SAS® GLOBAL FORUM 2018

USERS PROGRAM

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products
Zihan Fan & Xiaojing Zhao

Zihan Fan, Graduate Student, Clark University
Xiaojing Zhao, Graduate Student, Clark University

Zihan Fan is a master of science in business analytics in Clark University, who worked in Marketing Department in Amazon Kindle for one year, and started to learn SAS® last summer.

Xiaojing Zhao is now a business analytics student in Clark University, who has been worked in Bank of China, Beijing branch for 3 years. She started to learn SAS® since last summer on big data course.

Insert your Twitter handle if applicable.
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products

INTRODUCTION

• Motivation: We found that the body sunscreen products of the US market usually merely have UV protection and broad spectrum. So, we wonder what are the main concerns for US consumers to purchase specific products.

• Purpose: This study is to analyze the crucial factors that affect consumers’ preference of sun care products in different states.

• Crucial factors(variables): Average hour on sunshine in four seasons in 2016, median annual household income in each state in US, population distribution by race in US and reviews written in 2016 from Amazon which categorized by each state in US.

• Research questions:
  - What are consumers’ main concerns when choosing sunscreen products?
  - What makes consumers keep loyalty to specific types of sunscreen products?
  - Is there a difference in the demand for sunscreens’ features among different race lived in the same region in US?
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products

**METHODS**

Data Collection: Amazon, kff.org, currentresults.com

Data Preparation: data cleaning, aggregated data

Data Exploration: visualization, summary statistics

Statistical Analysis: t-test & Linear regression
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products

Figure 1 shows that the southwestern region has the highest average hours of sunshine within four seasons of a year.

Figure 2 illustrates that the range of each data is around 450, which indicates that there is obvious variation from state to state.
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products

Figure 3 shows that Sum Quantity of reviews = - 0.00056729 * Median annual household income + 4.56386 * percentage of Asian resident + 60.47743
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products

H₀: \( \mu_{\text{Body}} = \mu_{\text{Face}} \)
Hₐ: \( \mu_{\text{Body}} \neq \mu_{\text{Face}} \)
The results show that the p-value is 0.0001, which is less than \( \alpha(0.05) \). We can reject the null hypothesis because we have strong statistical evidence from our sample that there is difference between sales of body sunscreen products and face sunscreen products.

H₀: \( \mu_{\text{SPF50}} = \mu_{\text{SPF55+}} \)
Hₐ: \( \mu_{\text{SPF50}} \neq \mu_{\text{SPF55+}} \)
The results show that the p-value is 0.0020, which is less than \( \alpha(0.05) \). We can reject the null hypothesis because we have strong statistical evidence from our sample that there is difference between sales of SPF50-sunscreen products and SPF55+ sunscreen products.

H₀: \( \mu_{\text{Lotion}} = \mu_{\text{Spray}} \)
Hₐ: \( \mu_{\text{Lotion}} \neq \mu_{\text{Spray}} \)
The results show that the p-value is 0.0010, which is less than \( \alpha(0.05) \). We can reject the null hypothesis because we have strong statistical evidence from our sample that there is difference between sales of lotion sunscreen products and spray sunscreen products.

H₀: \( \mu_{\text{Sensitive}} = \mu_{\text{Normal}} \)
Hₐ: \( \mu_{\text{Sensitive}} \neq \mu_{\text{Normal}} \)
The results show that the p-value is 0.0002, which is less than \( \alpha(0.05) \). We can reject the null hypothesis because we have strong statistical evidence from our sample that there is difference between sales of sensitive sunscreen products and normal sunscreen products.
Understanding the Factors that Affect Customers’ Choice of Sunscreen Products

