ISSN 2959-6157

The Contributing Factors of Individual Carbon Footprint based on Linear Regression

Haoxi Geng

Shanghai Foreign Language School Affiliated to SISU, Shanghai, 200000, China
Corresponding author: cindyhx888@outlook.com

Abstract:

Public concerns have been raised by recent climate changes. The temperature rise has and will result in more extreme weather and the melting of the Atlantic's ice, which must be paid attention to. Different organizations have stressed the importance of achieving carbon neutrality after conducting investigations on the environment. The purpose of this paper is to address the issue of carbon emissions and seek solutions to meet the government's objectives. This article selects several daily activities and correlates them with carbon emissions. Using the linear regression model as well as the stepwise regression model, the collected data can demonstrate how various daily activities affect an individual's carbon emissions. The result shows that almost all daily activities that use fuel will eventually have an impact on an individual's carbon footprint, as found in this paper. Moreover, this paper suggests several possible ways for future development to be achieved in order to achieve the realization of carbon neutrality.

Keywords: Carbon emission; individual's carbon footprint; linear regression; step-wise linear regression.

1. Introduction

Nowadays, climate change has gradually attracted people's attention. The temperature on the surface of the Earth has increased by 1.05 Celsius compared to a century ago. As one of the severest tasks humanities will ever face, the decisions made today could have a major impact on Earth's future. Recently, the phrase carbon emission appeared frequently in various news articles and newspapers about the Earth's atmosphere. Carbon emission refers to the amount of carbon dioxide emitted into the air, which has a direct relationship to global warming. Many nations have made plenty of efforts to safeguard the planet where humans live.

According to data from the Intergovernmental Panel on Climate Change (IPCC), the crucial goal to ensure the safety of human living conditions is to restrict the growth of temperature to 1.5 degrees Celsius. From this point forward, rising temperatures will eventually cause more extreme weather and the melting of the Atlantic's ice. According to official data released by the British government, both individual carbon emissions and carbon footprints appeared to be upward trending. The Chinese government has responded to the requirement by making efforts to achieve carbon peaking at or before the year 2030 and become carbon neutrality thirty years later. Not only has China set plans, but countries from all over the world have also shown their determination towards carbon neutrality. Government support is useful in promoting firm carbon emission efficiency, but an individual's effort to reduce carbon emission is, at the same time, necessary [1].

There have been many studies focusing on different aspects to reduce carbon emissions. A study by Rokhmawati used the Stochastic Impacts by Regression on Population, Affluence, and Technology model (STIPRAT model) to reflect the impact of a carbon tax on carbon emission in Indonesia and concluded that with the help of renewable energy, the government will be able to realize the targets on time [2]. This paper offers a new approach to limiting carbon emissions through choosing different kinds of energy, emphasizing the emission produced by traditioned fuels is Indonesia's main concern. China is the biggest carbon dioxide emitter in the world; hence the Earth's atmosphere will be greatly impacted by any reduction in carbon emissions. The use of fuels and Gross Domestic Product (GDP) per capita were identified as the two primary emission causes by Lin's study [3]. Methods such as controlling the urbanizing speed as well as increasing energy efficiency will be effective in facing this issue [3]. From the two papers mentioned above, carbon dioxide emitted when consuming fuels is a crucial factor that should raise people's concerns. Focusing on this perspective, a study conducted has revealed the effect of speed

limitations through a model based on energy conservation law [4]. With the increase of speed restrictions from 80 to 120 kilometers per hour, the average amount of carbon emitted will result in an increase of approximately 33% to 38%, reflecting a positive relationship between carbon emission and car speed [4].

