Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/38/20240550

Research on factors influencing household consumption-take wine as an example

Yuxuan Wang

School of Economics and Management, Beijing University of Technology, Beijing, 100000, China

xiaohama@emails.bjut.edu.cn

Abstract. It has been a great increase on wine consumptions these years. This article aims to identify those of the factors that have an impact on the wine consumptions and to perform clustering to summarize customer segmentation markets. The method of stratified multiple linear regression and stepwise multiple linear regression are used in this article. This article analysis fourteen important factors that may have an effect on wine consumption of a survey on 2240 registered consumers. Through the research in this article, it can be concluded that the consumption is strongly related to age, education, income, the number of times one visits the store. Meanwhile, the number of children in the family has a significant negative impact on alcohol purchase. What's more, the study also finds that making some promotions could have a positive impact on wine consumption. This article also makes a guideline to the company to better understand their customers.

Keywords: Wine consumption, stratified regression, stepwise regression.

1. Introduction

Consumption is the ultimate link that effectively realizes all economic links and is a lasting driving force for promoting economic growth. Especially in the post-pandemic era, against the backdrop of economic globalization being hindered, driving domestic demand has become an important theme. However, residential consumption is finally reflected in consumer purchases. Therefore, predicting consumer purchasing power and willingness to purchase has become an important issue in promoting economic development. Meanwhile, it is of great significance for enterprises to accurately determine customer purchasing power from basic information. Enterprises can not only better understand customers, but also accurately market and launch the products they need based on customer portraits and their different needs.

Currently, some scholars mainly focus on analyzing factors related to household consumption levels and predicting consumption levels. Based on the information consumption data of Beijing from 2010 to 2018, Zhou Yang et al. used a multiple linear regression model to analyze the influencing factors of Beijing residents' information consumption level. The article explores objective factors such as per capita disposable income, population age structure, years of education, level of information industry development, and information infrastructure construction, and demonstrates the different impacts of each factor on the level of information consumption [1].

DOI: 10.54254/2753-8818/38/20240550

Song et al. used the data on Chinese residents' consumption levels from 1992 to 2021 to combine the ARIMA model and the quadratic curve model to construct a combined model for predicting Chinese residents' consumption levels. The absolute value of the error and the MAPE between the predicted value and the actual value of the combined model were calculated [2].

As for in the field of wine consumption, Campos Arteaga Nicole et al. established a n extended model of the Theory of Planned Behavior, aiming to determine a relationship between purchase intention and perceived behavioral control. The study evaluates the antecedents of Italian local wine consumption [3]. Kruger and Viljoen identified the preferences of wine consumers based on age cohorts to identify differences in consumption in developing country. The research filled the current literature gap regarding the characteristics of the different age cohorts and indicates that wine consumption and preferences may change over time [4]. Monica et al. drew the conclusion that people with higher income have a disposition to more valuable wine, and experienced people in the domain inclined to pay more for a bottle of wine [5]. Wang et al. found that the age group of consumers has a significant impact on drinking frequency, wine tasting knowledge level, number of drinks, and drinking duration. Consumers in different age groups have significant differences in their drinking frequency [6]. Li went into descriptive analysis and factor analysis on the pre-factors of wine purchasing behavior such as wine knowledge level, purchasing motivation, and purchasing willingness of Chinese consumers, and believed that education and income are both factors that affect wine consumption level [7, 8].

Most of the above studies lack in-depth exploration of factors such as age, discounts, etc. on consumer purchase intention [9]. Meanwhile, the previous data was mostly based on macroeconomic data, and there was also a lack of analysis on consumer personality. By summarizing the research results of the above scholars, this article analyzes 2240 samples of respondent, and studies the influencing factors of the purchasing power of registered users on a certain shopping platform. It can better understand the problem of consumption stratification and has certain reference value for enterprises to find target customers and precision marketing. This study focuses on finding the influences of demographic factors on actual consumption levels.

2. Methodology

2.1. Data source

The data comes from Kaggle. This article refers to the company's ideal customers: marketing strategy survey database conducted by AMAN CHAUHAN et al. on 2240 registered consumers, and conducts research and analysis on the consumption and related data of registered users on a shopping platform. It is divided into four modules: basic information, products, promotion, and place, with a total of 26 rows and 2240 valid data items.

2.2. Indicator selection and explanation

Referring to the classification method of Peng Liang et al. and combining literature materials, it is found that the main influencing factors of household consumption involve 10 primary indicators and 14 secondary indicators [10].

Indicator selection: The dependent variable is divided into two parts: total food consumption and total gold consumption. The independent variables related to total food consumption include age, education level, marital status, family member status, family income, number of family descendants, purchasing channels, etc.

Among them, the status of family members is divided into the number of children in the family and the number of young people in the family. The discount section includes the number of purchases made during the 1-5th discount. The selected indicators are in table 1.

DOI: 10.54254/2753-8818/38/20240550

Table 1. List of variables.

