MathWorks Math Modeling Challenge 2023

Conestoga High School

Team #16424, Berwyn, Pennsylvania

Coach: Allison Youndt

Students: James Johnson, Evan Lu, Daniel Tu, Vincent Yao, Evelyn Yu

M3 Challenge FINALIST—$5,000 Team Prize

JUDGE COMMENTS

Specifically for Team # 16424 —Submitted at the Close of Triage Judging:

COMMENT 1: Very good summary. Great discussion on the modeling presented (including assumptions, strengths and weaknesses as well as sensitivity analysis). Improving citation skills would be helpful.

COMMENT 2: The team provides a good executive summary that includes the predicted results and insights into the methods the team used. The paper is well organized and it is easy to follow. References citations are included. The model revision and discussion sections provide insights on the methods used to address the questions.

The response to the first question is strong. An exponential model is used for the projection of e-bike sale in US. Insights on the method used are discussed, including sensitivity analysis, weaknesses and strengths. The response to the second question is also strong. The team did not simply analyze the correlation between e-bike sales and potential factors using coefficient of determination but also tried to determine the causality. Figure 2.5 and 2.6 are the same (wrong plot is included in Figure 2.6). The discussions are thorough.

The team uses a multivariate linear regression for the third question. The response is good but is not as strong as the previous two questions. The data used to determine the coefficients are not clearly explained or indicated and citations are not specific.

COMMENT 3: Well written and highly organized. Model is presented and explained.

COMMENT 4: You answered the questions and presented good critical and analytical thinking. Well done!

Pedaling Into The Future: Exploring The Current Surge In Electric Bike Usage

Executive Summary

To the U.S. Department of Transportation:

In recent decades, a growing interest in sustainable and eco-friendly modes of transportation facilitated the rise of the electric bicycle around the world. In 2021, the United States witnessed a 70% increase in annual e-bike sales. Gaining popularity among commuters, delivery services, and casual riders alike, e-bikes are now the most demanded electric vehicle on the market. In this paper, we aim to explore the reasons behind the growing popularity of e-bikes and their potential to revolutionize the transportation industry.

The first foundational question regarding surging e-bike popularity lies within their potential for growth. How many e-bikes will be sold in the coming years? The first section of this report proposes a model to predict growth in e-bike sales and estimates the number of e-bikes sold two and five years from now. After tempering American e-bike purchase data with European data, we created an exponential regression equation to predict e-bike sales in the United States. Our model predicts the sale of 1,560,080 e-bikes in 2025 and 3,045,720 e-bikes in 2028. This suggests a continuing demand for e-bikes as a method of transportation.

Recognizing the underlying causes of e-bike growth is essential for understanding their future role in transportation. What factors—such as gas prices and environmental awareness—contribute most to the growth of e-bike usage? We identified three primary factors—gas prices, personal income, and lithium battery prices—that may play significant impacts on e-bike usage growth. Using a two-stage instrumental variable regression method, we determined causal relationships between each of the three primary factors and e-bike usage. Our analysis indicated that there was a strong positive causal relationship upon e-bike usage from real household income levels ($R^2 = 0.8313$), a moderate negative causal relationship with the cost of lithium-ion batteries ($R^2 = 0.4885$), and no significant causal relationship with the cost of gas ($R^2 = 0.2582$).

Finally, the rise in e-bikes may affect other elements of human existence. How will the growth in e-bike sales impact aspects of our daily lives? We investigated the impacts of e-bike usage on two primary factors: carbon emissions and traffic congestion. By implementing a multivariate linear regression analysis and a Monte Carlo simulation to analyze the impacts of e-bike sales and other confounding variables on carbon emissions and traffic congestion, we concluded that e-bike sales indeed have an impact on both these factors. Specifically, a sale of an e-bike correlates with a decrease of 0.1099 metric tons of carbon emissions per year and a decrease of 0.0032 minutes or 0.192 seconds in the average commute time.

We hope these models and predictions are used to educate and guide policymakers in the process of legislating e-bikes into sustainable energy plans, such as removing cars from roads, adding tax incentives for e-bike usage, and investing in bike lanes.

Overall, this research paper aims to provide a comprehensive analysis of the increasing use of electric bikes and its potential impact on the environment, economy, and society.

Contents

Introduction

As an increasingly fast, reliable, convenient, and cheap avenue of transportation, electric bikes are emerging as a popular alternative to conventional transportation options like public transit and motor vehicles. This technology threatens to facilitate a paradigm shift in transportation, revolutionizing the way in which humans spend their time while posing potential consequences for the environment and public health.

The first problem required the generation of a model to predict American e-bike sales in the future. Using the model, we estimated the number of e-bike sales in 2025 and 2028 in the United States.

The second problem challenged us to evaluate the influence of individual factors on the rising trend in e-bike sales. By determining the extent to which a factor impacted the trend, we learned about the underlying causes of growing e-bike popularity.

The third problem sought to analyze the resulting impacts of increased e-bike transportation on important real-world factors. Two primary factors we investigated were traffic congestion and carbon emissions.

The answers to these problems bear serious implications for lawmakers seeking to establish regulatory pathways for this emerging technology.

1 Q1: The Road Ahead

1.1 Defining the Problem

In this problem, we were tasked with creating a model to predict growth in e-bike sales. E-bikes are defined as bicycles equipped with an electric bike motor to assist in pedaling; this definition does not include mopeds and sit-down scooters. We then used the model to predict e-bike sales in 2025 and 2028 in the United States.

1.2 Assumptions

1. **There will not be any technological advancements in e-bikes, motor vehicles, conventional bicycles, or any other form of urban transportation that significantly alter the ability of humans to navigate land terrains.**
   Justification: While technological changes can revolutionize the transportation industry in terms of convenience and pricing, breakthrough technologies take years to develop, and mass adoption is unpredictable.

2. **The trends for future e-bike sales growth and decline are the same as past trends.**
   Justification: Short-term fluctuations have the potential to dramatically alter the transportation landscape. Notably, the COVID-19 pandemic forced the temporary closure of many public transportation systems in cities, resulting in a sudden rise in e-bike sales as a substitute method for transportation. Such variations are impossible to predict, so our model assumes that changes in e-bike sales will follow the same long-term trends shown in past data; we assume that a significant transportation-altering event will not occur.

3. **The trend for e-bike sales follows a compound growth rate.**
   Justification: Companies measure their own growth through compound growth because a percentage of their revenue is reinvested into their production, and they must see an equal return on the investment in future years in order to maintain financial success. This effect is especially prominent in the early stages of a company's growth. The e-bike market is still young.

4. **The U.S. and European data measure e-bike sales using the same definition of "electric bicycle."**
   Justification: As e-bikes are relative newcomers to the grand stage of transportation, there

remains some disagreement with regard to the specific definition of an e-bike. For example, while some bikes merely assist the rider's organic pedal power, others add a throttle, integrating moped-style functionality. Since differing definitions may result in inconsistent data, we assume that data was collected with the same underlying definition.

5. **Time is the only variable necessary for projecting future sales.**
Justification: Projecting future sales requires finding the change in sales as a function of time. Although sales may be affected by an indefinite number of variables, each of these factors is also dependent on and change with respect to time. Adding up all of these effects, or summing the derivatives of each factor with respect to time, will give us the total derivative of sales with respect to time. We are investigating this total summation of the change in e-bike sales, so it is unnecessary to solve for each factor individually.