Simultaneously, research about how different aspects of people's lives except vehicles influence carbon emissions is available. With the development of digital technology, this technology will result in the growth of carbon emission in residential areas [5]. Additionally, a paper that focuses on the emission of greenhouse gases in houses in the United States uses the Life Cycle Assessment (LCA) method to calculate the carbon footprint revealing that the footprint mainly is due to transportation and electricity used in the family [6]. According to a survey regarding the second-hand trend in the fashion business, people's preferences for eco-friendly clothing are growing [7]. What is euphoria is that individuals around the world have become more aware of the protection of the planet. Through gathered data, another study made a comparison between vegetarian and omnivorous which has a significant influence on the environment [8]. To reduce carbon emission, less consumption of meat products as well as the increase in the production of organic food should be promoted [9, 10].

These studies have concentrated on summarizing or evaluating a specific viewpoint. This paper aims to focus on the influence of different factors on carbon footprint by using the linear regression model. Using the linear regression model can not only find the influence level of different variables on the carbon footprint but also anticipate future trends. Determining the primary source of carbon footprint will enable more beneficial implementation and help determine the order of importance for various measures.

2. Methods

2.1 Data Source

The data set was published by Mesut Duman on the website Kaggle. In the creation of the dataset, the author gathered information from authoritative papers and documents to generate the most close-to-life data.

2.2 Variable Selection

This data set lists factors such as an individual's vehicle type as well as his or her frequency of taking a shower. Using numeric and text statistics, a person's action, which will more or less impact the environment, is listed. The dependent variables as well as their descriptions will be shown in Table 1 below.

In this data set, 10000 people's living habits are involved to determine the dominant factor contributing to the individual carbon footprint. Since the population of the human race is significantly larger than 10,000, using the 10% rule in statistics, it can be assumed that the sampling is done with replacement. Since there is an efficient number of data and various variables used to define the weight, these dependent variables can be excellent indicators of how human activities affect carbon emission.

Table 1. Variable introduction

Variable	Logogram	Description
'Body Type'	x_1	body type
'Diet'	x_2	vegan (1), vegetarian (2), pescatarian (3), omnivore (4)
'How Often Shower'	x_3	less frequent (1), daily (2), twice a day (3), more frequent (4)
'Heating Energy Source'	x_4	natural gas (1), wood (2), electricity (3) Coal (4),
'Transport'	x_5	walk/bicycle(1), public(2),private(3)
'Vehicle Type'	x_6	none (1), LPG (2), hybrid (3), petrol (4), diesel (5)
'Social Activity'	x_7	never (1), sometimes (2), often (3)
'Monthly Grocery Bill'	x_8	the money spent on groceries monthly in dollars
'Traveling by Air'	x_9	never (1), rarely (2), frequently (3), very frequently (4)
Vehicle Monthly Distance	x_{10}	the distance traveled by individual monthly in kilometers

'Waste Bag Size' 'Waste Bag Weekly Count' 'TV PC Daily Hour' New Clothes Monthly	x_{11} x_{12} x_{13} x_{14}	small (1), medium (2), large (3), extra-large (4) the total amount thrown away in a week the amount of time watching PC or TV each day the amount of new clothes bought each month
Internet Daily Hour 'Energy efficiency'	x_{15} x_{16}	the amount of time using the Internet each day yes(1), sometimes(2), no(3)

2.3 Method Introduction

To identify the significance of different factors affecting individual carbon emission, this paper uses the model multiple linear regression (MLR) to satisfy the need. The expression for the multiple linear regression can be expressed as:

$$y = k^T x + b \tag{1}$$

In this function, x refers to an input variable that includes multiple features, in other words, dependent variables. k^T is the weight of the features, while b is the intercept or offset of the model. The result y is the output of this model or the anticipated goal variable.

The y variable in this function selected for this paper is the carbon footprint. The association between each variable and the carbon footprint can be found by computing the regression coefficient for each variable. Thus, the formation mechanism of y can be explained, and the prediction of y will also be accessible. The final function is:

$$y = k_1 x_1 + k_2 x_2 + \cdots + k_{16} x_{16} + b \tag{2}$$

The variables $x_1, x_2 \ldots x_{16}$ are defined in Table 1.

