# Applications of Multiple Linear Regression in Social Medias Related Marketing

Wenxi Zhang\*

Woodside Priory School, Portola Valley, CA, United States of America

#### \*Corresponding author: 200318@yzpc.edu.cn

**Abstract.** In a market, what the sellers care the most is their budgets and their sales. They need to know how to make more money. In order to do this, they need to find ways to compare different plans and the results they get from them. For selling products, advertisement is indeed an important part. This paper compares three different social medias, trying to decide which is the best way to invest for advertisement: newspaper, facebook, or youtube. By examining the regression statistics, this paper shows how to use multiple linear regression to make the best marketing decisions when it comes to advertising. Particularly in this case, the preferable order of spending budgets on the social medias for advertisement is facebook or youtube, and lastly newspaper. It also compares single and multiple linear regression, and it is found that the latter is more important. This paper provides people with helpful information to sell their products properly.

Keywords: Multiple Linear regression; Social medias; Marketing.

# 1. Introduction

When it comes to making marketing decisions, the business people usually compare different methods to see which is the most profitable or efficient. These decisions should all be related to the money they make eventually, whether it is their sales, products, or even time consumed. Linear regression is a really good way to find relationship between variables, so this is the major focus. Single linear regression is the most straightforward way to find those relationships, but it is limited to one variable. Thus, the author wants to know how to use multiple linear regression to make the best marketing decision. In real world, actual business situations, it is almost impossible to have only one variable. In most cases, there are a lot of different means of achieving their goals, so people need to consider all the factors and compare them. Therefore, multiple linear regression is needed. It is more complicated, but also will be a useful tool. In this research, the author is trying to examine the relationship between the budget spend on advertisement for three social medias, that is youtube, facebook, and newspaper; their relationship with the total sales [1].

Why do people use linear regression analysis, not other graphs or plots? The author would argue that this is because regressions can clearly show the relationship between independent and depend variables, which is very useful in different aspects of economy. "Regressions are used to quantify the relationship between one variable and the other variables that are thought to explain it; regressions can also identify how close and well determined the relationship is". [2] For example, in the research, the author examines the relationship between the budgets putting into advertisement in three different social medias with the total sales. In this case, the total sales can be explained by the different budgets putting into the social medias. Then, one can examine this relationship, and decide which set of data---which way of spending the budget on advertisement, will give the most sales, that is, the best market decision.

By computing the experimental data, people can get numerical results that explains all the relationships. They can then generate formulas that can be used to predict other situations which can help them make the best choices. In short, regressions analysis can help economists to value assets and make assumption, and help investors to understand the relationship between factors such as price and sales, which is the foundation of earning money [3].

### 2. Methods and Theory

The author will introduce the data set chosen, explain the reason of using regression analysis, and the similarities and differences with other people who uses single linear regression text.

#### 2.1. Data

Table 1 are some example sets of the variables from the entire data, which is the advertising experiment between social media budgets and sales in thousand dollars. There are a total of 171 data sets [1], the table below is just the very first few sets. This paper will show the preferable order of investing budgets in three different social medias and explain the reasons.

Youtube	Facebook	Newspaper	Sales
84.72	19.2	48.96	12.60
351.48	33.96	51.84	25.68
135.48	20.88	46.32	14.28
116.64	1.8	36	11.52
318.72	24	0.36	20.88
114.84	1.68	8.88	11.40
348.84	4.92	10.2	15.36
320.28	52.56	6	30.48

 Table 1. Advertising experiment between Social Media Budget and Sales (in Thousands \$) [1]

#### 2.2. Single vs. Multiple Linear Regression

Instead of single, why would people choose to use multiple linear regression, that is more complicated? It is true that people can directly see the relationship between the two, when only having one depend and one independent variable [3]. There is a positive strong relationship between them. However, in most situations, especially in the real world, there are more factors to consider. For example, when a seller wants to know the relationship between a product and its demand, price is not the only determinant. There are other factors that could impact the result, such as consumer preferences and number of sellers. Therefore, people need to introduce more variables, and use multiple linear regression.[4]

One of its main advantages is that it can better represent the real-world phenomena by including multiple independent variables. People can account for more factors that influence the dependent variable, and reduce the error and bias in the estimates [5]. In a similar research paper, Paik also use multiple linear regression model to address his problem, the influential factors on consumption [6]. From this the author can see that he includes multiple variables, such as oil price and tax rate. Indeed, it is impossible, in the real life, to only have one factor that impacts consumption. The methodology is similar with his, where the author includes three different social media budgets, and find the relationship between them and the sales.