1.3 Variables

Symbol	**Definition**	**Units**
S_A	Total unadjusted predicted number of e-bike sales in the United States for a given year t	Number of e-bikes sold
S_E	Total predicted number of e-bike sales in the EU for a given year t	Number of e-bikes sold
S_U	Total adjusted predicted number of e-bike sales in the United States for a given year t	Number of e-bikes sold
t	Number of years	Years since 2000

1.4 The Model

Given the limited e-bike sales data provided for the U.S., we sourced additional data published by Statista's research department for the years 2012 – 2016 (Statista Research Department).

Year	**E-bike sales in the U.S. (in thousands)**
2012	70
2013	159
2014	193
2015	130
2016	152

Combining the given data with the additional data, we performed exponential regression by year for e-bike sales in the U.S. and EU and found the following equations (E-Bikes - U.S. Sales 2016):

US total predicted e-bike sales (unadjusted)

$$S_A = 5.915e^{0.2243t} \tag{1}$$

$$R^2 = 0.9484 \tag{2}$$

EU total predicted e-bike sales

$$S_E = 51.416e^{0.2217t} \tag{3}$$

$$R^2 = 0.9934 \tag{4}$$

We observed that the exponents of the U.S. and EU exponential regression equations were similar, at 0.2243 and 0.2217, respectively. However, we determined two important factors underlying the nature of the data that should be acknowledged and accounted for. Firstly, the EU data did not contain entries from any COVID-19 pandemic years, which affected economic conditions drastically and altered the volume of market sales, including e-bikes. In comparison, the more recent U.S. data reflected these changes, evident in the decrease in sales in 2020, and the steep increase in 2021. Secondly, the data for the U.S. was very sparse compared to the EU's (eBicycles). With fewer data points and a gap within the data, the U.S. data was more likely to be subject to economic fluctuations that may not reflect the long-term growth trend of the e-bike market. Furthermore, the R^2 coefficient of determination value for the EU model was higher than that of the U.S. at 0.99, and 0.95.

Given the previously stated factors, we decided that an adjusted model that accounted for these changes would more accurately predict e-bike sales than the one determined from the raw regression for the US. Additionally, since the economic conditions of the two countries are relatively similar, it would be reasonable to use the EU data to help normalize the U.S. data. Doing so would help make the resulting model more accurately project into the future. We averaged the exponents of the two equations and came up with the following adjusted equation to model e-bike sales in the US:

$$S_U = 5.915e^{0.223t} \tag{5}$$

Since data concerning e-bike sales was provided only for the EU as a whole, and specific data on the UK was not provided nor readily available online, we chose not to use our model to project future e-bike sales in the UK to avoid over-generalizing for this specific population.

We decided not to factor in the data from China and Japan, as both these countries have extremely different economic conditions compared to the U.S. and the EU. Furthermore, China and Japan are further along in the economic trend; more of the inhabitants in these countries already own e-bikes, and thus have a stabilized e-bike market and will naturally see relatively less growth. Conversely, in the U.S. and EU, e-bikes still function as an emerging technology, which means that markets are still adopting and growing much faster in comparison.

1.5 Results

Graphical representations of the exponential regressions for the adjusted U.S. model and EU model are is shown in Figure 1.1 and 1.2:

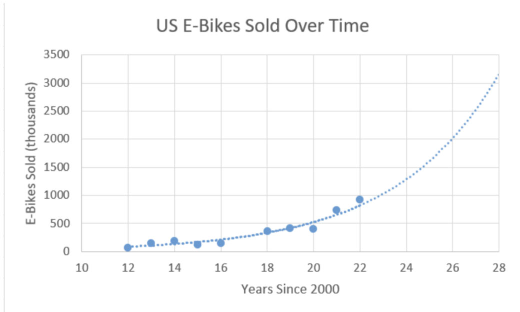

Figure 1.1: U.S. E-bikes Sold vs. Time (years since 2000, adjusted)

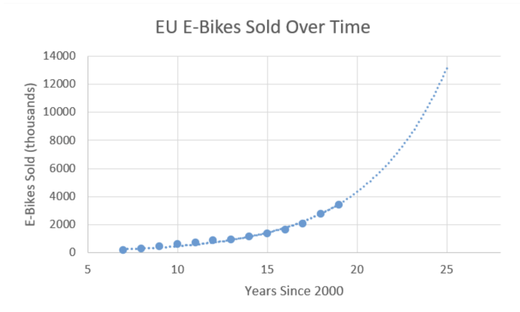

Figure 1.2: EU E-Bikes Sold vs. Time (years since 2000)

Using our model, the following table shows the predicted volume of e-bike sales in the U.S. for the years 2025 and 2028:

Year	**Predicted volume of e-bike sales in the U.S. (in thousands)**
2025	1560.08
2028	3045.72

1.6 Model Revision

Initially, we created a model using only e-bike sales data for the United States. However, given the high variability of the U.S. data points, we tempered the American data using the European e-bike sale data provided to create a more comprehensive and accurate model. After ensuring that the United States and European e-bike markets experienced similar rates of exponential growth over the past 15 years, we incorporated the European data to reduce the variability in our model and form a stronger estimate.

After initially attempting a linear regression model, we switched to an exponential regression upon realizing that technology grows in an exponential manner. Once the exponential model was adopted, we found more accurate results that better correlated with the data.

1.7 Discussion

Our model predicted that there would be a significant increase in the total sales of e-bikes in the U.S. within the next two years and even greater growth within the next five years. Given that e-bikes are an emerging technology in U.S. markets, this predicted increase is expected, as it would follow trends in other countries when e-bikes started to become more useful and accessible. The utility of e-bikes as an alternative to public transportation and automobiles has increased as urban centers increase in number and urban populations grow larger through natural population growth.

Strengths: Our model is cost-effective and can easily be scaled to map other countries if more data becomes available. The variability of our model for the U.S. is also reduced using EU data.

Weaknesses: Since our model only varies with respect to time, we cannot take into account external factors that cannot be predicted by time, such as fluctuations in the economy, global recessions, or pandemics. Developments like these would significantly impact e-bike sales, but we would not be able to predict them with our model. We also failed to account for potential major technological advancements that might make production or distribution of e-bikes or alternative substitutes cheaper, affecting supply and market volume. Furthermore, our model does not take into account natural limits of growth through the implementation of a logarithmic model. As such, we do not account for the market's natural carrying capacity and growth cap. Accordingly, our model should not be used to project extensively far into the future. If we had more time, we would have liked to conduct more research to estimate this threshold to better adjust our model.

1.8 Sensitivity Analysis

The part of our equation that had the largest impact on the model was the coefficient of the exponent. If we had run the un-adjusted U.S. model without tempering the equation with the EU regression, we would have seen an increase from 1560.08 thousand e-bike sales to 1611.61 thousand e-bike sales in 2025 and an increase from 3045.72 thousand e-bike sales to 3158.62 thousand e-bike sales in 2028.

1.9 Technical Computing

Conducting an exponential regression manually would be very time-consuming and inaccurate, so using a computer program to conduct the regression was justified.

We used the Excel polynomial regression function to initially generate our equations for the U.S. and EU models.

In addition, we created code as an easy way to solve the model, which can be found in the appendix.

