Correlation Analysis Between Cereal Food Consumer Ratings and Nutritional Composition

Rui Luan[1,a,*]

[1]*School of Management Science and Engineering, Central University of Finance and Economics, Beijing, 102206, China*
a. 2020311463@email.cufe.edu.cn
*corresponding author

Abstract: Since contemporary information-retrieval systems rely heavily on the content of titles and abstracts to identify relevant articles in literature searches, great care should be taken in constructing both. With the popularization of food safety and health knowledge, consumers are increasingly concerned about the nutritional value of food. Research has shown that when consumers purchase food products, the nutritional composition table on the food packaging can affect their purchasing intention. Therefore, it is important to explore the impact of various factors in the ingredient table on consumer ratings and purchase intentions. The paper takes nutritional data of cereal products as a sample to investigate customer ratings of different products and conduct regression analysis and correlation research between ratings and nutritional components of cereal products. Based on the obtained regression equation, the paper predicts the rating of cereal products on the market. The paper finds that there is a correlation between consumer ratings and nutritional composition in the cereal product. Using regression equations to predict the test set, the prediction accuracy is good, which indicates that we can refer to the nutritional composition table of cereal products to roughly predict the consumer ratings of products in the market. Based on research, cereal manufacturers can attract consumers to purchase products by optimizing brand promotion strategies, enhancing brand image, and other means. Also, manufacturers can also choose high protein and low fat cereal raw materials to meet consumer demand and expand market consumption of cereal products.

Keywords: cereal products, consumer ratings, regression analysis, prediction

1.   Introduction

With the improvement of consumer economic level and the popularization of nutrition and health knowledge, more and more consumers are paying attention to the nutritional composition of food [1]. When purchasing food, consumers are no longer only satisfied with purchasing food that can satisfy their appetite but tend to purchase similar kinds of foods with higher nutritional value. Existing research of cereal products has shown that multiple factors such as food brand, safety attributes, nutritional composition, and taste collectively affect consumers' purchasing preferences and satisfaction with food [2]. In terms of nutritional composition, consumers have a strong intention to purchase certain foods with prominent nutritional components, such as high protein and low fat [3]. Based on existing research, the paper conducts research on the factor of nutritional composition, using

quantitative methods to comprehensively analyze the impact of various nutritional components on consumer ratings.

It can be found that there are many studies on the relationship between consumer preferences and nutritional composition. When conducting such research, some literature considers nutritional composition as one of the independent variables that affect consumer preferences, and uses multiple independent variables such as nutritional composition, brand image, and advertising to conduct correlation analysis on consumer preferences. On the basis of existing literature, the paper starts with the independent variable of nutritional composition and specifically analyzes the correlation between nutritional components such as protein and fat in the nutritional composition table and consumer preferences. Meanwhile, the paper quantifies consumer preferences in the form of ratings, facilitating correlation analysis and linear regression.The paper is based on the perspective of data analysis, using the cereal product dataset as a sample, and using regression analysis methods to analyze the correlation between various nutritional components such as calories, protein, trace elements, and consumer ratings. Based on regression analysis, the paper selects grain products on the market as the test set and uses regression analysis formulas to predict consumer ratings of products based on their nutritional composition.

By analyzing the relationship between consumer ratings and cereal nutritional composition, we can understand the impact of cereal nutritional composition on consumer purchasing preferences, thereby helping cereal manufacturers better conduct consumer positioning evaluations. Cereal manufacturers can upgrade and improve their products through research results, produce products that are more in line with consumer preferences, increase consumer ratings, and thereby enhancing market competitiveness and occupy more market share. The production improvement of cereal products often drives healthy competition in the grain food industry, achieving overall upgrading of grain products, thereby improving the overall level of the cereal product market and the nutritional value of consumers purchasing products, and achieving a win-win situation between the supply and reception ends.