3. Results and Discussion

3.1 Data Visualization

In the data set, the variables are either presented as text, describing their classification in a certain category, or as numbers. The data represented by text under each category is evenly distributed according to the classification contained in that section. For instance, the data for different people's diets are shown in figure 1.

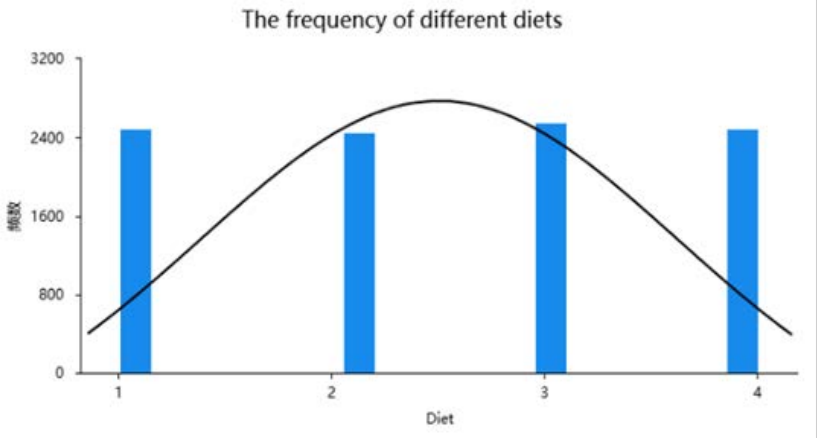

Fig. 1 Box plot of the frequency of different diets

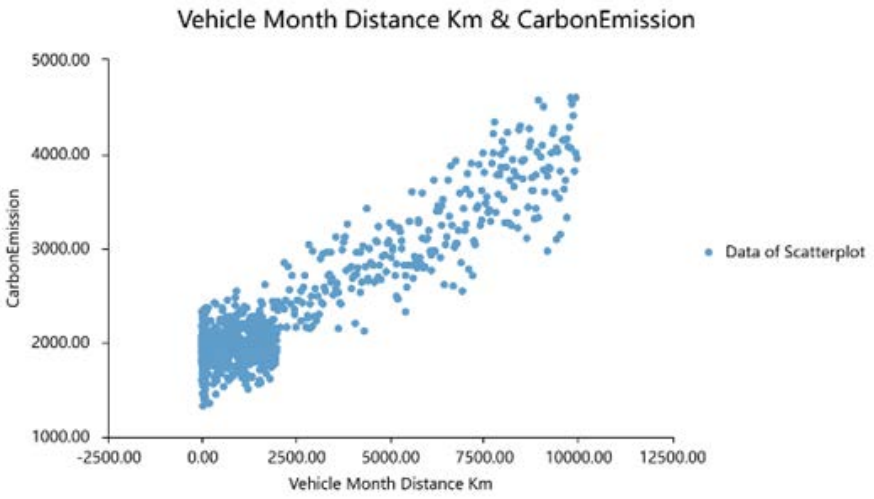

Fig. 2 The scattered chart of the monthly distance travelled by vehicles and the carbon emission

Numerical data is another kind of data used in this paper. Monthly mileage driven by a car is a representative type of data of this kind. The variable representing the distance traveled by the vehicle each month, as shown in figure 2, is mainly clustered in the range above 0 but below 2500. In this range, carbon emissions are mostly below 3,000. With an increase in distance, there is an upward tendency for carbon emissions. How strongly this is correlated will be discussed later in this paper.

In this research, the dependent variable is carbon emission. Based on the data, the median quantity of emission, as shown by the violin graph in figure 3, is 2080. Between 1538 and 2748, half of the data were collected; in the meantime, 2.5% of the data on carbon emissions exceed 4867.

