MG-220 Final Course Project

From: Kerrigan de Meij
To: Dr. Craig
Date: December 9th, 2022
Subj: The Statistical Analysis of Demographic Variables' Impacts on Recycling to Assess Current Environmental Protection Methods

---

Section 1: Introduction

Recycling has a generally substantial, positive impact on environmental protection, the economy, and the overall flow of the production process of businesses. Without recycling, there would be additional increases in already excessive amounts of toxic wastes dumping into oceans and landfills, decreases of labor supply in the realm of manufacturing industries that use recycled goods, and diminished usage of domestically-found supplies which come from recycled parts. Because of these consequences, people in modern-day society are more aware of the benefits that environmental protection has, and are therefore voluntarily willing to recycle on a daily basis. However, many people believe that, realistically, recycling alone is not enough to protect the environment. Additionally, the recycling that does take place does not have effective enough outcomes that the world desperately needs it to have. The goal of this statistical analysis is to hopefully gain more insight on what variables have an effect on the amount of recycling that occurs, and from that information begin to formulate methods to improve the way in which we recycle. If people around the world believe that recycling is not currently reaching acceptable standards and it must be changed, we can't make any logical decisions about the problem until we first learn more about what variables have effects on recycling, and then confront the issue with this foundational knowledge. Perhaps it is a matter of the level of education of a country's citizens, people with more money personally available allowing them to recycle more as they have more resources, or the age of the people if younger or older generations are more knowledgeable about recycling and the environment which causes them to recycle more.

<u>Section 2: Statement of Hypotheses</u>

After researching data regarding recycling methods, and overall rates of environmental protection around the world, I began contemplating variables that could affect how much recycling occurs in various geographical regions. With the objective of gaining more insight on where government/society can improve recycling methods, I selected four variables to analyze. These variables may provide foundational knowledge on aspects that could possibly have an influence on the amount of recycling that occurs. The base question is: does GDP Per Capita (by country) have a significant influence on the amount of plastic waste collected for recycling (by country). Depending on what the results display, and with a certain level of confidence, we may conclude that GDP Per Capita does have significant influence on the amount of plastic waste collected for recycling, and thus the government may see this as another reason to prioritize the wealth/financial comfort of its citizens when brainstorming alternative methods to keeping our planet safe and healthy. In addition to analyzing the correlation of these two main variables, I found two other variables which may have an enlightening influence on the amount of plastic waste collecting for recycling: the high school graduation rate in 2018-2019 by country, and the median age of each country's population.

Hypotheses:

1. <u>Null Hypothesis (H_0)</u>: There *is not* a correlation between the GDP Per Capita by country and the amount of plastic waste collected for recycling by country.

a. Alternative Hypothesis (H_1): There *is* a correlation between the GDP Per Capita by country and the amount of plastic waste collected for recycling by country.

2. Null Hypothesis (H_0): The amount of plastic waste collected for recycling by country *is* significantly influenced by the high school graduation rate in 2018-2019 by country, the GDP Per Capita by country, and the median age of each country's population.

a. Alternative Hypothesis (H_1): The amount of plastic waste collected for recycling by country *is not* significantly influenced by the high school graduation rate in 2018-2019 by country, the GDP Per Capita by country, and the median age of each country's population.

