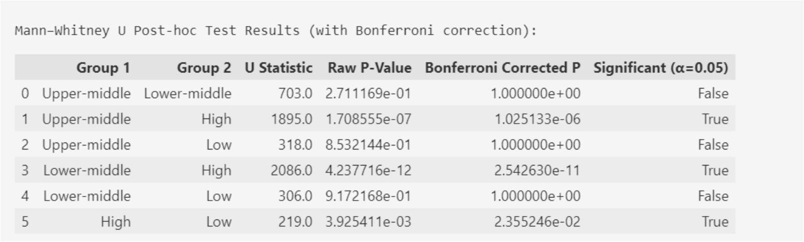
3.3.2 Mann–Whitney U Test Results

The post-hoc analysis revealed **that three out of six** comparisons remained statistically significant after applying the Bonferroni correction:

- Between **High** and **Upper-middle** income groups: $p \approx 1.0 \times 10^{-6}$
- Between **High** and **Lower-middle** income groups: $p \approx 2.5 \times 10^{-11}$:
- Between **High** and **Low** income groups: $p \approx 0.023$:

In contrast, no statistically significant differences were found between the Low, Lower-middle, and Upper-middle income groups.

איור 7: Mann-Whitney U



```
Mann-Whitney U Post-hoc Test Results (with Bonferroni correction):
```

	Group 1	Group 2	U Statistic	Raw P-Value	Bonferroni Corrected P	Significant (α=0.05)
0	Upper-middle	Lower-middle	703.0	2.711169e-01	1.000000e+00	False
1	Upper-middle	High	1895.0	1.708555e-07	1.025133e-06	True
2	Upper-middle	Low	318.0	8.532144e-01	1.000000e+00	False
3	Lower-middle	High	2086.0	4.237716e-12	2.542630e-11	True
4	Lower-middle	Low	306.0	9.172168e-01	1.000000e+00	False
5	High	Low	219.0	3.925411e-03	2.355246e-02	True

3.3.3 Main Conclusion from Post-Hoc Analysis

The results suggest that the most significant differences in plastic waste per capita are concentrated between the High-income group and all other groups.
The High-income group stands out as significantly different from each of the other three income categories.

3.3.4 One-Sided Hypothesis Testing: Do Richer Countries Produce More Plastic?

To further investigate, we conducted one-sided Mann–Whitney U tests to examine whether richer countries **actually produce more plastic per capita**, or perhaps significantly less, compared to lower-income groups.

$$H_0 = (UpperMiddle/LowerMiddle/Low) \geq High$$

$$H_1 = (UpperMiddle/LowerMiddle/Low) \leq High$$

This tests whether high-income countries produce more plastic.
Results :

- High vs. Upper-middle : $p_{value} = 0.9999 \rightarrow$ Fail to reject H_0
- High vs. Lower-middle : $p_{value} = 0.9999 \rightarrow$ Fail to reject H_0
- High vs. Low: : $p_{value} = 0.9981 \rightarrow$ Fail to reject H_0

These very high p-values indicate no evidence that high-income countries produce more plastic than the other groups.

3.3.5 Testing the Opposite Hypothesis: Do Richer Countries Produce Less Plastic?

We then tested the reverse hypothesis:

$$H_0 = (UpperMiddle/LowerMiddle/Low) \leq High$$

$$H_1 = (UpperMiddle/LowerMiddle/Low) \geq High$$

This tests whether high-income countries produce less plastic per capita.
Results :

- High vs. Upper-middle : $p_{value} = 0.000000 \rightarrow p < 0.05$, reject H_0
- High vs. Lower-middle : $p_{value} = 0.000000 \rightarrow p < 0.05$, reject H_0
- High vs. Low : $p_{value} = 0.000000 \rightarrow p < 0.05$, reject H_0

3.3.6 Final Interpretation

These results strongly suggest that middle- and low-income countries exhibit significantly higher per capita plastic waste. This may be due to factors such as:

- Rapid industrialization
- High population growth
- Lack of advanced recycling infrastructure

In contrast, high-income countries often benefit from more efficient waste management systems, better public awareness, and stronger environmental regulations, which likely contribute to their lower levels of mismanaged plastic waste per capita.