CONCLUSIONS

• The reviews on Amazon of sunscreen products are negatively correlated with median annual household income, which might be caused by that consumer with high income would require higher quality of service.
• Reviews on Amazon of sunscreen products are positively correlated with percentage of Asian residents within a state.
• In each state, reviews on Amazon of sunscreen products with different features are different, so consumers have preferences toward different features.
• Based on our descriptive analysis, we recommend focusing on Southwestern region and trying to explore more types of demand about sunscreen products in this region.
• According to our t-test analysis and linear regression model, we recommend that the manufacturers could pay more attention to targeting Asian consumers’ needs in each state to acquire more sales potential.
• We believe that this project will help the sunscreen product manufacturers come up with more accurate manufacturing strategies and marketing plans.
SAS®
GLOBAL FORUM
2018
April 8 – 11 | Denver, CO
Colorado Convention Center
ABSTRACT:
Recently, we found that the body sunscreen products of the US market usually merely have UV protection and broad spectrum. So, we wonder what are the main concerns for US consumers to purchase specific products. The purpose of this study is to analyze the crucial factors that affect consumers’ preference of sun care products in different states. We come up with several factors and collect the relative data from historical data on sunscreens and study the relationship between the factors and sun care products sales in each state in US. We use SAS Enterprises Guide to perform analysis. We believe that this project will help the sunscreen product manufacturers come up with more accurate manufacturing strategies and marketing plans. In our research, we use data from websites and integrate the information in the form we need. According to the data analysis, we can mainly give the manufacturer three suggestions, which are, firstly, focusing on Southwestern region and trying to explore more types of demand about sunscreen products in this region. Secondly, focusing more on daily sunscreen products, rather than other featured type of sun care products. Thirdly, the manufacturers could pay more attention to targeting Asian consumers’ need in each state to acquire more sales potential.

INTRODUCTION:
Northeastern Asians pay more attention to sunscreen, since over sunlight may cause tanning, accelerating skin aging and sunburn, even skin diseases. Recently, we found that the body sunscreen products of the US market usually merely have UV protection and broad spectrum. So, we wonder what is the main concern for US consumers to purchase specific sunscreen products. We believe that this conclusion will make the products more in line with market expectations of the sunscreen product and provide a more effective way to prevent skin diseases. Besides, we also believe that this project will help the sunscreen product manufacturers come up with more accurate manufacturing strategies and better marketing plans. Moreover, people now pay increased attention to health-related problems. Skin diseases as a prevalent health problem, may lead to skin cancer. It is important to promote more effective sunscreen products to prevent skin diseases caused by sunburn.
The management question we focus on is what marketing and production strategy should manufactures take to improve their sales in each state in the United States and help consumers prevent skin diseases caused by sunburn. Consequently, the research questions we have taken into consideration are listed below:

1. What are the factors that cause the different preferences of consumers in different regions, i.e. what are their main concerns when choosing sunscreen products?
2. How do the consumers weigh the performance of the sunscreen products?
3. What is the main reason for consumers to choose the specific products? In other words, what makes them keep loyalty to specific types of sunscreen products?
4. How do people in targeting region perceive sunscreen products?
5. Are existing sun care products meeting all of the most important customer needs in targeting regions?
6. Is there a difference in the demand for sunscreens’ features among different race lived in the same region in US?

In the article “Sunscreen Product Performance and Other Determinants of Consumer Preferences”, the writers determined the characteristics and the most commonly cited positive and negative features of highly rated sunscreens described by consumers on Amazon. The analysis result could be used as suggestions to dermatologists making recommendations to their consumers. In the “Assessing the current market of sunscreen: A cross-sectional study of sunscreen availability in metropolitan counties in the United States”, the writer's analyzed sunscreen availability in three large metropolitan counties to determine the relationship between availability and community demographics. As for “Trends in sunscreen recommendation among US physicians”, to evaluate trends in sunscreen recommendation among physicians to determine whether they are following suggested patients - education guidelines regarding sun protection, and to assess data physician sunscreen recommendations to determine the association with patient demographic, physician specialty, and physician diagnosis.

As far as what we have dug into in this field, we have found several studies about the relationship between sunscreen products and skin diseases, and classification of different features on sunscreens that are popular in different regions, as well as analysis on several factors that affect consumers’ preferences. However, we are focusing on the main factors that would mostly influence consumers’ preference of choosing sun care products in different regions of the United States, where the weather and climate are different, population compositions are various, and the economic development is on different level. Based on the former researches, we are aiming at coming up with a consulting result for manufacturers to
help them change their current marketing strategy to improve their sales and profit, as well as people’s acceptance toward sunscreen products.

**METHODS AND ANALYSIS**

We have mined average hours of sunshine, median annual household income and population distribution by race, as well as quantity of reviews on Amazon of sunscreen products with different features, i.e. used for face or body, whether it is sensitive or not, SPF under 50 or above 55, spray or lotion, in each state for our data to be used. We look into review writers’ profiles on Amazon to find their state information. All the data is from the www.currentresults.com, www.amazon.com and www.kff.org websites.