2 Q2: Shifting Gears

2.1 Defining the Problem

In this problem, we were tasked with identifying potential underlying factors that contributed to e-bike growth. We elected to analyze income levels, gas prices, and lithium-ion battery prices as possible contributing factors. We then evaluated the significance of each factor on the growth of e-bike usage.

2.2 Assumptions

1. **The instrumental variables do not have a causal effect on the response variables: educational attainment, the cost of crude oil, and the number of engineering PhDs awarded yearly do not cause changes in e-bike sales.**
   Justification: There is no direct logical relation to be drawn between any of the listed instrumental variables with e-bike sales. None of the variables pose any direct relation to e-bike sales; all effects are a by-product of a confounding variable.

2. **The instrumental variables do have a causal effect on the predictor variables: educational attainment, the cost of crude oil, and the number of engineering PhDs awarded yearly cause changes in income, gas prices, and cost of batteries respectively.**
   Justification: By definition of the instrumental variable analysis model, the instrumental variables must be assumed to have a causal effect on the predictor variables in order to establish a causal relationship between the predictor variables and the response variables.

3. **The instrumental variables do not have any causal effects on any confounding variables: educational attainment, the cost of crude oil, and the number of engineering PhDs awarded yearly have no significant effect on environmental awareness, health and fitness, perceived "coolness," or any other potential underlying cause behind e-bike sales growth.**
   Justification: There is no direct effect of the listed instrumental variables on other potential underlying causes. None of the variables impact outside causes; the correlation of the instrumental variables can then be isolated to their respective treatment variables.

2.3 Variables

Symbol	Definition	Units
E	Educational attainment level in terms of the percentage of U.S. adults older than 25 years with a bachelor's degree	Percentage
I_r	Real disposable personal income per capita by year	U.S. dollars
I_p	Predicted disposable personal income per capita by year	U.S. dollars
S_I	E-bike sales in the U.S. for a given I_p	Number of e-bikes sold
C	Average annual OPEC crude oil price	U.S dollars per gallon
G_r	Real yearly average price of regular-grade gasoline in the U.S.	U.S. dollars per gallon
G_p	Predicted yearly average price of regular-grade gasoline in the U.S.	U.S. dollars per gallon
S_G	E-bike sales in the U.S. for a given G_p	Number of e-bikes sold
A	Number of yearly engineering PhDs awarded in the U.S.	Number of degrees
B_r	Real yearly cost of a lithium-ion battery pack in the U.S.	U.S. dollars per kilowatt of usable energy
B_p	Predicted yearly cost of a lithium-ion battery pack in the U.S.	U.S. dollars per kilowatt of usable energy
S_B	E-bike sales in the U.S. for a given B_p	Number of e-bikes sold

2.4 The Model

Seeking to understand the underlying causes of the growth in e-bike sales, we identified and analyzed 3 factors that may have contributed: disposable personal income per capita, price of gasoline, and cost of lithium-ion battery packs.

While the data provided allows for the establishment of correlations between certain factors and e-bike sale growth, there is no way to prove a causal relationship. We decided to use instrumental variable regressions for causal inference, also known as two-stage least squares regressions, in order to demonstrate causation. This technique is useful for understanding whether or not some predictor variable affects a response variable. Oftentimes, there are other variables that affect the relationship between predictor and response; instrumental variable regression allows for the isolation of the predictor-response relationship.

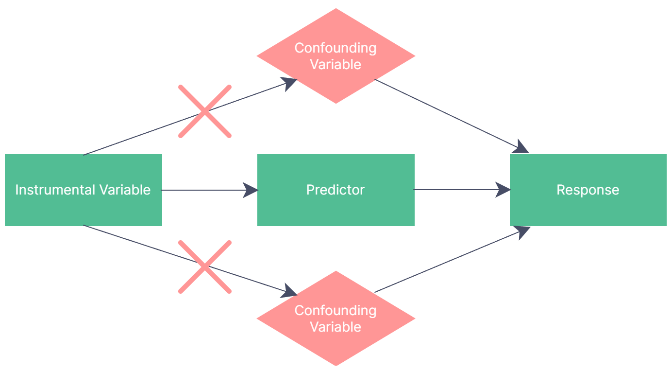

Figure 2.1: A visual representation of the logic behind instrumental variable regression

The two-stage regression model first utilizes the following generalized function in which f represents the predictor variable and x represents the instrumental variable:

$$f = \alpha * x + b \tag{6}$$

Using the function f generated by the first regression, predicted values for the predictor variable are calculated along the least squares regression line between f and x. Then, using the predicted values for the predictor and the true values for the response variable, a second regression is performed to yield the following generalized function in which y represents the response variable and z represents the projected predictor variable:

$$y = \beta * z + c \tag{7}$$

The regression coefficient β accurately captures the effect of the predictor variable on the response variable since the instrumental variable isolates the impact. As a third variable uncorrelated with the response, the assumption of independent assignment for the instrumental variable is key. Since we solely use the instrumental variable to calculate the projected values for the predictor variable, and because the instrument is not correlated with the response, any significant correlation in the second stage regression can be attributed to the predictor.

In the analysis of disposable personal income per capita, we predicted that rising income levels in the U.S. (the predictor variable) contributed to the accompanying rise in e-bike sales (the response variable). We identified educational attainment levels, measured by the percentage of U.S. adults older than 25 years with a bachelor's degree, as the instrumental variable in this relationship (see Figure 2.2). Educational attainment is highly correlated with income($R^2 = 0.9403$) since higher education helps people earn higher salaries, but has no direct effect on e-bike usage or any other potential confounding factor (U.S. Census Bureau).

In the analysis of gas prices, we predicted that inconsistent gas prices in the U.S. (the predictor variable) did not contribute to the accompanying rise in e-bike sales (the response variable). We identified crude oil prices, measured by the average annual OPEC crude oil price, as the instrumental variable in this relationship (see Figure 2.2). Crude oil price is highly correlated with gas price ($R^2 = 0.9403$) since crude oil is an ingredient in gasoline, but has no direct effect on e-bike usage or any other potential confounding factor (Aizarani).

In the analysis of lithium-ion battery costs, we predicted that decreasing battery costs in the U.S. (the predictor variable) contributed to the accompanying rise in e-bike sales (the response variable). We identified the number of yearly engineering PhDs awarded in the U.S. as the instrumental variable in this relationship (see Figure 2.2). Engineering PhDs are highly correlated with battery costs ($R^2 = 0.9005$) since greater STEM research leads to greater development and cheaper, more efficient technology, but has no direct effect on e-bike usage or any other potential confounding factor (Sabastian, Kang).

