

Investigation for Smartfood's Market Ability Based on Regression Analysis

LI Runshi

Beihang University, Beijing, China

SHEN Tzuying

Shanghai Soong Ching Ling School International Division, Shanghai, China

TING Tzuyun

Shanghai High School International Division, Shanghai, China

Product Development at Smartfood is considering adding a new low-carbohydrate food product labeled K-Pack to their product line. Hence, the company needs to estimate the product's profit potential and choose the best marketing strategy. In order to investigate the different factors' impact on sales, we establish a multiple regression model. In our model, we consider selling price, advertisement payment, and selling location as independent variables and the sale is a dependent variable. Based on the empirical test, we found that the sales of the product increase as the selling price and the ads investment increase, and locations near bakery sections lead to higher sales. We also found that the final profits increase as the selling price and the ads investment increase. As a result of the analysis made with the model, we provide the company with a full-scale estimation of sales and suggestions on their marketing strategy: higher price, higher advertisement payment, and selling locations near the stores' bakery sections. Compared with the original prediction made by the company's staff, the methods followed are more accurate and provide suggestions more specifically.

Keywords: multiple regression, analysis model, marketing strategy, advertisement payment

Introduction

The Smartfood Company considers launching a new product labeled K-pack, a kind of low carbohydrate diet food. Through a small-scale marketing test, the director estimated that the annual sales are 750,000 cases per year, and the net margin is 1.55 million dollars, with a price of 50 cents per package (12 dollars per case) and an advertising cost of 3 million dollars. To ensure the result of the marketing test, a four-month market test study, including numerous variables, is conducted. We need to analyze the data collected and build a model to calculate the estimated total sales and profits.

Based on the market test, we figured out many factors that would affect the sales, such as price, location,

These authors contributed equally.

LI Runshi, Beihang University, Beijing 100000, China.

SHEN Tzuying, Shanghai Soong Ching Ling School International Division, Shanghai 200000, China.

TING Tzuyun, Shanghai High School International Division, Shanghai 200000, China.

Correspondence concerning this article should be addressed to SHEN Tzuying, Hangtian Second Apartment No. 3 1001, Qingzhou North Rd. 939, Xuhui District, Shanghai 200000, China.

advertisement, store volume, calculated the annual sales through the four-month sales, and computed the profit with diverse conditions. Then we try to establish models to evaluate the different factors' impact in this study.

Multiple linear regression model is a statistical technique used to determine the relationships between the independent variables and the dependent variables. It is widely used in marketing to predict a business's future path, and within the regression model, researchers have to resolve many complicate issues (Mishra & Min, 2013). Appropriate regression diagnostics, estimation, and test procedures are developed to suit the models and achieve a favorable outcome (Fisher & Lee, 1992). Winship and Radbill (2016) emphasize the importance of the biased and inconsistent result or error caused by insufficient sampling weight. They state that the standard errors will be calculated incorrectly and prefer the unbiased unweighted OLS estimates (Winship & Radbill, 2016). Moreover, suitable statistical software and proficient operation are required to access and validate the postulated model (Hoang & Ngo, 2014). A correlation between the factors and the outcomes can be clarified and leads to a general idea of the marketing environment (Fisher & Lee, 1992).

Net margin is a value that shows the efficiency of a business and how well the multiple regression analysis has been done. It's influenced by many numerous factors, including both quantitative and qualitative variables. The sale earnings, merchandise costs are quantitative variables affecting the net margin. It's possible to increase the sales amount while decreasing your net margin. When marketing needs are different, the marketing mix must be altered to accommodate customers' demands (Peter & Olson, 1998). A study shows that the price reductions of 10%, 25%, and 50% on lower fat snacks will increase sales by 9%, 39%, and 93%, respectively, compared with usual price conditions (French, 2003). While some of the factors are controllable, many qualitative factors, such as leadership, training programs for employees, and marketing strategies, are variables you can't control at all. Marketing strategies are built based on an environmental analysis (Douglas & Craig, 2016), consumer behavior, and regional marketing target (Peter & Olson, 1998). Lone, Dan, Levi, Chan, & Bianco-Simeral (2010) suggest that healthy food is only attractive to two groups of people and proposed two marketing methods. It's recommended to sell special emphasis on shaping brand image, marketing mix, and product labeling (Mitić & Gligorijević, 2012). According to Kearney (2010), a pattern that most of the population is experiencing a dietary shift and nutritional transition occurs. It's suggested to consider agricultural and healthy sectors as musts (Kearney, 2010). Later in this paper, we will determine the influencing factors of sales and maximize our profits to reach higher efficiency.