2. Process of Experiment

2.1. Data Source

The paper selects 77 cereal products as samples from seven manufacturers: American Home Food Products, General Mills, Kelloggs, Nabisco, Post, Quaker Oats, and Ralston Purina. For each cereal product, the paper records the content of each product's nutrients such as calories, protein, fat medium, fiber, carbon, sugar, potato, vitamins, and quantifies them with specific content. The paper obtains the average rating of consumers on products through consumer reports, and quantifies their preference for cereal products by using consumer ratings, with a rating range of 0-100. The higher the rating, the higher the consumer's preference for the product.

2.2. Research Methods

Firstly, the paper uses Excel to carry out descriptive statistical analysis on the data, calculates the average score and calories of the whole sample. The paper also classifies the products according to the manufacturers, calculates the average rating and calories of different manufacturers' products, to get a general view of the sample and understand the Central tendency of product.

Subsequently, to find out the correlation relationship between nutritional components and consumer ratings, the paper uses SPSS for multiple regression analysis to obtain a regression equation between consumer ratings and nutritional composition and tests the fitting degree of the regression equation to consumer ratings. So, the paper divides the sample data into two parts, with 54 products for regression analysis and 23 products test and prediction, conduct correlation analysis between

consumer ratings and nutritional components, calculates Pearson correlation coefficient, selects components with strong correlation as independent variables, and conducts multiple regression analysis. Pearson correlation coefficient is used to measure the linear correlation between variables, and the value range is -1 to 1. The greater the absolute value of Pearson correlation coefficient, the greater the correlation between variables [4]. Then The paper also tries to analyze the obtained multiple regression equation, evaluate its R-square, p-value and F-statistics and adjust the selected independent variables to obtain the multiple regression equation that meets the requirements [5]. After that, the paper uses the obtained regression equation to test the sample data, calculates the MAE, MSE and MAPE of the test set to analyzes the prediction effect.

3. Research Results

3.1. Descriptive Analysis

By analyzing the average ratings of cereals from different manufacturers, we can also know the general impressions from different manufacturers on consumers. The ratings of most cereal products are between 18 and 54, with an overall average of approximately 43. The average ratings of cereals from American Home Food Products, Kellogg's, Quaker Oats and Nabisco are higher than the overall average, whereas other three scored lower than the overall average.

3.2. Regression Analysis

The paper uses SPSS to do Pearson correlation analysis on ratings and calories, proteins, fat, sodium, fiber, carbo, sugars, potass, vitamins to obtain the corresponding Pearson correlation coefficients. When the Pearson correlation coefficient is 0-0.2, it indicates extremely weak correlation or no correlation. Therefore, as shown in the table, there is no correlation between consumer ratings and the carbon and potass of nutritional components.

Table 1: Pearson Correlation Analysis.

Correlation Efficiency										
		calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	rating
rating		-.638**	.320*	-.417**	-.587**	.317*	.149	-.777**	.105	-.316*	1

After removing carbon and potassium, multiple linear regression is performed using calories, proteins, fat, sodium, fiber, sugar, vitamins as independent variables, and ratings as dependent variables. The significance test showed that the explanatory power of calories to the dependent variable was relatively low (P=0.538), so multiple regression was performed again by removing calories.

The multiple regression results of removing calories show that the adjusted judgment coefficient R^2 of the model is 0.96, close to 1, and has a high goodness of fit; The significance test of the equation has an F-statistic of 215.661, and the p-value corresponding to the F-statistic is<0.001, indicating strong significance. The significance test and multicollinearity test of independent variables both satisfy the hypothesis of multiple regression (significance<0.05 and VIF value<10).

Table 2: Multiple Regression Analysis.

		Sum of Squares	Free Degree	Mean Square	F	Significance	Adjusted R-square
	Regression	7956.367	6	1326.061	215.661	.000b	
	Residual	288.995	47	6.149			
	Sum	8245.361	53				

Table 3: Related indicators of multiple regression.