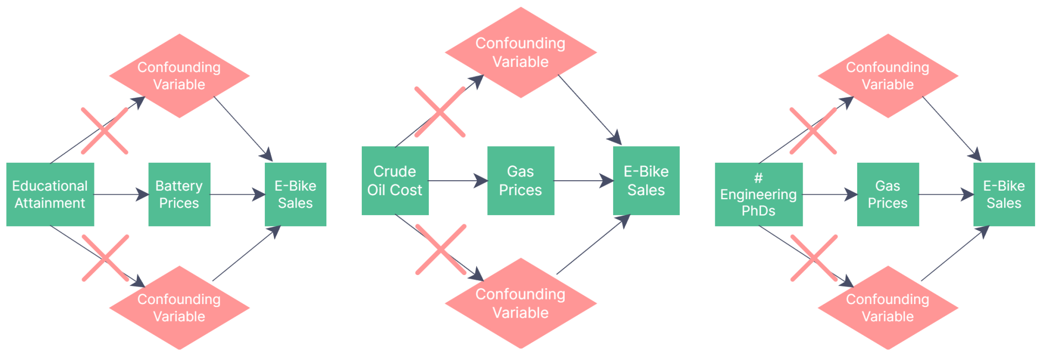

Figure 2.2: A visual representation of the instrumental variable regression pathways for the three variables analyzed

2.5 Results

For each of the three predictor variables investigated, we derived two linear regression models: one establishing the correlation between the instrumental and predictor variables; another establishing the correlation between the predictor and response variables.

Linear regressions for real income vs. educational attainment (Figure 2.3) and e-bike sales vs. predicted income (Figure 2.4) were performed and resulted in the following equations. The equations and their graphical representations are shown below:

Real personal income as a function of educational attainment

$$I_r = 1258.2x - 49.443 \tag{8}$$

$$R^2 = 0.969 \tag{9}$$

U.S. e-bike sales as a function of personal income as predicted by Equation 8.

$$S_I = 0.0601 I_p - 2280.2 \tag{10}$$

$$R^2 = 0.8313 \tag{11}$$

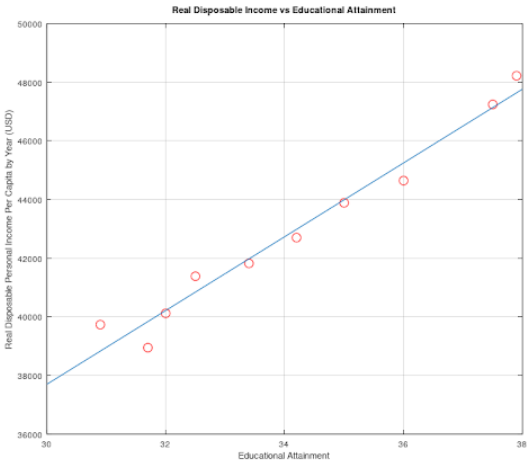

Figure 2.3: Real Disposable Income vs. Education Attainment

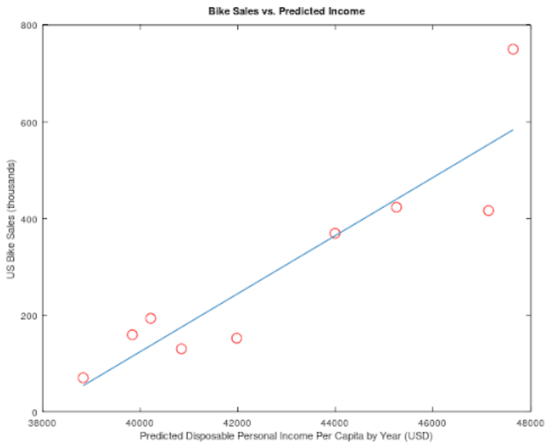

Figure 2.4: E-bike Sales vs. Predicted Disposable Income

Linear regressions for real gas prices vs. crude oil prices (Figure 2.5) and e-bike sales vs. predicted gas prices (Figure 2.6) were performed and resulted in the following equations. The equations and their graphical representations are shown below:

Real gas prices as a function of crude oil prices

$$G_r = 0.0229C + 1.2283 \tag{12}$$

$$R^2 = 0.9403 \tag{13}$$

U.S. e-bike sales as a function of gas prices as predicted by Equation 12.

$$S_G = -0.0015 G_p + 3.267 \tag{14}$$

$$R^2 = 0.2582 \tag{15}$$

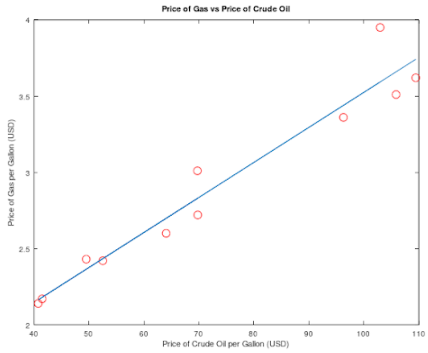

Figure 2.5: Gas Prices vs. Crude Oil Prices

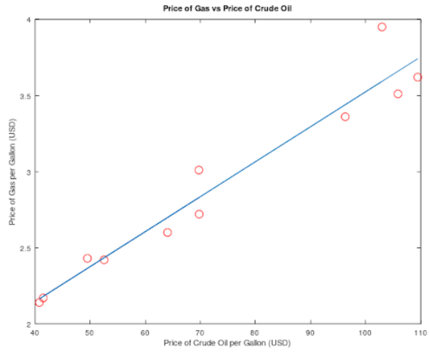

Figure 2.6: E-bike Sales vs. Gas Prices

Linear regressions for real lithium-ion battery prices vs. engineering PhDs awarded (Figure 2.7) and e-bike sales vs. predicted gas prices (Figure 2.8) were performed and resulted in the following equations. The equations and their graphical representations are shown below:

Real lithium-ion battery prices as a function of engineering PhDs awarded

$$B_r = -0.1034A + 1233.3 \tag{16}$$

$$R^2 = 0.9005 \tag{17}$$

U.S. e-bike sales as a function of lithium-ion battery prices as predicted by Equation 16.

$$S_B = -1.8967B_p + 789.29 \tag{18}$$

$$R^2 = 0.4885 \tag{19}$$

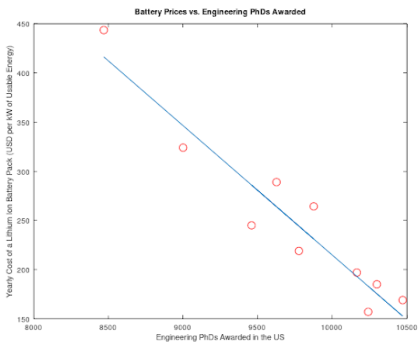

Figure 2.7: Lithium-ion Battery Prices vs Engineering PhDs Awarded

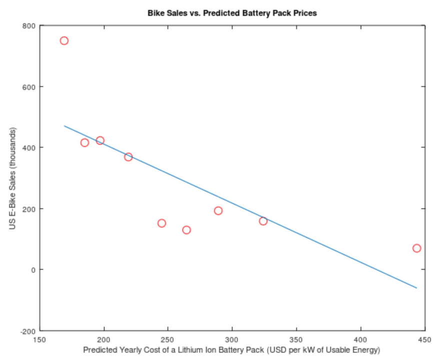

Figure 2.8: E-bike Sales vs Lithium-ion Battery Prices

The $R^2 = 0.8313$ value for predicted income vs. e-bike usage indicates a strong positive relationship. The $R^2 = 0.2582$ value for predicted gas prices vs. e-bike usage indicates a weak or no causal relationship. The $R^2 = 0.4885$ value for predicted lithium-ion battery costs vs. e-bike usage indicates a moderate negative relationship. These results establish income, battery costs, and gas prices (in that order) as the most to least important contributing factors analyzed.