Logogram	Variable	Explanation
x1	Age	Customer's age
x2	Education	Customer's education level
x3	Marital_Status	Customer's marital status
x4	Income	Customer's yearly household income
x5	Num WebVisitsMonth	Number of visits to website in last month
x6	Num Children	Number of children in customer's household
x7	Num Deals Purchases	Number of purchases made with a discount
x8	Num WebPurchases	Number of purchases made through the s website
x9	Num CatalogPurchases	Number of purchases made using a catalogue
x10	Num StorePurchases	Number of purchases made directly in stores
x11	AcceptedCmp1	Accept when the first discount is offered
x11	AcceptedCmp2	Accept when the second discount is offered
x12	AcceptedCmp3	Accept when the third discount is offered
x13	AcceptedCmp4	Accept when the forth discount is offered
x14	AcceptedCmp5	Accept when the fifth discount is offered
Y	MntWines	Costumer's wine consumption

2.3. Method indication

For each secondary indicator, determine their weights and calculate the primary indicator. Establish a stepwise linear regression model based on primary indicators and conduct hypothesis testing, including multicollinearity testing, etc.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (1)$$

Here, $\beta_0, \ldots, \beta_p$ are the regression coefficients.

3. Results and discussion

3.1. Data processing

Data cleaning was conducted to remove 26 items of missing income data, resulting in 2,214 valid data items. The original data contains the birth year of each customer, and based on this information, the customer's age was calculated. A simple analysis of the above data reveals that wine-related consumer goods are the most expensive category in the sample, with a per capita consumption of 305.0 dollars over the past two years.

3.2. Hierarchical regression results

For wine-related consumer goods, it can be seen the Pearson correlation coefficient between these factors and customer consumption. The impact on residents' consumption in the field of diet is diverse. Taking the consumption of wine products by customers in the past two years as an example, these data are divided into four categories: the first layer of basic information, including Age, Education, Kidhome, Income. The second layer of purchase-related information, including Num DealsPurchases, NumWebPurchases, NumCatalogPurchases, NumstorePurchases, NumWebVisitsMonth. The third layer of discount information, including AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5. The forth layer contains Complain, Response, Recency. Then the author makes a hierarchical regression of these contents, n=2216. The specific results are as follows in table 2:

DOI: 10.54254/2753-8818/38/20240550

Table 2. Results of stratified regression.

	Layer1		Layer2		Layer3		Layer4	
	B	p	B	p	B	p	B	p
Constant	-1.611	0.956	-364.90**	0	-337.05**	0	-347.05**	0
x1	0.182	0.7	0.521	0.18	0.855*	0.013	0.877*	0.0
x2	25.113**	0	22.00**	0	22.057**	0	21.784**	0
x3	0.006**	0	0.003**	0	0.002**	0	0.002**	0
x4	-190.921**	0	-47.20**	0	-44.491**	0	-44.402**	0
x5			-13.79**	0	-5.154*	0.035	-5.243*	0.03
x6			22.93**	0	21.066**	0	20.935**	0
x7			22.71**	0	18.736**	0	18.636**	0
x8			31.17**	0	29.725**	0	29.937**	0
x9			38.46**	0	31.007**	0	30.675**	0
x10					37.669*	0.016	35.400*	0.02
x11					180.562**	0	178.492**	0
x11					98.715**	0.007	96.524**	0.00
x12					52.403**	0.004	50.036**	0.00
x13					269.801**	0	266.562**	0
Complain							-50.9	0.21
Response							12.287	0.33
Recency							0.202	0.14
R^2	0.42		0.617		0.701		0.702	
F	F (4,2211)=399.803		F (9,2206)=395.365		F (14,2201)=369.090		F (17, 2198) =304.33	

3.3. Stepwise regression results

According to the hierarchical regression, Education, age, and income have strong positive correlations with alcohol consumption. “Kidhome” has a negative correlation with alcohol consumption. Considering that it is not clear which indicators have a more significant impact on the dependent variable, the stepwise regression method was used to study 2194 valid data with ID greater than 100 in the sample. After screening, Age, Education, Income, Kidhome, NumDealsPurchases, AcceptedCmp5, AcceptedCmp4, AcceptedCmp3, NumWebPurchases, AcceptedCmp2, NumCatalogPurchases, AcceptedCmp1, NumStorePurchases, NumWebVisitsMonth are list in the model. The results of the model after 14 iterations are shown in the following table 3:

The table 3 shows the regression coefficients of the multiple linear regression equation model. The adjusted p-values for all independent variables in the T-test do not exceed 0.005. Therefore, it can be concluded that the fourteen independent variables have a significant impact on the dependent variable Y, and the above linear regression has statistical significance. Based on the above data, the relevant multiple linear regression equation can be obtained. The model formula is:

$$Y = -335.333 + 0.832X_1 + 22.301X_2 + \cdots + 18.772X_{14} \quad (2)$$

The Multiple linear regression analysis in this paper must pass the collinearity test. In this linear regression, the VIF values of all independent variables are below 5, indicating that there is no multicollinearity in these data and no significant mutual influence between the data. The R-square in the model is 0.700, meaning that the above fourteen variables can explain 70.0% of the reasons for changes

DOI: 10.54254/2753-8818/38/20240550

in wine consumption. And the D-W value is near the number 2, indicating that the model does not have autocorrelation and there is no correlation between the sample data, making the model sound.