This paper discusses the relationship between the total sales or total profits (or net margin) of the product to be sold and the independent variables that influence the marketing strategy of Smartfood Company. The other is to provide the company's board of directors with the most profitable marketing mix and the profits potential of the new product. The researching method within this passage is to first process the data collected in the marketing test and then build a regression model according to the data. Finally, we will build another model to calculate the total profits under different circumstances and decide which advice to provide for the company. As for the conclusions, we found that the sales and profits are likely to have a positive near linear relationship with price and advertising investments, and the best marketing mix strategy for the company is 70 cents price, 3.5 million dollars advertising investment, and positions nearby bakery sections. Under such marketing strategy, the total profit is 3.19 million dollars, which is much higher than the predicted 1.55 million dollars.

There are four main parts of the article. The background part introduces the problem the whole article is going to discuss. The method part shows the process we made to analyze the problem. The analysis result part

displays all results we get from our analysis, and the discussion part shows the conclusion we made from all these results.

Background

The Smart Food Company is now considering introducing a new product called K-Pack. The CEO of the company, Mr. Donovan, has chosen 215 stores in different cities to sell. However, he is still confusing with the marketing mix of K-Pack and does not know their future plan and possible achievement. K-pack is a kind of low carbohydrate diet food, which looks like a typical candy bar. Since most snack products have short life cycles, the CEO of Smartfood wonders the demand and marketing mix of K-pack and has conducted a small pre-test market study with Mary's marketing director. The initial estimate of annual sales is 750,000 cases of K-pack (24 packages in a pack), the net margin is 1.55 million dollars, with a price of 50 cents per package (12 dollars per case) and an advertising cost of 3 million dollars. Mr. Donovan was satisfied with the pre-test result and asked the director to conduct a four-month test market study including numerous variables to ensure the market conditions in the future are reliable. This paper will use a multiple regression model and function to examine the affecting sales factors, including price, advertising investment, and location. We will analyze the relationship between these variables and the sales, then find out a function based on the three variables. We will use the function to estimate the highest profit available for K-Pack and provide the best way for Smartfood Company to increase their net margin.

Method

According to the result of the marketing test, there are a couple of variables affecting sales of K-pack, which are prices, locations, advertisements, city index, store volume. On account of K-pack being considered to sell throughout the nation, we did not take the city index as a variable. The same for store volume should not be taken as a variable because the number and volume of stores already decided by the CEO are not allowed to change. Through building a regression model with diverse variables, we know that price is the factor that affects sales the most. Instead, location is the last one. Additionally, calculation of annual sales would be available. In this way, we can get the total profit by combining the pricing of each product.

Profits Calculation

Table 1 represents the frequency distribution of store volume, the scale of the grocery stores: a total of 215 stores dispersing across the countries. Table 2 shows the market test, including the testing city, price, location, advertisement, and store volume through a four-month test.

Table 1

*Number of Grocery Stores by Store Volume**

Store volume (\$00,000)	Frequency (No. of stores)
40-45	20
45-50	35
50-55	50
55-60	50
60-65	35
65-70	25
Total	215