	Unstandardized Coefficient		Standardized Coefficient	Significance	Multicollinearity Statistics	
	B	Standard Error	Beta		Tolerance	VIF
constant	58.260	1.279		.000		
protein	1.515	.414	.126	.001	.627	1.595
fat	-4.410	.458	-.302	.000	.760	1.316
sodium	-.050	.004	-.355	.000	.799	1.251
fiber	2.458	.252	.312	.000	.729	1.371
sugars	-1.655	.086	-.602	.000	.766	1.306
vitamins	-.037	.014	-.077	.014	.835	1.198

Therefore, based on the above data, with consumer rating as the dependent variable and the above nutritional composition as the independent variable, a multiple regression equation for y (rating) on x_1(protein), x_2(fat), x_3(sodium), x_4(fiber), x_5(sugars), x_6(vitamins) established as follows:

$$y = 1.515x_1 - 4.410x_2 - 0.050x_3 + 2.458x_4 - 1.655x_5 - 0.037x_6 + 58.260$$

3.3. Prediction of Data

Table 4: Prediction and Residual.

	Rating	Predicted Rating	Residual	MAE	MSE	MAPE
1	68.40	67.14	1.27	1.55	3.98	0.039
2	33.98	31.68	2.30			
3	59.43	59.83	0.41			
4	93.70	90.81	2.90			
5	34.38	30.76	3.62			
6	29.51	29.68	0.17			
7	33.17	33.40	0.23			
8	37.04	34.24	2.80			
9	49.12	45.86	3.26			
10	53.31	55.40	2.08			
11	18.04	19.17	1.13			
12	50.76	46.37	4.40			
13	19.82	20.23	0.40			
14	40.40	39.39	1.01			
15	22.74	23.93	1.19			

Table 4: (continued).

16	41.45	41.40	0.05			
17	45.86	45.01	0.85			
18	35.78	36.95	1.17			
19	22.40	23.93	1.53			
20	40.45	39.90	0.55			
21	64.53	61.26	3.27			
22	46.90	46.86	0.04			
23	36.18	37.32	1.14			

In order to test whether the regression model can be better used to predict the consumer score of grain products in the market, the multiple regression equation is used to predict the consumer ratings of 23 products in the sample, and the MAE, MSE and MAPE are calculated to analyze the prediction effect of the model [6]. The results are shown in the table. It can be seen from the chart that the prediction result of the multiple regression model is good, with MAE = 1.55, MSE = 3.98 and MAPE = 0.039. Therefore, when a cereal product and its nutritional composition table are given, we can use the multiple regression equation to estimate the score of the product in the consumer group, and then make further analysis on the product sales and improvement direction.

4. Discussion

Through descriptive statistical analysis, it can be concluded that there are certain differences in the nutritional composition of cereal products produced by different manufacturers, which is shown as the difference between the average calories of each manufacturer and the overall average calories. From the perspective of manufacturers, food brands do indeed affect consumers' purchasing preferences. The brand awareness, brand image, and consumers' perception of the use of brand products all affect consumers' impression of the product and even the brand, as well as their purchasing decisions [7]. There are also differences in consumer ratings of products from different manufacturers. Among them, the higher the average calories of the manufacturer's products, the lower the average rating obtained. The Pearson correlation coefficient (-0.638) between the score and calories can also prove the strong negative correlation between the two. Consumer ratings were also negatively correlated with fat, sodium, sugar and vitamin contents; It was positively correlated with protein, fiber, carbon content.

Therefore, in order to improve product quality and promote the development of the grain product market, manufacturers can start from the brand image and enhance consumers' liking and adhesion to the brand. In terms of product sales strategies, manufacturers should pay attention to market demand, continuously enrich the structure of miscellaneous grain varieties, cultivate and introduce excellent and new miscellaneous grain varieties, and provide distinctive and multi-level miscellaneous grain food [8].

Through regression analysis, it can be concluded that consumers' ratings on cereal products can be fitted by multiple regression with some components in the nutrition table, and the regression has good fitting effect. Assuming that the consumer rating in the sample is the standard to measure consumers' preference for cereal products, we can use the above multiple regression equation to predict the product rating based on the content of protein, fat, sodium, fiber, sugar and vitamin in the product nutrition table. As many studies have shown, most consumers pay attention to the nutritional composition of food when purchasing, and consumers are particularly concerned about the calorie and fat content of food [9]. Once certain elements in the nutritional composition table do not meet

consumers' expectations, consumers will also change their overall impression of the product and purchase decisions [10]. In terms of improving products, manufacturers can approximate consumer evaluations based on the nutritional composition of the product, and improve the product based on this. They choose high nutrient materials for production, and promote the product based on its nutritional composition to meet the demand of consumers for healthy grain products [11].