2.6 Model Revision

When we first approached the problem, we attempted a linear regression to determine the correlation between the possible contributing factors and e-bike sales. We planned to analyze the coefficient of determination R^2 to measure the proportion of the variance in e-bike sales that can be explained by each contributing variable in the regression model. However, team discussion led us to realize that while the R^2 describes the strength of the relationship, it reveals nothing about the causality. As a result, we pivoted to the instrumental variable regression method described above.

2.7 Discussion

Our model proves that there is a strong positive causal relationship between real household income levels, moderate negative causal relationship with the cost of lithium-ion batteries, and no causal relationship with the cost of gas. These three results are consistent: if e-bikes are primarily purchased by consumers with more disposable income, then small shifts in prices caused by their batteries will not sway many more people away from or towards buying them. Prices will have some impact, but income variability is far greater than bike price variability. Furthermore, if those with less disposable income are not capable of buying e-bikes, increasing gas prices will not push them to change their form of transportation. For these individuals, driving and gas are necessities, so the demand for them is rather inflexible. They cannot afford to make large lump-sum purchases and they also cannot afford to go without transportation for the time needed to save the money to purchase an e-bike.

Strengths: Our model uses instrumental variable regression to demonstrate causality instead of relying solely on strong correlations. It also is able to demonstrate that three predictive variables had varying levels of contribution to the growth in e-bike sales.

Weaknesses: Since the regressions are compounded over multiple levels of analysis, the model is sensitive to any errors made in the early stages of calculation. If any of the assumptions, particularly those made about the instrumental variables, are false, the model will no longer hold true. It does not contain redundancies to mitigate inaccurate data. Our data is also only drawn from relatively small data sets because there are not many available statistics on the e-bike industry in the United States. Due to the emerging nature of the e-bike industry, there are few years during which inventory has been taken on the sales of e-bike sales. This means that even though we can compute strong regressions for the instrument and the predictor, they lose some of their value when compared against the more inconclusive response data.

2.8 Sensitivity Analysis

One key assumption that was made in our model was that the instrumental variable had a causal effect on the predictor variable. If this was not held true, then we would no longer be able to infer a causation relationship between the predictor variable and the response variable. Instead, the relationship would revert back to a correlation.

Another key assumption made was that the instrumental variables did not have a causal relationship with either the response variables or the confounding variables. If this was not held true, the true effect of the predictor variable on the response variable would be indeterminate due to the fact that the impacts of the instrumental variable on the response variable through the confounding variables would be unknown. Hence, we would no longer be able to determine the causational relationship between the predictor variables and response variables.

In summary, the three major assumptions that we made had significant implications on our results. While slight complications in our assumptions may not lead to drastic consequences in our results, they would definitely decrease the reliability in our solutions. However, given that we were tasked to determine causation between two variables through an observational study, the instrumental variable analysis model was the most feasible and accurate.

2.9 Technical Computing

Finding the line of best fit manually is an intensive and tedious process that consumes time and results in inaccuracies. Therefore, determining the least squares regression lines merited the use of a computer.

We used the built-in MATLAB "polyfit" function to create our linear extrapolations for each regression: educational attainment vs. real income, predicted income vs. e-bike sales, crude oil price vs. real gas price, predicted gas price vs. e-bike sales, engineering PhDs awarded vs. real lithium-ion battery cost, and predicted lithium-ion battery cost vs. e-bike sales. The data for all instrument and predictor variables were inputted into separate arrays, which were later fitted with a linear model. Once the stage one linear regressions were performed, the least squares regression line was used to create a set of projected predictor variable values. These projected values were then inputted into individual arrays alongside the response variable data, after which the stage two regressions were performed.

The process was repeated 6 times to encompass all regressions.

3 Q3: Off the Chain

3.1 Defining the Problem

In this problem, we were tasked with modeling potential impacts of rising e-bike usage on various related factors. We evaluated the impacts on the following: carbon emissions and traffic congestion.

3.2 Assumptions

1. **A linear relationship between the dependent and independent variables**

   For this model, we use a multivariate linear regression. This automatically assumes a linear relationship between the variables. We chose this for simplicity's sake - it would not be feasible to make a multivariate nonlinear linear regression given the constraints with time and data. If the data is found to be nonlinearly correlated, the user will need to transform the data using statistical software.

2. **The independent variables are not highly correlated with each other**

   This is a requirement of multivariate linear regression. When independent variables show multicollinearity, there will be problems figuring out the specific variable that contributes to the variance in the dependent variable. See discussion.

3. **All variables suggested in the model explain the outcome**

   For the coefficients to be the most accurate, all major variables must be incorporated in the regression. We suggest that the user consult relevant literature to make the most accurate choices of what to include in the model. The differences between direct and indirect variables must be considered - if you include one indirect variable, you cannot also include its intermediate direct variable (assumption 2), so you must include all subsequent indirect variables.

3.3 Variables

General Variables

Symbol	**Definition**	**Units**
G	General function of "impact"	N/A
x_n	Variables G is directly dependent upon	N/A
B	e-bike sales	number of e-bikes
β_n	Slope coefficients for each variable x_n	N/A
ϵ	Model Error	N/A

Specific Variables

Symbol	Definition	Units
E	Carbon Emissions	grams CO2
C	Commute Time	minutes
V	Total Gasoline Vehicles	Vehicles
A	Emissions from Agriculture	grams CO2
F	Emissions from electric power production	grams CO2
T	Emissions from other transportation sources	grams CO2
c	Commute time (dependent variable)	time
I	Car Accidents	car accidents
J	Annual Precipitation	inches

3.4 The Model

For this problem, we chose to use a multivariate linear regression. This is to find the equation for the general function for any "impact" (ex. carbon emissions):

$$G(x_1, x_2, x_3, ...x_n) \tag{20}$$

This defines the general function of a given impact in terms of all of the variables that have significant effect on it. To choose variables, the user of the model must thing both logically and evaluate literature relevant to G. However, it is very important to note the differences between direct and indirect relationships of the variables. Due to the requirement of a multivariate linear regression to have the independent variables not highly correlated with each other (assumption 2), using an indirect variable means you cannot also use the direct "intermediate" variable in the function and vis versa. At the same time, every major impact variable should be included (assumption 3), so all of a counterpart indirect variables must be included.

Usually, therefore, only direct variables should be considered. However, one of the independent variables must be e-bike sales, as we are ultimately trying to find the partial change in G with respect to e-bike sales (shown below).

$$B = x_i$$

Therefore, if it is indirect in the situation of the problem, all other indirect variables that impact its subsequent intermediate variable should be included in the model.