The author used the formula to evaluate outputs

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \epsilon.$$
 (1)

In this research, Y represents the output, which is the sales,  $\beta_0$  is the intercept,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are the corresponding coefficients for each of the variables. The variable  $X_1$ ,  $X_2$ , and  $X_3$  represent the social medias Youtube, Facebook, and Newspaper, respectively.  $\epsilon$  is the error term but it is not included in this research [4].

#### 3. Results and Application

The author used 171 sets of result from advertising experiment between three different types of social media budgets and sales (see table 1) to examine their relationships [1].

#### 3.1. Individual Plots and guesses

Instead of seeing to the numerical outputs directly, the author firstly generates plots separately for each of the social medias. Based on these plots, their individual relationships with the sales can be clearly seen. The Fig. 1 shows there is a positive strong linear relationship between youtube advertising budgets and sales. Similar with Facebook, shown also in Fig. 1, the relation is weaker comparing with youtube. On the other hand, for newspaper, there seem to be a weak, or maybe not linear relationship with sales. The data points are really scattered in Fig. 2.





From these plots, the guess for the linear relationship strength for the three social medias with the sales is (from strongest to the weakest): Youtube, Facebook, Newspaper. And the author will see if the numerical outputs match this guess.



Fig. 2 The sale of Newspaper at different advertising budgets.

#### 3.2. Results

The author used excel to generate the numerical outputs. It is found that value of Multiple R is 0.94895095, value of R square is 0.90050791, and value of adjusted R square is 0.89870986. In addition, the standard error is 2.01283467. The coefficients, standard error, and other quantities are shown in Table 2.

The upper table tell people about the media budget as a whole and its relationship with the sales. The author mainly focuses on the multiple R and R square. It is called multiple R here since here are three different variables, instead of only one in single linear regression. The author is looking at the relationship of all three social medias. R represent three correlations between independent variables and dependent variable. In this research it is the budgets of the social medias and the sales. It is a number between -1 to 1, where close to -1 indicates a strong negative correlation, one variable increase and the other one decrease, and close to 1 indicates a strong positive correlation, both x and y decrease or increase together. If the R is close to zero, the dependent variable and independent variable may have a weak or no linear relationship [7]. From the table, it is clear that the multiple R is very close to 1 (0.949), meaning that the social media budgets have a positive strong relationship with the sales, which is what the author expected.

media	coefficients	Standard error	T stat	P-value
Intercept	3.48498971	0.40128124	8.68465636	3.4408E-15
Youtube	0.04531485	0.00152266	29.7602253	1.9098E-68
Facebook	0.1887033	0.00903935	20.8757582	2.7223E-68
Newspaper	0.00398915	0.00652334	0.61151962	0.54169219

Table 2. The coefficients, standard error, and other quantities.

The R square, also known as the coefficient of determination, explains the variations of the dependent variable that is caused by the independent variable. It is a number between 0 and 1, since it is squared from R [7]. The author can see that from the results, where a R of 0.94895095 result in a R square of 0.90050791. R square is usually state as a percentage in interpretations. For example, in this research, this R square means that, about 90.1% of the changes of the budgets for advertisement of the three social medias is accounted for by the sales [7].

The author also generated a formula using the intercept and coefficients. This function can be used to predict sales when investing differently for each social media, and it is given by

$$Sales = 3.485 + 0.045YT + 0.189FB + 0.004NS$$
(2)

Moreover, the author also examined the individual relationship of the three social media and the sales by looking at their P-values and coefficients. In short, a smaller p-value gives a more convincing coefficient. From the table, both youtube and facebook has a small p-value, which means their coefficient should be reliable, that they won't be affected easily. At the meantime, facebook has a higher coefficient, so investor would get more sales if they put more budgets into facebook advertising. However, Newspaper has a high p-value, higher than the common ones the author uses, such as 0.05, 0.01, and 0.1. This implies that it will be easily influenced by other factors, that it will be risky to invest on. Therefore, based on this observation, the preferable order of spending advertising budget should be: facebook, youtube, newspaper.