4 Results

The statistical analysis revealed a significant association between a country's income level and the amount of mismanaged plastic waste per capita.

The Shapiro–Wilk test indicated that the data are not normally distributed ($p_{value} = 3.217 \cdot e^{-14}$), and therefore non-parametric methods were selected for further analysis.

The Kruskal–Wallis test, used to detect differences between the four income groups, yielded a statistically significant result ($Q = 52.1901$, $p_{value} = 2.7283 \cdot e^{-11}$), indicating that at least one pair of groups differs significantly.

A post-hoc analysis using the Mann–Whitney U test with Bonferroni correction showed that the High-income group is significantly different from each of the other three groups, in that it has a lower level of mismanaged plastic waste per capita. However, no significant differences were found between the Low, Lower-middle, and Upper-middle income groups.

One-sided hypothesis tests further supported this conclusion: In all comparisons, high-income countries were found to produce significantly less mismanaged plastic waste per capita than countries in the other income groups.

5 Conclusions and Discussion

The findings of this study indicate a statistically significant gap between a country's income level and the amount of mismanaged plastic waste per capita.
The statistical analysis showed that high-income countries actually produce less mismanaged plastic waste per capita compared to the three lower income groups.

At first glance, this result may seem counterintuitive, as it is well-known that wealthy countries tend to consume larger quantities of plastic products. However, this is not a paradox; rather, the finding aligns with the fact that high-income countries generally possess more efficient recycling and waste management infrastructure, stricter environmental regulations, and a more environmentally conscious consumer culture.

On the other hand, in low- to middle-income countries, despite having lower average per capita consumption, a substantial portion of waste becomes an environmental pollutant due to a lack of advanced infrastructure, regulatory oversight, and waste treatment capacity. This means that the amount of mismanaged plastic per capita is actually higher in those countries, reflecting operational gaps rather than behavioral ones.

Therefore, an effective environmental policy must emphasize a dual approach: On one hand, encouraging reduced consumption in high-income countries, but equally important, enhancing the capacity of developing countries to manage existing waste. Investments in waste collection, sorting, and recycling infrastructure, environmental education, and accessible waste management technologies are essential steps toward reducing global plastic pollution.
The results of this study underscore the importance of not only measuring levels of consumption, but also evaluating a country's ability to manage the waste it produces.

6 Python Notebook

During the quantitative analysis phase of the study, a Python notebook was used to process and merge data from multiple sources, with a particular focus on low- and lower-middle-income countries.

The dataset underwent cleaning procedures, including the handling of missing values and normalization of relevant variables.

As an extension of the core analysis, a linear regression model was implemented in Python to assess the relative contribution of various factors- including literacy rate, urbanization level, the Environmental Performance Index (EPI), and population size - to per capita plastic waste generation.

This analysis made it possible to identify key drivers of plastic pollution, providing a foundation for practical policy recommendations aimed at developing countries.

7 Bibliography

1. **Jambeck, J.R.,Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady , A., ...** & **Law, K. L.** (2015). Plastic waste inputs from land into the ocean. Science, 347(6223), 768–771. https://doi.org/10.1126/science.1260352
2. **World Bank** (2023). World Bank Country and Lending Groups – Country Classification. Retrieved from: World Bank Country and Lending Groups
3. **OECD** (2022). Global Plastics Outlook: Economic Drivers, Environmental Impacts and Policy Options. OECD Publishing, Paris.Global Plastics Outlook
4. **World Bank** (2018). What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. Urban Development Series.WHAT A WASTE 2.0
5. **Our World in Data**. (2023). Mismanaged Plastic Waste per Capita. Retrieved from: Plastic Pollution
6. **Shapiro, S. S.,** & **Wilk, M. B**. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591–611.
7. **Kruskal, W. H.,** & **Wallis, W. A.** (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583–621.
8. **Mann, H. B.,** & **Whitney, D. R.** (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50–60.