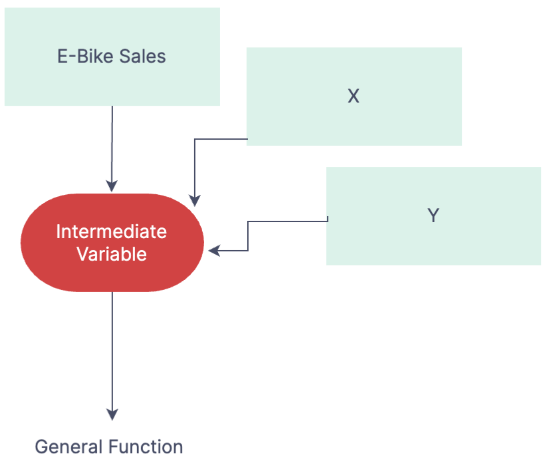

Figure 3.1: Indirect Variables

Multivariate linear regression gives the following equation for G:

$$G = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...\beta_n x_n + \epsilon \tag{21}$$

The coefficients β_1 through β_n can be determined using a software or by a Monte Carlo simulation. We decided upon a Monte Carlo simulation in order to estimate these coefficients through numerous random sampling. By doing so, we derive an equation in which the coefficients β_1 through β_n solve the equation in a way that minimizes the error on G. Below is a matrix which displays how the Monte Carlo simulation worked (see Figure):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ 1 & x_{31} & x_{32} & \dots & x_{3q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Figure 3.2: System of Matrices for the Monte Carlo Simulation

As stated before, we are ultimately trying to find the partial change in some general impact G with respect to the change in e-bike sales, which by the definition of a partial derivative is:

$$\frac{\partial G}{\partial x_n} = \beta_n$$

$$\therefore \quad \frac{\partial G}{\partial B} = \beta_i \tag{22}$$

Overall, the impact due to e-bike sales is the coefficient β_i of the e-bike variable x_i.

3.5 Results

To demonstrate this model, we chose three important impacts that e-bikes sales will have. First is Carbon Emissions E. We began with e-bike sales B, which we determined indirectly impacted carbon emissions through its change to the amount people drive. As stated before, we then also included the other variables that impacted this, such as commute time C and total gasoline vehicle use V. Other major variables we found through the EPA - agriculture emissions A, electricity production F, industrial production Q, residential power T, and and other transportation T. Therefore, we have:

$$E(B, C, V, A, F, T) \tag{23}$$

To construct our regression, we found data in each of the areas from 2006 to 2019. Data on total energy consumption was taken from the EPA. Data from e-bike sales was the same as Q1. Data for commute time was collected from the Census Bureau. Data from total gasoline vehicle use was collected from Statista. Data from agriculture emissions, electrical energy, industrial production, household power, and other transportation was collected from the EPA.

When we implemented this data into the Monte Carlo simulation it gave us the following coefficients for the equation:

$$E = 10558.8134 - 0.1099B + 499.0037C + 27.293V + 6.8324A + 532.2105F + 8.3208T + \epsilon$$

Since we are only concerned with the coefficient of B because the partial derivative with respect to B. Overall we get the change in carbon emissions due to e-bike sales:

$$\frac{\partial E}{\partial B} = -0.1099 tons$$

This means that the change in the total carbon emissions due to is approximately a decrease 0.1099 tons for every e-bike sold, with an r-squared of 0.9863.

The second area to demonstrate our model was in traffic congestion c measured by time on the road (census bureau) from 2012 to 2019. E-bike sales is a direct variable of this, as time spent on an e-bike is less spent on the road. Other variables we identified were car accidents I and poor weather J, with data from IIHS and Weather Underground, respectively.

Plugging this in, the data gives us the function:

$$c(B, I, J) = 0.0032B + 0.0002I + 0.0014J + 19.6 + \epsilon$$

Overall, we get the change in traffic congestion with respect to e-bike sales:

$$\frac{\partial c}{\partial B} = 0.0032$$

This means that for every 1 e-bike, the average commute time across the U.S. fell by 0.192 seconds (0.0032 minutes), with an r-squared of 0.9651.

3.6 Model Revision

This model underwent the most change during our work. We started with a general function of what we were trying to find, the partial of G with respect to e-bike sales. We thought this function could be described by multiple variables, some of which in turn could be described by e-bike sales as a variable. Therefore, we could use the chain rule to find our ultimate answer. We wrongly assumed we could do a simple regression of two variables to find each partial, however, so we scratched that idea and instead tried to use the methodology from Q2 to make a function based off a completely independent variable. Again, this logic was flawed as it did not give us an accurate function. We finally turned to multi variable linear regression. Within this model, we changed the variables we used for examples several times, based off of overall effect, availability of data, and more.

3.7 Discussion

Strengths Overall, this is a particularly important section as it explains many of the real-world effects of the increased sales of e-bikes. This model has many uses, from business executives in the travel industry making company decisions to environmental scientists. As this model is further developed, it can prove particularly insightful for how the world feels the effects of e-bike transitions. Furthermore, the r-squared values of our regressions were very high, and the model overall seemed to make sense. It is also all-encompassing to any general impact G that the user would want to implement.

Weaknesses To improve this model in the future, we realized we may want to address steps to account for if B was both a direct and indirect variable. All that would change is when taking the partial the user would end with $\frac{\partial E}{\partial B} = \beta_{i1} + \beta_{i2}$. Furthermore, the linear assumption is not ideal because clearly not all of the relationships would necessarily be linear. A more in depth use of the model would include translations where necessary. Also, this tells us little about how these might change with major events (technological improvement, economic disaster, etc.), which is a weakness of most mathematical models.

We would also like to point out that we made an assumption that commute time is not effected by e-bike sales, but then later say it is. We made this assumption originally because we were not able to find completely independent variables that effected the same variable as the e-bike sales. In the future we will consider possibly performing another iteration of the multivariate linear regression, and then use the chain rule of partial derivatives to obtain the solution.

For assumption 2, the user can use Variance Inflation Factor method to check if they satisfy the assumption.

3.8 Sensitivity Analysis

Each of our assumptions were essential for this question. Any changes to them provide a large logical fallacy to our final result. In fact, a multivariate linear regression cannot be performed without meeting these necessities. In most cases, what we assume should be true, and with much statistics testing we could prove it. The few assumptions it did change were reasonable considering their nature (i.e. not having enough data), and should not have an immense impact on the overall result. The largest impact would be from missing a variable entirely - especially if it had one of the largest effects on the outcome in the end. This is why it is imperative that the user looks at literature in the area to make sure they implement the most important variables and obtain the most accurate final result.

4 Conclusion

We provided an estimate for the projected total volume of e-bike sales in the U.S in 2025 and 2028 utilizing a model generated using an exponential regression. The initial raw model from the exponential regression for the US was potentially flawed due to economic fluctuations and the COVID-19 pandemic. We adjusted this model using a corresponding model generated from data from e-bike sales in the EU. This new adjusted model was used to project our final estimate of e-bike sales in 2025 and 2028.

Next, we identified potential underlying factors that could have contributed to the growth of e-bike sales. In order to demonstrate causation with the data provided, we used instrumental variable regressions for causal inference, also known as two-stage least squares regressions. The implementation of instrumental variable regression allowed for the isolation of the predictor-response relationship. Conducting linear regressions on the data for each of the factors allowed us to determine the existence and strength of the causal relation between each variable and e-bike sales.

Finally, we developed a method to predict the impact of the increase in e-bike usage on various related factors. We implemented a multivariate linear regression to find a general equation to model the impact of e-bikes on any given factor. In doing so, we were able to isolate the partial derivative of a given general impact with respect to the change in e-bike sales.

In the face of growing environmental awareness and an increasing interest in sustainable alternatives to emission-heavy transportation sources, e-bikes have emerged as a popular choice for cheaper, faster, and more accessible transportation. Properly estimating their new role in the transportation market and their impacts on related factors is crucial, as based on our data, the demand for e-bikes will only continue to increase. By gaining a greater understanding of the factors underlying these changes and their effects on our daily lives, policymakers can properly take advantage of and facilitate these shifts to ensure a smooth transition from existing transportation trends to more sustainable practices.