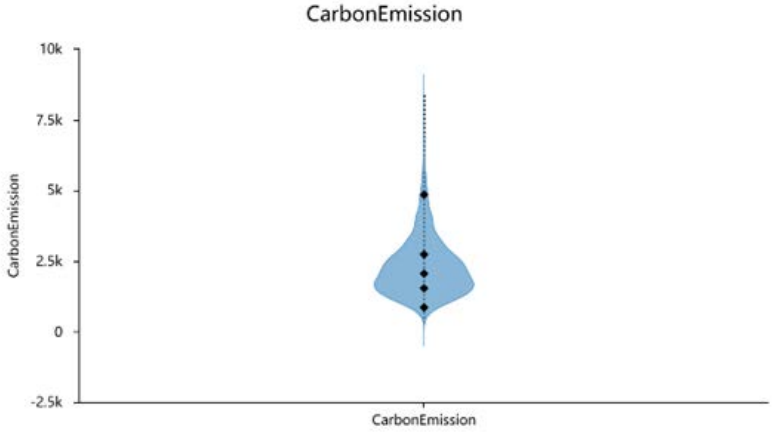

Fig. 3 The violin chart for the distribution of the amount of carbon emission

Since there are sixteen variables representing different aspects of daily activity that influence carbon emission, an analysis of the correlation between these variables is necessary. Pearson correlation is used in this section to show the relationships between the variables. The extent to which each variable is related is displayed in figure 4.

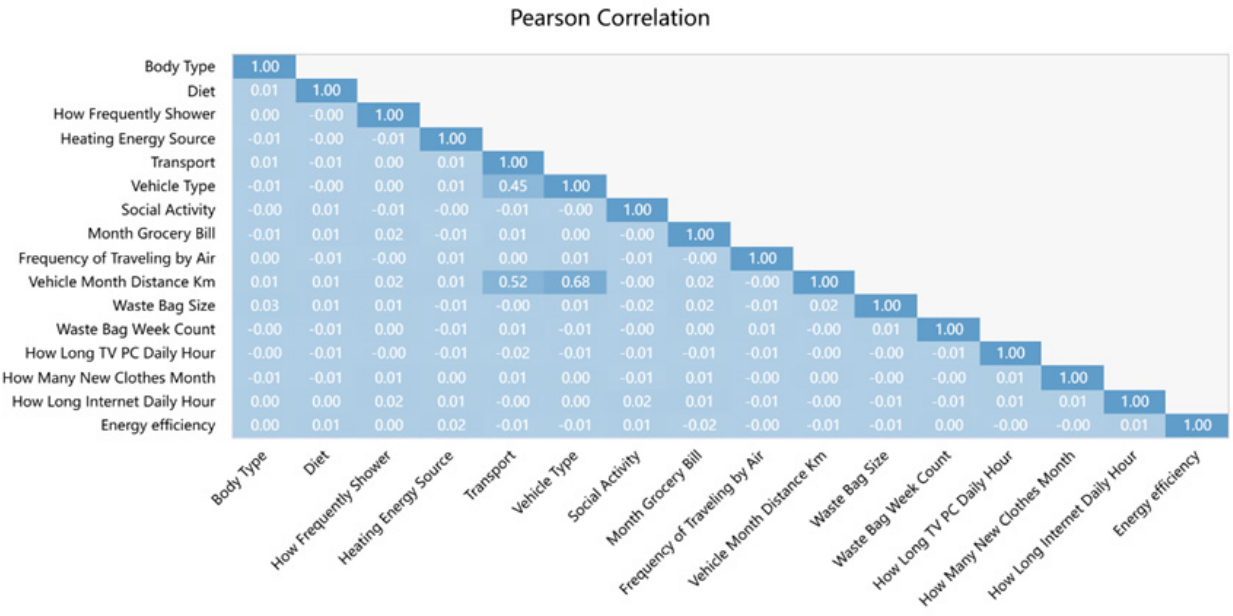

Fig. 4 Pearson Correlation of the variables

When the correlation coefficient shown in figure 4 is close to zero, the variables are not related to each other. The correlation coefficient ranges from zero to one, the closer to one, the stronger relation is proposed. Therefore, from figure 4, most of the variables are not related to others and they are presented to be independent. However, this is not the case for three variables: vehicle type, vehicle month distance, and transport. Since whether people choose private or public transportation is directly related to whether they have a vehicle, it is not surprising to learn that the three variables have a close relationship. By discovering this, the link may indicate that transportation plays a significant role in predicting a person's carbon emissions. It also aided in the following examination of the significance of the sixteen variables that influence a person's carbon footprint.