For variables, we chose the average hour on sunshine in four seasons in 2016, median annual household income in each state, population distribution by race in US as independent variables to figure out what influence will the sunshine has on consumption of sunscreen products, and reviews written in 2016 from Amazon which categorized by each state in US. We combined the data of different seasons in one sheet and deleted the invalid data points that cannot be estimated in an appropriate way.

We use paired T-Test to do the hypothesis test and linear regression analysis to figure out whether the factors have relationship with each other. The analysis is processed via SAS Enterprises Guide.
RESULT

The Figure 1 shows the average sunshine hours in nine different regions in the United States. It is clear that the southwestern region has the highest average hours of sunshine within four seasons of a year. So, it is wise for producers to focus on this region and try to explore more types of demand about sunscreen products in this region.

![Average sunshine hours in each region](image)

**Figure 1 Average sunshine hours in nine regions in US**

The Figure 2 illustrates that the range of each data is around 450, which indicates that there is obvious variation from state to state. As a consequence, we suggest that the producers’ marketing strategy should different according to the different environmental situation of different states.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>spring</td>
<td>746.32</td>
<td>91.58</td>
<td>607.00</td>
<td>1079.00</td>
<td>472.00</td>
<td>50</td>
</tr>
<tr>
<td>Summer</td>
<td>898.39</td>
<td>100.66</td>
<td>729.00</td>
<td>1209.00</td>
<td>480.00</td>
<td>50</td>
</tr>
<tr>
<td>Fall</td>
<td>603.12</td>
<td>101.03</td>
<td>358.00</td>
<td>888.00</td>
<td>530.00</td>
<td>50</td>
</tr>
<tr>
<td>Winter</td>
<td>465.91</td>
<td>106.08</td>
<td>232.00</td>
<td>764.00</td>
<td>532.00</td>
<td>50</td>
</tr>
</tbody>
</table>

**Figure 2 Statistics of average sunshine hours in four seasons in US**

The descriptive analysis in Figure 3 shows that the population distribution composition is white, black, Hispanic and Asian. Among these four races, the white accounts for 69.32%. It can be concluded from the chart above that the main consumers of sunscreen products are white people. As a result, the manufacturer should focus on white people's preferences and their special need toward sunscreen products. However, to improve the market share, we suggest the manufacturer to try to reach out to people from different races.
In 2016 annual median household income dataset, according to our statistical analysis result, which presented as Figure 4 and Figure 5, we can see the distribution is approaching normal, which means that most common annual household show in the interval between 50,000 and 60,000. Consequently, we suggest that manufacturer can notice this point when they make their pricing strategy.

In figure 6, it is obvious that the sales of sunscreen products aimed to body are more than to face. The sunscreen lotion is more popular than sunscreen spray. The sales of sunscreen products spf50- are more than those of spf55+. In a conclusion, we suggest manufacturers to produce normal spf50- sunscreen lotion.
From figure 7, we make the following analysis.

H₀: β₁=β₂=β₃=β₄= 0

Hₐ: At least one of the regression coefficients is ≠ 0

Conclusion: The overall model is significant as p-value is less than α (0.05). So, we reject the null and say at least one of the regression coefficients is not equal to zero. Unstandardized estimate is the change in the amount of dependent variable if we change the independent variable by one unit whereas standardized estimate is the change in the amount of the standard deviation of dependent variable due to a change of one standard deviation of independent variable. In this model unstandardized estimate of Asian is 2.81, that means one-unit change in sales of lotion sunscreen product will change quality of life by 2.81 unit. The estimate is different from the previous model since this one is a multiple regression model keeping other three predictors constant.
From figure 8, we make the following analysis.
$H_0$: $\beta_1=\beta_2=\beta_3=\beta_4= 0$
$H_a$: At least one of the regression coefficients is $\neq 0$

Conclusion: The overall model is significant as p-value is less than $\alpha$ (0.05). So, we reject the null and say at least one of the regression coefficients is not equal to zero. Unstandardized estimate is the change in the amount of dependent variable if we change the independent variable by one unit whereas standardized estimate is the change in the amount of the standard deviation of dependent variable due to a change of one standard deviation of independent variable. In this model unstandardized estimate of Asian is 0.876, that means one-unit change in sales of spray sunscreen product will change quality of life by 0.876 unit. The estimate is different from the previous model since this one is a multiple regression model keeping other three predictors constant.