5 References

"A New Era - WFSGI." WSFGI Magazine, https://wfsgi.org/wp-content/uploads/2021/03/MAG2021_NEWERA.pdf.

Aizarani, Jessica. ""OPEC Crude Oil Price Statistics Annually 1960- 2022."" Statista, 3 Feb. 2023, https://www.statista.com/statistics/262858/change-in-opec-crude-oil-prices-since-1960/.

BEA Interactive Data Application, https://apps.bea.gov/iTable/ reqid=19&step=3&isuri=1&select_all_years=0&nipa_table_list=58&series=a&first_year=2000&last_year=2022&scale=-9&categories=survey&thetable=.

Bureau, US Census."Census Bureau Releases New Educational Attainment Data." Census.gov, 24 Feb. 2022, https://www.census.gov/newsroom/press-releases/2022/educational-attainment.html

Carlier, Mathilde. "U.S. Vehicle Sales 1976-2021." Statista, 14 July 2022, https://www.statista.com/statistics/199983/us-vehicle-sales-since-1951/.

E-Bikes - U.S. Sales 2016." Statista, Statista, 12 Oct. 2022, https://www.statista.com/statistics/326124/us-sales-of-electric-bicycles/.

eBicycles. "Useful Facts & Stats of e-Bikes [for 2023] + Infographic." EBicycles, 2 Feb. 2023, https://www.ebicycles.com/ebike-facts-statistics/.

EPA. "Sources of Greenhouse Gas Emissions." EPA, Environmental Protection Agency, https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions.

Kang, Kelly, and Jacqueline Falenheim. "Data Tables." NSF, https://ncses.nsf.gov/pubs/nsf23300/data-tables.

"North America E-Bike Market Analysis - Industry Report - Trends, Size & Share." North America E-Bike Market Analysis - Industry Report - Trends, Size & Share, https://www.mordorintelligence.com/industry-reports/north-america-e-bike-market.

"Petroleum & Other Liquids." U.S. Regular All Formulations Retail Gasoline Prices (Dollars per Gallon),https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPMR_PTE_NUS_DPG&f=A.

Sabastian, Xavier. "Price of Lithium-Ion Battery Pack through 2010 - 2021." Way Blog, 29 Mar. 2022, https://www.way.com/blog/lithium-ion-battery-pack-price-through-2010-2021/.

Statista Research Department. "E-Bikes - U.S. Sales 2016." Statista, 12 Oct. 2022, https://www.statista.com/statistics/326124/us-sales-of-electric-bicycles/.

"Traffic Volume Trends December 2021 - Policy: Federal Highway Administration." Traffic Volume

Trends December 2021 - Policy | Federal Highway Administration, https://www.fhwa.dot.gov/policyinformation/travel$_monitoring/21dectvt/$.

6 Code Appendix

Problem One

```
    //years of data
int[] yearsUS = new int[] {12, 13, 14, 15, 16, 18, 20, 21, 22};
//thousands of bikes sold
int[] USBikes = new int[] {70, 159, 193, 130, 152, 369, 423, 416,750, 928};
//correlation
int[] ratioUS = new int[9];

for (int i = 0; i < yearsUS.length; i++) {
ratioUS[i] = USBikes[i]/yearsUS[i];
}
```