In addition, all of the coefficients through the Pearson Test are below 0.8, demonstrating that the multicollinearity problem, which might affect the accuracy of the model, does not exist.

3.2 Model Results

By using the data collected, Table 2 shows how each dependent variable affects carbon emissions and estimates how effective this model is. Table 2 serves as an indicator for the curve fitting of this model. This reveals that the sixteen variables can explain 72.9% of the change in carbon emission. This model passes the test when it is applied to the F-test in statistics. (F (16,9983) = 1674.297, p = 0.000). The F-test indicates that among the sixteen variables listed, at least one will affect the carbon emission increase or decrease, which means the establishment of this model is meaningful. In addition, by testing the multicollinearity of the variables in this model, all of the VIF values are below five, which eliminates the possibility of having a multicollinearity problem. In addition, the D-W value is 2.035, a value almost equal to 2, reflecting that autocorrelation does not exist in this model and there is not any correlation between the data collected. The best fit function of this model is:

$$f(x) = -1496.569 + 223.835x_1 + \ldots + 6.446x_{15} + 25.972x_{16} \quad (3)$$

The x variables are explained in Table 1. By reviewing the P value in this model, all of the variable except how often an individual takes a shower have a significant impact on the carbon emission by a person.

Table 2. Linear Regression Model results

	Unstandardized Coefficients		Standardized Coefficients	t	p	collinearity diagnostics	
	B	Std. Error	Beta			VIF	Tolerance
Constant	-1496.569	48.664	-	-30.753	0.000**	-	-
Body Type	223.835	6.381	0.183	35.080	0.000**	1.002	0.998

	Unstandardized Coefficients		Standardized Coefficients	t	p	collinearity diagnostics	
	B	Std. Error	Beta			VIF	Tolerance
Diet	53.153	4.754	0.058	11.180	0.000**	1.002	0.998
How Frequently Shower	6.972	4.774	0.008	1.460	0.144	1.001	0.999
Heating Energy Source	70.087	6.406	0.057	10.941	0.000**	1.001	0.999
Transport	-138.876	13.214	-0.065	-10.510	0.000**	1.397	0.716
Vehicle Type	111.542	5.874	0.137	18.988	0.000**	1.926	0.519
Social Activity	81.539	6.475	0.066	12.592	0.000**	1.001	0.999
Month Grocery Bill	1.031	0.074	0.073	14.020	0.000**	1.002	0.998
Frequency of Traveling by Air	437.736	4.750	0.481	92.157	0.000**	1.001	0.999
Vehicle Month Distance Km	0.194	0.003	0.529	69.971	0.000**	2.105	0.475
Waste Bag Size	99.619	4.082	0.127	24.402	0.000**	1.002	0.998
Waste Bag Week Count	81.915	2.667	0.160	30.712	0.000**	1.001	0.999
How Long TV PC Daily Hour	3.032	0.747	0.021	4.058	0.000**	1.001	0.999
How Many New Clothes Month	13.760	0.361	0.199	38.096	0.000**	1.001	0.999
How Long Internet Daily Hour	6.446	0.730	0.046	8.834	0.000**	1.001	0.999
Energy efficiency	25.972	6.568	0.021	3.954	0.000**	1.001	0.999
R 2	0.729						
Adj R 2	0.728						
F	F (16,9983)=1674.297,p=0.000						
D-W	2.035						
Dependent Variable=Carbon Emission							
* p<0.05 ** p<0.01							

3.3 Further analysis

By using stepwise regression, which will automatically eliminate factors that are not significant, a new model is established. (How frequently shower) is disregarded in this case. The function of the new one will be:

$$f(x) = -1479.364 + 223.872x_1 + 53.125x_2 + \ldots + 6.463x_{15} + 25.980x_{16} \quad (4)$$

This model can explain 72.8% of the changes in carbon emission. Since the variables are almost the same, the difference between the two models is not so significant.