Table 2

Test Market Results

Average sales per month, store (cases)				Price (cents)	Advertising (1 if high, 0 otherwise)	Location (1 if bakery, 0 otherwise)	Store volume (\$00,000)	City index
First month	Second month	Third month	Fourth month					
S1	S2	S3	S4	P	A	L	V	C
170	255	244	229	50	0	0	45	1
213	239	339	242	50	0	0	45	2
195	223	275	264	50	1	0	43	3
241	193	293	336	50	1	0	45	4
179	242	199	309	50	0	1	49	1
115	255	305	197	50	0	1	44	2
190	343	364	364	50	1	1	49	3
317	355	407	360	50	1	1	54	4
95	238	269	353	60	0	0	54	1
157	153	160	371	60	0	0	50	2
205	243	216	349	60	1	0	44	3
156	202	321	334	60	1	0	49	4
160	242	259	193	60	0	1	41	1
160	217	210	269	60	0	1	54	2
65	248	300	282	60	1	1	41	3
210	299	322	265	60	1	1	47	4
41	199	235	135	70	0	0	49	1
107	298	352	250	70	0	0	58	2
98	216	255	296	70	1	0	40	3
169	291	395	266	70	1	0	59	4
101	275	252	222	70	0	1	53	1
212	257	190	203	70	0	1	52	2
123	415	232	313	70	1	1	43	3
194	221	287	317	70	1	1	57	4

Since there is an uncertain proportion between prices and sales, we cannot determine the maximum profit through Table 2. Thus, to figure out the maximum profit and the best condition for selling, calculating the profit takes a significant place.

The model equation of sales equals to:

$$\text{Profit} = \beta_1 \times \text{Price} + \beta_2 \times \text{Advertising} + \beta_3 \times \text{Location} + \beta_4 \times \text{Volume} \quad (1)$$

Regression Model

We use the model mentioned above to build a regression model and range of the independent variables: Price is equal to 50, 60, or 70 cents; Advertising and location are dummy variables, store volumes are between 40 and 70. Though building different models by removing each variable, we could impact various variables on profit.

Through building a four-regression model with a different equation, the relationship between independent variables and the dependent variable is known clearly. By removing each variable and observing the data of R square, we know which factor is the most important. Profit always is the dependent variable in each equation below.

$$\text{Profit} = \beta_1 \times \text{Price} + \beta_2 \times \text{Advertising} + \beta_3 \times \text{Location} \quad (2)$$

By removing store volume, we get Equation (2). Price, advertisement, and location are the independent variables.

$$\text{Profit} = \beta_1 \times \text{Price} + \beta_3 \times \text{Location} + \beta_4 \times \text{Volume} \tag{3}$$

After removing the advertisement, we get Equation (3). Price, location, and store volume are the independent factors.

$$\text{Profit} = \beta_2 \times \text{Advertising} + \beta_3 \times \text{Location} + \beta_4 \times \text{Volume} \tag{4}$$

Remove price and get Equation (4). Take advertising, location, and store volume as the independent variables.

$$\text{Profit} = \beta_1 \times \text{Price} + \beta_2 \times \text{Advertising} + \beta_4 \times \text{Volume} \tag{5}$$

By removing location, we get Equation (5). Price, advertisement, and store volume are independent variables.

Then we input our sales data into these four-regression models. A series of results would be obtained in our next section.

Analysis Report

In this part of the article, we display the results of some specific statistics and the parameters of our linear regression model. The first part is about statistics of the revenue of different individual stores. The second part shows the regression results. The third part calculates different total profits under different circumstances and informs about the best market mix.

Revenues

We have calculated the Total sales and Revenues of each shop that has been surveyed in the market test. We can find out from the table that the total revenue goes higher obviously when we increase the price or the advertising investment. The detailed result is shown as follows (Table 3).

Table 3
Result of the Marketing Test and Total Revenue

Price (cents)	Advertising (1 if high, 0 otherwise)	Location (1 if bakery section, 0 otherwise)	Store volume (\$00,000)	City index	Total sales for four months (cases)	Total revenue for four months (dollars)
50	0	0	45	1	898	449
50	0	0	45	2	1033	516.5
50	1	0	43	3	957	478.5
50	1	0	45	4	1063	531.5
50	0	1	49	1	929	464.5
50	0	1	44	2	872	436
50	1	1	49	3	1261	630.5
50	1	1	54	4	1439	719.5
60	0	0	54	1	955	573
60	0	0	50	2	841	504.6
60	1	0	44	3	1013	607.8
60	1	0	49	4	1013	607.8
60	0	1	41	1	854	512.4
60	0	1	54	2	856	513.6
60	1	1	41	3	895	537
60	1	1	47	4	1,096	657.6

Table 3 to be continued

70	0	0	49	1	610	427
70	0	0	58	2	1,007	704.9
70	1	0	40	3	865	605.5
70	1	0	59	4	1,121	784.7
70	0	1	53	1	850	595
70	0	1	52	2	862	603.4
70	1	1	43	3	1,083	758.1
70	1	1	57	4	1,019	713.3

According to Table 3, we can see that both total sales and total revenue go up as price goes up, advertising goes up, and store volume goes up as well.