Section 3: Data & Analysis Methodology

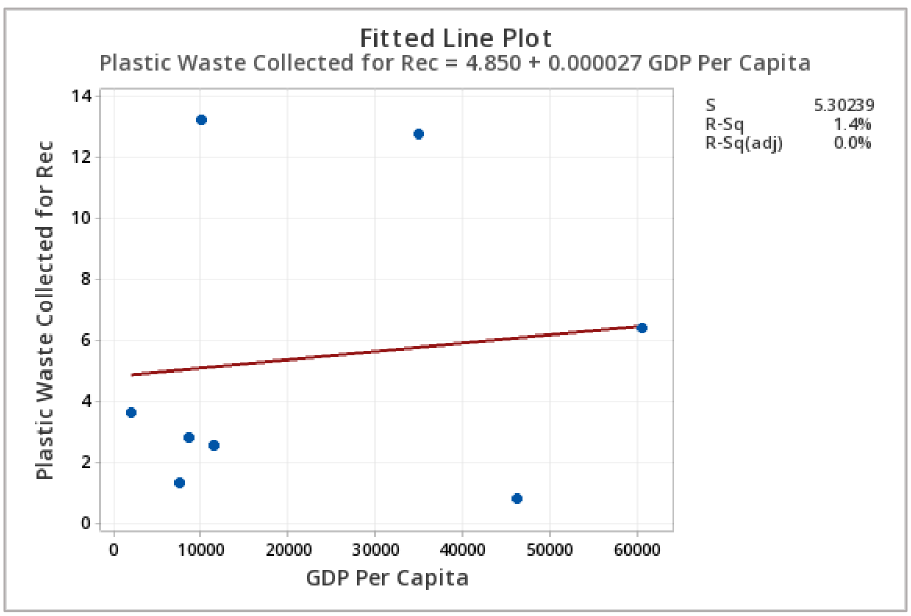
The response data being examined in this analysis, was collected from the Organization for Economic Co-Operation and Development's statistical database. The OECD database provided amounts of plastic waste collected for recycling, in millions of tons (t) of plastics, from the years 1990 to 2019, for various countries. From this raw data, I selected the data from the most recent year of 2019, and narrowed the study down to eight different countries/regions: United States, Canada, Europe, Asia (Pacific), Latin America, Middle East & North Africa, China, and India. Because the most recent data for this variable was from 2019, I chose to use this year for the other three variables in order to see a better comparison of the data. A majority of the data of high school graduation rates, Gross Domestic Product Per Capita, and median age of population for each of the eight countries, was found on World Bank Open Data, with the exception of Canada and China. China's secondary education graduation rate was found in an article written by Chen, Dandan, et al. "Progress and Challenges of Upper Secondary Education in China." Canada's high school graduation rate, in the 2018-2019 school year, was found in the "Statistics Canada" section of the Canadian Government's website. GDP Per Capita is a measure of a country's individual citizens' wealth and economic growth that each person contributes to the nation's economy. One important note about the data included in this statistical analysis, is that the variable GDP Per Capita is valued in current US dollars, in order to allow for a more "apples to apples" comparison between countries/regions.

A simple linear regression in Minitab will test the first hypothesis in this analysis. The response factor on the y-axis will be the amount of plastic waste collected for recycling (in millions of tons), and the predictor on the x-axis will be the GDP Per Capita. This simple linear regression test will display the level of dependance/correlation, and the predictive power that the model has, between the sample individuals' wealth and the amount of plastic waste collected. By looking more specifically at the Coefficient of Correlation, or the actual correlation value known as "r", the test will show a percentage value of how much correlation exists between the two variables. In addition to the "r" value, the test also has a certain level of explanatory power, known as the "r-squared" value. The r-squared value in this analysis, represents the percentage amount of variance in the amount of plastic waste collected for recycling in this sample, that can be explained by the GDP Per Capita. In this case, a high r-squared value and coefficient of correlation means that the model has significant explanatory power, and therefore the amount of plastic waste collected for recycling does depend on GDP Per Capita. If the model's outcome shows significant explanatory power, I would then suggest that instead of attempting to create unattainable new recycling methods, each country should prioritize encouraging well-rounded, financially successful citizens first, and as their available money/resources increases, the amount of plastic waste collected for recycling also increases, effectively protecting our planet in this way. This outcome would be a strong positive correlation, and the Alternative Hypothesis (H_1) would be accepted. If the test's outcome shows that as GDP Per Capita increases, the amount of plastic waste collected for recycling decreases, the variables would have a strong negative correlation, and the Alternative Hypothesis (H_1) would still be accepted. However, if the test results in no significant correlation between the two variables, we would accept the Null Hypothesis (H_0), and reject the Alternative Hypothesis. Such an outcome would mean, in this analysis, that the amount of plastic waste collected for recycling does not depend on citizens' wealth/financial status, shown as the variable GDP Per Capita.