![Linear Regression Results](image)

**Figure 8 Linear regression result of spray and different races**

From figure 9, we make the following analysis.
$H_0$: $\beta_1=\beta_2=\beta_3=\beta_4= 0$
$H_a$: At least one of the regression coefficients is $\neq 0$

Conclusion: The overall model is significant as p-value is less than $\alpha$ (0.05). So, we reject the null and say at least one of the regression coefficients is not equal to zero. Unstandardized estimate is the change in the amount of dependent variable if we change the independent variable by one unit whereas standardized estimate is the change in the amount of the standard deviation of dependent variable due to a change of one standard deviation of
independent variable. In this model unstandardized estimate of Asian is 2.69, that means one-unit change in sales of SPF50- sunscreen product will change quality of life by 2.69 unit. The estimate is different from the previous model since this one is a multiple regression model keeping other three predictors constant.

**Figure 9 Linear regression result of SPF50- and different races**

From figure 10, we make the following analysis.

H₀: β₁=β₂=β₃=β₄= 0

Hₐ: At least one of the regression coefficients is ≠ 0

Conclusion: The overall model is significant as p-value is less than α (0.05). So, we reject the null and say at least one of the regression coefficients is not equal to zero. Unstandardized estimate is the change in the amount of dependent variable if we change the independent variable by one unit whereas standardized estimate is the change in the amount of the standard deviation of dependent variable due to a change of one standard deviation of independent variable. In this model unstandardized estimate of Asian is 0.99, that means one-unit change in sales of SPF55+ sunscreen product will change quality of life by 0.99 unit. The estimate is different from the previous model since this one is a multiple regression model keeping other three predictors constant.
The null hypothesis in this case would be that there is no difference between sales of body sunscreen products and face sunscreen products. The alternate hypothesis would be there is difference between sales of body sunscreen products and face sunscreen products.

\[ H_0: \mu_{\text{body}} - \mu_{\text{face}} = 0 \]

\[ H_a: \mu_{\text{body}} - \mu_{\text{face}} \neq 0 \]

This is a two-tail hypothesis. Consider \( \alpha = 0.05 \). This hypothesis can be tested using the paired t-test. The results from the TTEST procedure are given below. The results show that the p-value is 0.0001, which is less than \( \alpha \). Therefore, we can reject the null hypothesis because we have strong statistical evidence from our sample that there is difference between sales of body sunscreen products and face sunscreen products.
The null hypothesis in this case would be that there is no difference between sales of SPF50- sunscreen products and SPF55+ sunscreen products. The alternate hypothesis would be there is difference between sales of SPF50- sunscreen products and SPF55+ sunscreen products.

\[ H_0: \mu_{SPF50-} - \mu_{SPF55+} = 0 \]
\[ H_a: \mu_{SPF50-} - \mu_{SPF55+} \neq 0 \]

This is a two-tail hypothesis. Consider \( \alpha = 0.05 \). This hypothesis can be tested using the paired t-test. The results from the TTEST procedure are given below. The results show that the p-value is 0.0001, which is less than \( \alpha \). Therefore, we can reject the null hypothesis because we have strong statistical evidence from our sample that there is difference between sales of SPF50- sunscreen products and SPF55+ sunscreen products.
The null hypothesis in this case would be that there is no difference between sales of lotion sunscreen products and spray sunscreen products. The alternate hypothesis would be that there is difference between sales of lotion sunscreen products and spray sunscreen products.

\[ H_0: \mu_{lotion} - \mu_{spray} = 0 \]
\[ H_a: \mu_{lotion} - \mu_{spray} \neq 0 \]

This is a two-tail hypothesis. Consider \( \alpha = 0.05 \). This hypothesis can be tested using the paired t-test. The results from the TTEST procedure are given below. The results show that the p-value is 0.0001, which is less than \( \alpha \). Therefore, we can reject the null hypothesis because we have strong statistical evidence from our sample that there is a difference between sales of lotion sunscreen products and spray sunscreen products.