Regression Results

Based on the above regression model, we capture the different variables' impact on sales. The related results are summarized in Table 4.

Table 4

Regression Chart of Sales per Store

	Only price (1)	Only price and ads (2)	Price, ads and location (3)	Price, ads, location and volume (4)
β_0	1363* (5.741)	1269* (6.640)	1242* (6.458)	718.1* (3.536)
β_1	-6.469* (-1.650)	-6.469* (-2.069)	-6.469* (-2.073)	-9.882* (-3.834)
β_2		188.2* (3.685)	188.2* (3.693)	216.5* (5.294)
β_3			53.33* (1.047)	49.64* (1.258)
β_4				14.76* (3.792)
Adj- R^2	0.1101	0.4596	0.4877	0.7083

Note: t statistics in parentheses. * $p < 0.05$.

According to the results shown in Table 4, we find that model (1) has a very low significance level (R squared) (0.1101), so it's completely not reliable. There is an obvious increase in R squared (to 0.4596) when we add advertisement investment as an independent variable. So we are sure that advertisement investment is a significant factor in sales. R squared doesn't increase much (to 0.4877) as we include location into the independent variables, so location is not an important factor in the sales. Similarly, we can find that store volume is a significant factor as well.

According to our regression analysis, we found that price, ads, and volume are crucial variables that significantly influence the sales of the product in a store because they have a significant influence on the significance level (R squared). At the same time, location seems to have little influence on it. When we put all the variables in it, the formula we get is:

$$Sales = \beta_1 \times Price + \beta_2 \times Advertisements + \beta_3 \times Location + \beta_4 \times Volume + \beta_0 \quad (6)$$

$$\beta_1 = -9.882, \beta_2 = 216.5, \beta_3 = 49.64, \beta_4 = 14.76, \beta_0 = 718.1$$

Max Profits

The total profits are calculated through the formula given above (Table 5). There are six mixtures of the

independent variables, which are given in the chart. They are the price as 50, 60, or 70 cents, and advertisement investment as a dummy variable to be 0 or 1.

Table 5

Total Profits and Corresponding Strategies

Price (cents)	Ad (0 = 3 million USD, 1 = 3.5 million USD)	Profits (dollars)
50	0	1,201,046.495
50	1	1,734,176.664
60	0	1,803,101.176
60	1	2,570,779.816
70	0	2,191,002.688
70	1	3,193,229.798

According to the models above, the max profits is 3,193,229.798 dollars, under the condition of 70 cents price and 3.5 million dollars Advertisement investment. The minimum profit is 1,201,046.495 dollars, under the condition of 50 cents price and 3 million dollars Advertisement investment. Both price and advertisement investment are positive factors that influence the total profits. Compared with the original estimation made by the company's staff in which the total profits or net margin is 1.55 million dollars, we found that she overestimated the profits under the marketing mix of 50 cents price and 3 million advertisement investment. Still, the profiting potential of the product is much higher when the company raises the price to 70 cents and the advertisement investment to 3.5 million dollars.

Discussion

According to all analyses above, the price of each pack and the advertising investment have the greatest influence over the total sales of the product since when analyzing them separately; they have rather large R squared. This parameter shows the influence of this factor on sales. When analyzing the influence of the location in the shops by taking it as the only variable in the regression model, we found that the parameter R squared is only 0.02, so this variable seems to have very little influence over the total sales.