3.4 Discussion

Based on the linear regression model as well as the stepwise regression model, they reached the same conclusion that 15 of the 16 variables attributed to carbon emission increase. As climate change and environmental problems are no longer far away from people's lives, the investigation of the significance of the factors can help to design an effective way to face these challenges. In daily life, the things that affect the carbon footprint are variable, however, that also indicates that people can start to change their living habits from simple affairs to help with the problem of climate change. For instance, buying fewer clothes each day or using environmentally friendly energy will contribute a lot to the process of controlling an individual's carbon footprint.

The limitation existing in this paper is that it considers some of the activities that have an impact on the environment. A more comprehensive analysis could be made with a greater dataset as well as including more factors.

4. Conclusion

Throughout this paper, the importance of different dependence variables is calculated and discussed by using linear regression as well as a stepwise regression approach. In doing so, more effective ways to cut carbon emissions can be proposed, and individual's awareness of how daily activities affect the environment can be raised. By using

the approaches mentioned, most of the activities, have a significant impact on the carbon footprint. The ways to reduce carbon emissions can be categorized into two ones: the investigation of more environmentally-friendly energy and use less energy if not necessary. These ways can reduce the level at which activities affect carbon emissions in person. This conclusion has also been reached in the article stressing sustainable fuels will be crucial to the lowering of carbon emissions in airline activity and influence an individual's carbon footprint greatly. Additionally, an article suggests that using carbon-neutral methanol fuel can not only increase the cars' efficiency but also benefit the environment. These articles could indicate that the general ways to reduce carbon emissions concluded by this paper are correct and effective.

References

[1] Jie Jin, Fan Wang. Impact of government support on firm carbon emission efficiency: The transmission channel of green innovation, Finance Research Letters, 2024, 68.

[2] Andewi Rokhmawati, Vita Sarasi, Lailan Tawila Berampu. Scenario analysis of the Indonesia carbon tax impact on carbon emissions using system dynamics modeling and STIRPAT model, Working paper, 2024.

[3] Boqiang Lin, Xiying Liu. Carbon Emissions in China's Urbanization Stage_Influencing Factors and Emission Reduction Strategies. International Review of Economics & Finance, 2022.

[4] Chao Gao, Jinliang Xu, Miao Jia, Zhenhua Sun. Correlation between carbon emissions, fuel consumption of vehicles and speed limit on expressway. Journal of Traffic and Transportation Engineering (English Edition), 2024.

[5] Zhengning Pu, Jingyu Liu, Mingyan Yang. The effect of digital technology on residential and non-residential carbon emission. International Review of Economics & Finance, 2024, 95.

[6] Goldstein B, Gounaridis D, Newell J P. The carbon footprint of household energy use in the united states. Proceedings of the National Academy of Sciences, 2020, 117(32).

[7] Jin Yang, et al. Predicting the significance of consumer environmental values, beliefs, and norms for sustainable fashion behaviors: The case of second-hand clothing. Asia Pacific Management Review, 2024, 29(2): 179-194.

[8] Baroni L, Cenci L, Tettamanti M, Berati, M. Evaluating the environmental impact of various dietary patterns combined with different food production systems. European Journal of Clinical Nutrition, 2007, 61(2): 279-286.

[9] Bofan Wang, Zhao Jia Ting, Ming Zhao. Sustainable aviation fuels: Key opportunities and challenges in lowering carbon emissions for aviation industry. Carbon Capture Science & Technology, 2024, 13.

[10] Chao Jin, et al. Study on the stability and properties of carbon neutral methanol in blends with diesel fuel. Fuel, 2024, 374.