Table 3. Results of stepwise regression.

	B	t	p	VIF	tolerance
Constant	-335.333	-12.037	0.000**	-	-
Age	0.832	2.394	0.017*	1.107	0.903
Education	22.301	6.949	0.000**	1.022	0.979
Income	0.002	7.336	0.000**	2.059	0.486
Kidhome	-45.455	-4.577	0.000**	1.811	0.552
NumDealsPurchases	-5.2	-2.114	0.035*	1.431	0.699
AcceptedCmp5	269.29	14.827	0.000**	1.436	0.696
AcceptedCmp4	181.549	10.656	0.000**	1.271	0.787
AcceptedCmp3	37.213	2.363	0.018*	1.086	0.921
NumWebPurchases	21.166	11.022	0.000**	1.766	0.566
AcceptedCmp2	98.798	2.655	0.008**	1.15	0.87
NumCatalogPurchases	30.914	15.619	0.000**	2.139	0.468
AcceptedCmp1	52.062	2.831	0.005**	1.303	0.768
NumStorePurchases	29.591	17.108	0.000**	2.021	0.495
NumWebVisitsMonth	18.772	7.738	0.000**	2.211	0.452
R ^2			0.7		
Adjusted R ^2			0.699		
F			F (14,2179)=363.968,p=0.000		
D-W			2		

dependent variable: MntWines

* p<0.05 ** p<0.01

3.4. Line thickness

Next, this paper needs to verify the accuracy of the linear regression model for the given data. This article used 17 groups of data (customers with IDs less than 75) that were not subjected to linear regression in the survey database to conduct a posterior difference test. The specific steps are as follows: Compare the predicted data with the original data, and judge the quality of the data based on the degree of comparison. After analysis, the standard deviation of the residuals, S2, is 45.280, and the standard deviation of the original data S1 is 337.328. The ratio C is 0.134, which is less than 0.35, proving that the original linear regression model can better predict consumers' spending power and willingness to spend on food.

4. Conclusion

This study selects 2216 samples from the data set, which has 14 variables. During the analysis stage, the article uses Stratified multiple linear regression model and stepwise multiple linear regression model to find out the possible relationship between the variables and the total amount of money of food consumption. Through the research in this article, it can be concluded that the consumption is strongly related to some basic identity information (e.g., age, education). Meanwhile, the study found that the number of children in the family has a significant negative impact on alcohol purchase. However, due to the uncertainty of the consumption habits and so on, predictions made further away will become more inaccurate.

DOI: 10.54254/2753-8818/38/20240550

Through the predictions and research in this article, some suggestions can be provided for the policy maker of the company who sells food products. Making some promotions could have a positive impact on wine consumption, and companies can set discounts based on sales goals. By obtaining some basic information about customers, companies can understand more than 40% of their purchasing preferences and habits. Combined with past purchasing data (such as the number of purchases in the store, the number of web pages viewed, etc.), it is possible to effectively predict customers' wine consumption and other consumption.

References

[1] Zhou Y, Shi X X, Lv Q H and Xiao X W 2020 Analysis of the influencing factors of information consumption level of Beijing residents based on multiple linear regression. Shopping Mall Modernization, 21, 48-50.
[2] Song Q X, Tang Y Z and Meng J 2024 Prediction of Resident Consumption Level Based on Combined Models. Northern Economic and Trade, 51-54.
[3] Campos A N, et al. 2022 Prediction of Consumption of Local Wine in Italian Consumers Based on Theory of Planned Behavior. Sustainability, 14(22), 14769.
[4] Kruger M and Viljoen A 2022 That old saying about wine and age: identifying South African age-cohort preferences. International Journal of Wine Business Research, 34(4), 495-522.
[5] Monica A B, et al. 2022 Determinants of Choice and Wine Consumption Behaviour: A Comparative Analysis between Two Counties of Romania. Foods, 11(8), 1110.
[6] Wang Y B, et al. 2019 Research on the differences in wine consumption behavior based on different age segments. China Brewing, 38(3), 210-214.
[7] Li J G 2015 Research on Wine Consumers' Behavior in China. Northwest A&F University.
[8] Peng L and Li L 2023 Analysis of Factors Influencing the Consumption Capacity of the Middle Class: A Case Study of Hunan. Industrial Innovation Research, 70-72.
[9] Huang Y Y, Lu X H and Xu B 2023 Epidemic and Intercity Consumption Flow: From the Perspective of Urban Consumption Function and Industry Digitalization. Journal of Management Science, 26(5), 248-270.
[10] Lv Q H, Lin B K and Huang M 2023 Research on the Impact of Online Initial Review Fit on Consumer Purchase Intention. Journal of Henan Animal Husbandry Economics College, 36(1), 8.