The data and calculation above show the max profits Smartfood Company can get if K-Pack is selling nationally at the 215 stores fixed by the CEO, which is 3,193,229.798 dollars. According to all analysis above, if the situation is under the fixed 215 stores, the price of each pack and the advertising investment have a large R squared comparing to location. This, in other words, means that there is a greater influence performed by the price and advertising investment on the total sales of the product. However, if the CEO has not fixed the stores that K-Pack will be selling in the future, it's recommended to check the influence of the city index and store volume. The city index and store volume have a larger effect greater than the price and the advertising investment through our models. When the city index and store volume are considered variables, R squared value reached 0.7, but when these two factors are eliminated, R squared value is about 0.4. Therefore, if the condition is allowed, it's suggested to check and reset the 215 stores in a better city to reach higher profit. Furthermore, the company's CEO can also conduct another test model that observes the market condition to help them determine better selling strategies.

We proposed some recommendations for the company:

- (1) Set the price to be 70 cents per package (14.4 dollars per case).
- (2) Set the advertising investment to be 3.5 million dollars.

(3) If possible, sell the products at the bakery sections of the stores.

(4) If possible, make further market tests to confirm the influence of all these variables since the confidence level is relatively low.

Conclusion

This paper investigates the relationship between the variables and sales. It mainly analyzes the future marketing strategy for Smart Food Company by constructing a multiple regression model and developing a function. We used the multiple regression model to find out the connection between various factors, including price, advertisement investment, selling location, and store volume, and sales. We used the function developed based on the result of the multiple regression models to calculate the highest profit available for the company.

We mainly got results on two aspects: the regression model and the profits. Firstly, the regression model provides us a method to estimate the sales of one store from a couple of factors. We also found that price, advertising investment, and store volume are all important positive factors to the sales, and location is a negligible factor. Then, about the profits, we found that according to all analysis above, the estimated profit potential of the initial plan is 1.20 million dollars, and the highest total profit potential predicted is 3.19 million dollars. The optimized marketing strategy is to sell the product at the price of 70 cents per package, spend 3.5 million dollars on advertising, and choose the location in bakery sections.

The deficiency of this paper includes mainly two aspects: the data and our calculation. First, the number of samples provided is too small to conduct an accurate and precise result. When the number of samples is less than 30, the *t*-value may be affected, affecting the final suggestion made. Furthermore, throughout the observation we've made, we find that the city index is an important factor of sales amount. In order to have a precise and lucrative suggestion for the company, it's recommended to do another model focusing on the habits people consume and what kinds of characters in a city are suitable for selling. For the second part, we fail to make a model for each variable, which may cause a lower accuracy net margin. In the future, we will make a model based on a larger number of samples, do research on consumerism, and analyze the affecting factors one by one to predict K-pack's future profits.

References

- Douglas, S. P., & Craig, S. C. (2016). *Global marketing strategy*. New York: McGraw-Hill.
- Fisher, N. I., & Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, 48(3), 665-677.
- French, S. A. (2003). Pricing effects on food choices. *Journal of Nutrition*, 133(3), 841S-843S.
- Hoang, T., & Ngo, D. (2014). The steps to follow in a multiple regression analysis: *SAS Global Forum 2012*.
- Kearney, J. (2010). Food consumption trends and drivers. *Philosophical Transactions of the Royal Society B Biological Sciences*, 365(1554), 2793-2807.
- Lone, T. A., Dan, P., Levi, A. E., Chan, K. K., & Bianco-Simeral, S. (2010). Marketing healthy food to the least interested consumers. *Journal of Foodservice*, 20(2), 90-99.
- Mishra, D. P., & Min, J. (2013). Analyzing the relationship between dependent and independent variables in marketing: A comparison of multiple regression with path analysis. *SSRN Electronic Journal*, 6(3), 113-120.
- Mitić, S., & Gligorijević, M. (2012). Global challenges and perspectives of marketing of healthy food products. *Marketing*, 43(3), 205-218.
- Peter, J. P., & Olson, J. C. (1998). *Consumer behavior and marketing strategy*. Dalian: Dongbei University of Finance & Economics Press.
- Winship, C., & Radbill, L. (2016). Sampling weights and regression analysis. *Sociological Methods & Research*, 23(2), 230-257.