5. Conclusion

The paper uses nutritional composition and consumer rating data from 77 cereal products in the United States to conduct regression analysis and prediction of nutritional composition and consumer rating, explore the impact of each nutritional component on consumer rating, and predict cereal product rating based on regression analysis. The conclusion drawn in this paper includes the following aspects: Firstly, there are differences in nutritional composition and consumer ratings among grain products from different manufacturers. This may be related to the manufacturer's selection of raw materials, processing methods, and the way they promote their products. Secondly, consumers' ratings of cereal products are influenced by their nutritional composition, and they tend to give higher ratings to high protein, high fiber, low fat, and low-calorie cereal products. Therefore, when promoting products, cereal brands can try to focus on promoting the high nutritional value of the product to cater to consumer preferences, thereby improving consumers' overall impression of the product and ultimately improving the overall product rating.

Although the regression results provide direction for manufacturers on how to improve their products, under the influence of various practical factors, achieving the ideal conditions mentioned in the paper still requires multiple efforts and technological improvements. In future research, more comprehensive and universal regression conclusions can be obtained by attempting to expand the sample size, geographical range, and other methods, and we can add more factors into consideration based on nutritional composition, so that we can further improving the accuracy of regression analysis.

References

[1] Chen Weiping, Niu Mingchan. Consumer's use of food Nutrition facts label and its influencing factors [J]. Journal of Renmin University of China, 2009, 23 (04): 105-113

[2] Wang Hongxin, Liu Yuhui. Empirical Study on the Factors Influencing Consumer Satisfaction of Online Purchasing of Fresh Agricultural Products [J]. Consumer Economy, 2015,31 (06): 81-86

[3] Chang Xiangyang, Hu Hao. Research on Consumer Food Safety Attribute Preference Behavior Based on Choice Experiment [J]. Food Industry Technology, 2014,35 (11): 273-277. DOI: 10.13386/j. issn1002-0306.2014.11.051

[4] Chen Yongxiu. Understanding the Meaning of Correlation Coefficients [J]. China Exam, 2011, No.231 (07): 15-19. DOI: 10.19360/j.cnki.11-3303/g4.2011.07.003

[5] Cai Wenting, Peng Yi, Chen Qiuji. Prediction of Air Transport Passenger Volume Based on Multiple Regression Models [J]. Aviation Computing Technology, 2019,49 (04): 50-53+58

[6] Hu Juhua. Diagnosis and Correction of Linear Regression Models Based on Residual Analysis [J]. Statistics and Decision, 2019,35 (24): 5-8. DOI: 10.13546/j.cnki.tjyjc.2019.24.001

[7] Yang Weiwen, Liu Xin. The impact of brand awareness on consumer purchasing behavior [J]. Business Research, 2010 (03): 158-162. DOI: 10.13902/j.cnki. syyj.2010.03.022

[8] Li Yuqin, Zhang Huijie. Analysis of consumer willingness to consume miscellaneous grains and influencing factors - Taking Wuhan consumers as an example [J]. Agricultural Technology and Economics, 2013 (07): 100-109. DOI: 10.13246/j.cnki.jae.2013.07.008

[9] Liu Huaiyu, Shan Chengdi, Lv Jing, etc. Investigation and research on the influencing factors of consumers' use of Nutrition facts label [J]. Chinese Journal of Disease Control, 2012,16 (01): 60-63

[10] Liu Lei. The impact of food Nutrition facts label on consumer attitudes [J]. Food Safety Guide, 2016, No.153 (30): 56. DOI: 10.16043/j.cnki.cfs.2016.30.042

[11] Ren Chaoran, Liu Fang, He Lu. Analysis of Services marketing strategies of low calorie food enterprises [J]. China Market, 2014 (48): 22-23+88.