This result also matches the guess, the individual plots above. The youtube and facebook graph all have a stronger positive linear relationship than the newspaper plot, which is more scattered. And between the youtube and facebook, facebook plots are more scattered. The newspaper vs. sales graph also explains that it could easily be impacted by other factors, since its relationship with sales is not that strong. Closer research can also be done, such as seeing other non-linear relationships of the three social medias with sales.

#### **3.3.** Applications

From the results, this research can also give ideas for other similar business problems. From the preferable orders, a clear trend of using technology is shown. People prefer using their technologies, thus the probability of seeing the advertisement from youtube and facebook is higher than newspaper. People can then assume that there are less people seeing the paper advertisement than see them online. If similar decisions are needed to make when sellers are trying to advertise their products, they can choose to use the internet, other than actual copies, or printing them on magazines. Also, facebook works more effective than youtube. This gives an idea for where and towards which population people should put their advertisement, and think about the pros and cons. For example, in general, there is a huge population and age range for people who uses youtube, but the author would say more younger people. However, much young people now do not use facebook. So, in this research, maybe the product is more targeting to middle-aged people, so facebook advertisement is more effective. Also, there is a lot of advertisement on youtube, which could be more competitive. Sellers can decide where to advertise based on their target consumers.

Other researchers also examine the importance of multiple linear regression in economics. For example, in another research, the authors chose a fuel distributing company, and considered seven factors that might impact the clients' satisfaction [8]. These factors are considered

independent variables, and the client's spending monetary value is the dependent variable. From their result, the authors find which factors impact the clients' satisfaction the most. Similar with the case considered, they can then decide how to improve the company, in order to increase the consumers; satisfaction, and thus their spending value [9]. The purpose of this paper, similar with the current one, is also to illustrate how to apply and interpret the linear multiple regression analysis within a marketing research based on primary, marketing theory.

# 4. Conclusion

In this research, the author found the best way to spend advertising budget based on the relationship between different social media and sales. The author used multiple linear regression analysis, individual plots and generated numerical outputs. However, there are still several limitations for this method. Firstly, unlike single linear regression, it is hard or even impossible to create a graph for multiple linear regression. For one or two variables, people can still generate 2D or 3D graphs, but what about 4D and more? Even if it is possible to do so, the graph would be hard to interpret by human eyes. It is true that people can just use technology, but the purpose of seeing a clear graph is lost. Second, the result can be hard to interpret when having multiple variables. When having many independent variables or complex interactions, people need to check linearity, normality, independence, with different tests and plots. This process can be very time-consuming and complicated. Lastly, Multiple regression can be overfitting, which is when the model fits the data too well, and it failed to generalize to new data. Overfitting can occur when having too many independent variables, or when the variables are highly correlated with each others. Overall, there are still many places of improvement when using multiple linear regression analysis.

# References

- [1] Nimon, Kim F., Frederick L. Oswald. Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients. Organizational Research Methods, 2013, 16(4), 650–74.
- [2] Taras Shevchenko National University of Kyiv, Yu. Braychevskyy. The use of regression analysis in human-geographic studies: the case of electoral behavior analysis. Ekonomichna ta Sotsialna Geografiya, 2014, 70: 66–72.
- [3] Casson, Robert J., and Lachlan Dm Farmer. Understanding and Checking the Assumptions of Linear Regression: A Primer for Medical Researchers. Clinical & Experimental Ophthalmology, 2014, 42(6): 590–96.
- [4] Pervez Asif and Irfan Ali. Robust Regression Analysis in Analyzing Financial Performance of Public Sector Banks: A Case Study of India. Annals of Data Science, 2024, 11(2): 677–91.
- [5] Anita Mfuh Y. Lukong and Yahaya Jafaru. Covid-19 Pandemic Challenges, Coping Strategies And Resilience Among Healthcare Workers: A Multiple Linear Regression Analysis. African Journal of Health, Nursing and Midwifery, 2021, 4(1): 1-18.
- [6] Paik, Songyi. The U.S. Consumption Analysis: Using a Linear Regression Model. Proceedings of GREAT Day, 2016, 2015(1): 21.
- [7] Bolin, Jocelyn H. Regression analysis in R: a comprehensive view for the social sciences. First edition, CRC Press, Taylor & Francis Group, 2023.
- [8] Luigi Dumirescu, Oana Stanciu, MihaiTichindelean, and Simona Vinerean. Additional contact information. Studies in Business and Economics, 2012, 7(2): 94-109
- [9] Tibshirani, R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267–288.