The null hypothesis in this case would be that there is no difference between sales of sensitive sunscreen products and normal sunscreen products. The alternate hypothesis would be that there is difference between sales of sensitive sunscreen products and normal sunscreen products.

\[ H_0: \mu_{sensitive} - \mu_{normal} = 0 \]
\[ H_a: \mu_{sensitive} - \mu_{normal} \neq 0 \]

This is a two-tail hypothesis. Consider \( \alpha = 0.05 \). This hypothesis can be tested using the paired t-test. The results from the TTEST procedure are given below. The results show that the p-value is 0.0001, which is less than \( \alpha \). Therefore, we can reject the null hypothesis because we have strong statistical evidence from our sample that there is a difference between sales of sensitive sunscreen products and normal sunscreen products.
Figure 14 Paired t Test of sensitive sunscreen products and normal sunscreen products

From figure 14 we can see that based on the p-value in analysis of variance, which is less than 0.0001, we can reject the null hypothesis that all beta in the model are equal. As a result, the R-square shows that there are 70.12% of the data can be explained by the model. In the parameter estimates chart, we can see that only two variables’ p-value are less than 0.05. Consequently, our linear model would be:

\[
\text{Sum Quantity of reviews} = -0.00056729 \times \text{Median annual household income} + 4.56386 \times \text{percentage of Asian resident} + 60.47743
\]

The explanation of the model is that with keeping the median annual household income as a constant, increasing one percentage of Asian resident would increase one unit of sum quantity of reviews on Amazon about sunscreen products. If we keep the percentage of Asian residents as constant, one unit increment in median annual household income will decrease 0.00056729 units’ reviews on Amazon about sunscreen product.

Also, we noticed that some of the variance inflation values that are larger than 10, such as average sunshine hour in fall, percentage of white people, black people and Hispanic people. So, we conclude that there is multicollinearity in our model.
Figure 15 Linear regression result of sum quantity and all variables in our model in full model

From the histogram and Q-Q plot in figure 15, we can see that the data of sum quantity is normally distributed. The RStudent chart shows that there are several outliers.
The results of our correlation analysis are shown in figure 17, from which we found out that average sunshine hour in spring 2016 is highly associated with average sunshine hour in fall and winter 2016. It has moderate association with average sunshine hour in summer 2016, percent of white people and Hispanic people. Median annual household income in 2016 and percentage of black people and Asian are low associated with it.

For average sunshine hour in summer 2016, average sunshine hour in fall and percentage of black people and Hispanic are moderately associated, while others are low associated.

For average sunshine hour in fall 2016, average sunshine hour in winter is highly associated, while percentage of white people and Hispanic are moderately associated. And median annual household income in 2016, percentage of black people and Asian are low associated.
CONCLUSION

From current analysis, first of all, the reviews on Amazon of sunscreen products are negatively correlated with annual household income, which might be caused by that consumer with high income would require higher quality of service. Secondly, reviews on Amazon of sunscreen products are positively correlated with percentage of Asian residents within a state. Thirdly, in each state, reviews on Amazon of sunscreen products with different features are different, so consumers have preferences toward different features. Moreover, Amazon reviews of sunscreen products with different features have a positive linear relationship with Asian residents in US. However, we cannot conclude the relationship between sunscreen sales and hours of sunshine, which is indicated by amount of reviews on Amazon in 2016.

Based on our descriptive analysis, we recommend, firstly, focusing on Southwestern region and trying to explore more types of demand about sunscreen products in this region. Secondly, focusing more on daily sunscreen products, rather than other featured types of sun care products. According to our t-test analysis and linear regression model, we recommend that the manufacturers could pay more attention to targeting Asian consumers’ needs in each state to acquire more sales potential. In the meantime, the various sale channels’ concentration should be weighed by states that have more races and states that have fewer races. What’s more, it is necessary for manufacturers to produce paired features sunscreen products, such as lotion/spray, sensitive/normal, SPF50-/SPF55+ and body/face. In addition, Amazon could try to improve their services in states that have rather high annual income and target consumers that have rather high income.
LIMITATION

From our current analysis, we cannot conclude the relationship between hours of sunshine and sunscreen sales, which is indicated by amount of reviews on Amazon in 2016. Besides, consumers do not tend to write reviews every time they consume products and not every customer fills in their Amazon profile with states. Even if they have states information, there is still a chance that they are not they were in 2016, when they wrote reviews. We have not collected all the data we plan to have, which are either real sales data of sunscreen products from various channel or some data like quantity of reviews from various websites. Whether the amount of reviews could reflect sales information properly should be reconsidered, since the reviews might neither be written by consumers nor write to show their opinions about the products. For future study, we would find sale data on various consuming channel to get the original targeted results.

REFERENCE


4. [https://www.currentresults.com](https://www.currentresults.com)

5. [https://www.kff.org](https://www.kff.org)