The second set of hypotheses will be tested using Multiple (Multivariate) Regression in Minitab, in order to evaluate the effect of each predictor (independent) variable on the response (dependent) variable, which is the amount of plastic waste collected for recycling. Once again, the r-squared value will display the percent of variance in plastic waste collected for recycling, that is explained by the three predictor variables of: high school graduation rate in 2018-2019 by country, the GDP Per Capita by country, and the median age of each country's population. Thus, the higher the r-squared percentage is, the more predictive power the model has. To evaluate the predictors' influence, individually, I will also examine the "p-value" of the test's results, which indicates if the predictor has a significant influence on the response or not. Regarding the p-value of the predictor, if it is less than 0.05, then it does have a significant influence on the response variable. For example, if the independent variable of high school graduation rates has a p-value of 0.01, this would mean that the high school graduation rates of each country does have a significant influence on the amount of plastic waste collected for recycling. For my personal prediction, I am unsure about the predictive power of the model I have created (depending on the r-squared value), but I am predicting that the high school graduation rate variable may have a lower p-value than the other variables (below 0.05), meaning it has more significant influence on the amount of plastic waste collected for recycling. I decided this prediction because if more people have graduated high school, they may more likely be educated on the benefits that recycling has on the environment. Therefore, with a p-value under 0.05 and a positive coefficient: as graduation rates increase, the amount of plastic waste for recycling also increases, showing a positive correlation relationship. If this prediction is wrong, and the p-value for each predictor is greater than 0.05, then they have no significant influence on the response variable, and the Alternative Hypothesis (H_1) will be accepted.

<u>Section 4: Results</u>

Simple Linear Regression Analysis: Plastic Waste Collected for Rec. (t) versus GDP Per Capita



Method

Correlation type	Pearson
Number of rows used	8

Correlations

	Plastic Waste Collected for Rec
GDP Per Capita	0.119

Opposing my own personal prediction, the simple linear regression test visually shows there is almost no correlation between the amount of plastic waste collected for recycling and GDP Per Capita. Additionally, besides the visual scattering of points in the graph above which shows no obvious pattern, the actual Coefficient of Correlation (r) of 0.119, further supports the result that these two variables have basically no correlation between them. With a perfect r value that shows no correlation being at exactly 0.00, this model's r value is extremely close to having absolutely no correlation. Again, furthering the absence of correlation between these two variables in the sample, the r-squared value is only 1.4%. This r-squared value means that only 1.4% of the variance in plastic waste collected for recycling can be explained by the GDP Per Capita. With such a low Coefficient of Correlation (r = 0.119) and explanatory power of only 1.4% for this statistical model, the Null Hypothesis (H_0) is accepted. In terms of what this result practically means, the amount of plastic waste collected for recycling does not significantly depend on people's monetary resources, thus society should focus on other variables that can be improved/altered in order to attempt to increase the amount we recycle for our environment.

Multiple Regression Analysis: Plastic Waste Collected (t) versus High School Grad. Rate, GDP Per Capita, Median Age of Population by Country

Regression Equation

Plastic Waste Collected for Rec = -27.4 + 9.7 High School Grad. Rate (2018-20
- 0.000156 GDP Per Capita
+ 0.808 Median Age of Population (2019)

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-27.4	24.7	-1.11	0.330	
High School Grad. Rate (2018-20	9.7	30.3	0.32	0.765	1.23
GDP Per Capita	-0.000156	0.000114	-1.37	0.243	2.21
Median Age of Population (2019)	0.808	0.403	2.01	0.115	2.33