Problem Two

```
»

 eduAttain = [30.9
31.7
32
32.5
33.4
34.2
35
36
37.5
37.9]
eduAttain =

   30.900
   31.700
   32.000
   32.500
   33.400
   34.200
   35.000
   36.000
   37.500
   37.900

>> rIncome = [39732
38947
40118
41383
41821
42699
43886
44644
47241
48219]
rIncome =

   39732
   38947
```

```
   40118
   41383
   41821
   42699
   43886
   44644
   47241
   48219

>> plot(eduAttain, realIncome)
error: 'realIncome' undefined near line 1, column 17
>> plot(eduAttain, rIncome)
>> title(Real Disposable Income vs Educational Attainment)
error: parse error:

  syntax error

>>> title(Real Disposable Income vs Educational Attainment)
                        ^
>> title("Real Disposable Income vs Educational Attainment")
>> xlabel("Educational Attainment")
>> ylabel("Real Diposable Income")
>> LinearRegression(eduAttain, rIncome)
error: 'LinearRegression' undefined near line 1, column 1
>> p = polyfit(eduAttain, realIncome, 1)
error: 'realIncome' undefined near line 1, column 24
>> p = polyfit(eduAttain, rIncome, 1)
p =

   1258.236    -49.443

>> plot(p)
>> plot(x, y)
error: 'x' undefined near line 1, column 6
>> plot(eduAttain, rIncome, "or")
>> plot(eduAttain, rIncome, "or", p)
>> plot(eduAttain, rIncome, "or")
>> p = ployfit(rIncome, eduAttain, 1)
error: 'ployfit' undefined near line 1, column 5
>> p = polyfit(rIncome, eduAttain, 1)
p =

   7.7009e-04   1.0969e+00

>> plot(eduAttain, rIncome, "or", p)
>>  plot(eduAttain, rIncome, "or")
>> plot(eduAttain, rIncome, "or", t, 1258.238*t-49.443)
error: 't' undefined near line 1, column 32
>> t = [30 31 32 33 34 35 36 37 38]
t =

   30   31   32   33   34   35   36   37   38

>>  plot(eduAttain, rIncome, "or", t, 1258.238*t-49.443)
```

```
>> title("Real Disposable Income vs Educational Attainment")
>> xlabel("Educational Attainment")
>> ylabel("Real Income")
>> ylabel("Real Disposable Personal Income Per Capita by Year (USD)")
>> ;
>> pIncome = [38828.937
39835.497
40212.957
40842.057
41974.437
43987.6
45245.8
47133.1
47636.3
]
pIncome =

   3.8829e+04
   3.9835e+04
   4.0213e+04
   4.0842e+04
   4.1974e+04
   4.3988e+04
   4.5246e+04
   4.7133e+04
   4.7636e+04

>> bSales = [70
159
193
130
152
369
423
416
750
]
bSales =

    70
   159
   193
   130
   152
   369
   423
   416
   750

>> p2 = polyfit(pIncome, bSales, 1)
p2 =

   6.0109e-02  -2.2802e+03
```

```
>> t2 = bSales
t2 =

    70
   159
   193
   130
   152
   369
   423
   416
   750

>> plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
>> t2 = pIncome
t2 =

   3.8829e+04
   3.9835e+04
   4.0213e+04
   4.0842e+04
   4.1974e+04
   4.3988e+04
   4.5246e+04
   4.7133e+04
   4.7636e+04

>> plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
>> title("Bike Sales vs. Predicted Income")
>> xlabel("Predicted Income")
>> ylable("Bike Sales")
error: 'ylable' undefined near line 1, column 1
>> ylabel("Bike Sales");
>> crudeOil = [109.45
105.87
96.29
49.49
40.76
52.51
69.78
64.04
41.47
69.72
102.97
]
crudeOil =

   109.450
   105.870
    96.290
    49.490
    40.760
    52.510
    69.780
```

```
    64.040
    41.470
    69.720
   102.970

>> gasPrices = [3.62
3.51
3.36
2.43
2.14
2.42
2.72
2.60
2.17
3.01
3.95
]
gasPrices =

   3.6200
   3.5100
   3.3600
   2.4300
   2.1400
   2.4200
   2.7200
   2.6000
   2.1700
   3.0100
   3.9500

>> p3 = polyfit(crudeOil, gasPrices, 1)
p3 =

   0.022968   1.227447

>> plot(crudeOil, gasPrices, "or", crudeOil,  0.022968*crudeOil +  1.227447)
>> title("Price of Crude Oil vs. Gas Prices")
>> xlabel("Price of Crude Oil per Gallon (USD)")
>> ylabel("Price of Gas per Gallon (USD")"
error: parse error:

  syntax error

>>> ylabel("Price of Gas per Gallon (USD")"
                                         ^
>> ylabel("Price of Gas per Gallon (USD")
>>  plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
>> >> title("Bike Sales vs. Predicted Income")
error: parse error:

  syntax error

>>> >> title("Bike Sales vs. Predicted Income")
```

```
    ^
>> >> xlabel("Predicted Income") plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
error: parse error:

  syntax error

>>> >> xlabel("Predicted Income") plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
    ^
>> >> title("Bike Sales vs. Predicted Income")
error: parse error:

  syntax error

>>> >> title("Bike Sales vs. Predicted Income")
    ^
>> >> xlabel("Predicted Income")
error: parse error:

  syntax error

>>> >> xlabel("Predicted Income")
    ^
>>  plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
>> >> title("Bike Sales vs. Predicted Income")
error: parse error:

  syntax error

>>> >> title("Bike Sales vs. Predicted Income")
    ^
>> plot(pIncome, bSales, "or", t2, 6.0109e-02*t2-2.2802e+03)
>> title("Bike Sales vs. Predicted Income")
>> xlabel("Predicted Disposable Personal Income Per Capita by Year (USD)")
>> ylabel("US Bike Sales (thousands)")
>> plot(crudeOil, gasPrices, "or", crudeOil,  0.022968*crudeOil +  1.227447)
>> title("Price of Gas vs Price of Crude Oil")
>> xlabel("Price of Crude Oil per Gallon (USD)")
>> ylabel("Price of Gas per Gallon (USD)")
>> pGasPrices = [
>> 3.734705
ans = 3.7347
>> 3.652723
ans = 3.6527
>> 3.433341
ans = 3.4333
>> 2.361621
ans = 2.3616
>> 2.161704
ans = 2.1617
>> 2.430779
ans = 2.4308
>> 2.826262
ans = 2.8263
>> 2.694816
```

```
ans = 2.6948
>> 2.177963
ans = 2.1780
>> pGasPrices = [3.734705
3.652723
3.433341
2.361621
2.161704
2.430779
2.826262
2.694816
2.177963
]
pGasPrices =

   3.7347
   3.6527
   3.4333
   2.3616
   2.1617
   2.4308
   2.8263
   2.6948
   2.1780

>> polyfit(pGasPrices, bSales, 1)
ans =

  -174.92   790.89

>> plot(pGasPrices, bSales, "or", pGasPrices, -174.92*pGasPrices + 790.89)
>> title("Bike Sales vs. Predicted Gas Prices")
>> xlabel("Predicted Gas Prices per Gallon (USD)")
>> ylabel("US Bike Sales (thousands)")
>> phDs = [8469
9000
9626
9875
9459
9776
10164
10298
10471
10240
]
phDs =

    8469
    9000
    9626
    9875
    9459
    9776
   10164
```

```
   10298
   10471
   10240

>> batteryP = [443.6
324
289
264.3
245
219
197
185
169
157
]
batteryP =

   443.60
   324.00
   289.00
   264.30
   245.00
   219.00
   197.00
   185.00
   169.00
   157.00

>> polyfit(pHds, batteryP, 1)
error: 'pHds' undefined near line 1, column 9
>> polyfit(phDs, batteryP, 1)
ans =

  -1.3170e-01   1.5317e+03

>> plot(phDs, batteryP, "or", phDs, -1.3170e-01*phDs+1.5317e+03)
>> title("Battery Prices vs. Engineering PhDs Awarded")
>> xlabel("Engineering PhDs Awarded in the US")
>> ylabel("Yearly Cost of a Lithium Ion Battery Pack (USD per kW of Usable Energy)")
>> polyfit(-1.3170e-01*phDs+1.5317e+03, bSales, 1)
error: polyfit: X and Y must have the same number of points
error: called from
    polyfit at line 116 column 5
>> predpHds(443.6
324
error: parse error:

  syntax error

>>> 324
      ^
>> 289
ans = 289
>> 264.3
```

```
ans = 264.30
>> 245
ans = 245
>> 219
ans = 219
>> 197
ans = 197
>> 185
ans = 185
>> 169
ans = 169
>> ppHds[443.6
error: parse error:

  syntax error

>>> ppHds[443.6
         ^
>> 324
ans = 324
>> 289
ans = 289
>> 264.3
ans = 264.30
>> 245
ans = 245
>> 219
ans = 219
>> 197
ans = 197
>> 185
ans = 185
>> 169
ans = 169
>> ppHds = [443.6
324
289
264.3
245
219
197
185
169
]
ppHds =

   443.60
   324.00
   289.00
   264.30
   245.00
   219.00
   197.00
   185.00
```

```
   169.00

>> polyfit(ppHds, bSales, 1)
ans =

    -1.9354   798.0983

>> plot(ppHds, bSales, "or", ppHds, -1.9354*ppHds + 798.0983)
>> title("Bike Sales vs. Predicted Battery Pack Prices")
>> "Predicted Yearly Cost of a Lithium Ion Battery Pack (USD per kW of Usable Energy)")
error: parse error:

  syntax error

>>> "Predicted Yearly Cost of a Lithium Ion Battery Pack (USD per kW of Usable Energy)")
                                                                                     ^
>> ylabel("Predicted Yearly Cost of a Lithium Ion Battery Pack (USD per kW of Usable Energy)")
>> xlabel("Predicted Yearly Cost of a Lithium Ion Battery Pack (USD per kW of Usable Energy)")
>> ylabel("US E-Bike Sales (thousands)")
>>
```

```
public class MathModeling {

    public static void main(String[] args) {

        //Initializing Instrumental Variables
        double educationalAttainment = 1.0;
        double crudeOilPrices = 1.0;
        double numberOfEngineeringPhDs = 1.0;

        //First-Stage Linear Regression Equations
        double realIncome = 1258.2*educationalAttainment - 49.443;
        double gasPrices = 0.0229*crudeOilPrices + 1.2283;
        double batteryPrices = -0.1034*numberOfEngineeringPhDs + 1233.3;

        //Second-Stage Linear Regression Equations
        double eBikeSalesIncome = 0.0601*realIncome - 2280.2;
        double eBikeSalesGas = -0.0015*gasPrices + 3.267;
        double eBikeSalesBattery = -1.8967*batteryPrices +789.29;

        // print the results
        System.out.println("realIncome: " + realIncome);
        System.out.println("gasPrices: " + gasPrices);
        System.out.println("batteryPrices: " + batteryPrices);
        System.out.println("eBikeSalesIncome: " + eBikeSalesIncome);
        System.out.println("eBikeSalesGas: " + eBikeSalesGas);
        System.out.println("eBikeSalesBattery: " + eBikeSalesBattery);
    }

}
```

Figure 6.1: Problem 2 - Solving Linear Regression Equations Code