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.39706	54.81%	20.91%	0.00%

After examining the results of the Multiple Regression test, the Null Hypothesis (H_0) is rejected, and we accept the Alternative Hypothesis (H_1) because there is no significant influence from any of the three predictor variables. This Alternative Hypothesis (H_1) is supported by the p-values of the high school graduation rate variable at 0.765, the GDP Per Capita at 0.243, and the median age of each country's population at 0.115, which are all well above the targeted p-value of 0.05 (5%) and therefore do not have any significant influence on the amount of plastic waste collected (the response variable). Regarding the coefficients for each of these variables, any influence that the high school graduation rate and the median age of the population do have, display a positive relationship, meaning that as these variables increase, so does the response variable of plastic waste collected. In contrast, the GDP Per Capita displays a negative coefficient relationship, meaning that as GDP Per Capita increases, the amount of plastic waste collected decreases, although these variables were already shown to not have a significant influence no matter the coefficient sign. These results disagree with my original predictions for the analysis, revealing that the amount of plastic waste collected for recycling is not significantly affected by a country's age, education, or wealth (in this sample's case). Regarding the r-squared value for this test at 54.81%, the common theme of this analysis is once again reinforced, being that society may need to consider the other factors that are outside of this model and unaccounted for, when examining improvements to environment protection methods such as recycling. These outside, unaccounted for factors presumably have more of a significant impact on the response variable than I had hoped for when selecting the variables for this analysis. This is quantified by the r-squared value of 54.81%, because only this amount of variance in plastic waste collected can be explained by this model, which leaves a considerable amount (45.19%) that is unexplained and unaccounted for. Overall, the three independent variables used were insignificant to the response variable, and not surprisingly, the study showed relatively low explanatory power as well.

Section 5: Discussion & Conclusion

The main theme throughout the results of this statistical analysis disagreed with my predictions. Originally, the goal was to find clarity on a few more interesting demographic variables that could possibly impact recycling, in order to understand the influence that they have on the amount that society recycles. Based on that improved knowledge, organizations could attempt to increase or change the environmental protection methods we currently use. However, the simple linear regression and multiple regression tests both resulted in relatively ineffectual findings, due to the Coefficient of Correlation (r) of 0.119 between plastic waste collected and GDP Per Capita, meaning that these two variables show basically no correlation, and the p-values of each variable also being insignificant at: high school graduation rate (0.765), GDP Per Capita (0.243), and the median age of each country's population (0.115). Although there was no impactful relationship between these predictor factors and the amount of plastic waste that is collected for recycling, it is still important for this information to be understood. After concluding the analysis, I found that it is that much more important for people to examine why our environmental protection methods are not effective enough, because this sample's results show that it is not in the fault of the country's education, age, or monetary resources. Nevertheless, it is equally important to recognize the shortcomings of this analysis that could exponentially improve the outcomes and understandings of these relationships, such as increasing the different independent variables utilized, accounting for geographical impacts, for example, general differences in towns/cities that are closer to oceans or forests, and increasing the number of countries in the analysis.

Works Cited

Baffes, John, et al. *World Bank Open Data*. 1 Dec. 2022. https://data.worldbank.org/.

Chen, Dandan, et al. "Progress and Challenges of Upper Secondary Education in China." *Open Knowledge Repository*. World Bank, Washington, DC. 1 Feb. 2019. https://openknowledge.worldbank.org/handle/10986/31574.

Elez, Nikola, and Klarka Zeman. "High School Graduation Rates in Canada, 2016/2017 to 2019/2020." Government of Canada, Statistics Canada. 20 Oct. 2022. https://www150.statcan.gc.ca/n1/pub/81-599-x/81-599-x2022002-eng.htm.

"Municipal Waste, Generation and Treatment." *OECD*. https://stats.oecd.org/Index.aspx?DataSetCode=MUNW.

PowerPoints and Notes from Professor Craig's Data Analysis MG-220 Class.