FALL 2019

BUAN 6337.501: Predictive Analytics using SAS

*A Report On*

**Peanut Butter Brand Marketing**

Under guidance of

Prof. B.P.S. Murthi

**Submitted by**

**Group 10 -**

Aarya Aarya

Adithya Ganesh

Neerja Deogaonkar

Jen-Hsin Huang

Shih-Yen Cheng

**1. INTRODUCTION**

Skippy is a brand of peanut butter manufactured by Unilever Best foods North America. It is one of the best sellers worldwide. With data available from Professor Prof. B.P.S. Murthi, we had made the following analysis to understand and based on these questions, we provide recommendation to each of the following questions:

1. What is the effect of price, display, feature and promotion on the sales of Peanut Butter?
2. What are the demographic features of the most valuable customers of the Skippy brand?
3. What are the brand preferences among different customers?

In this report, we have captured several insights that help determine the standing that Skippy establishes in the market against various other peanut butter brands.

Our initial analysis revealed that Skippy has the highest market share second only to JIF.

**2. PROBLEM STATEMENT-1**

Analyze the effect of price, display, feature and promotion on the sales of Peanut Butter

**Data Preparation:**

729 stores which had data for all 52 weeks were filtered to create a balanced data set for Panel regression.

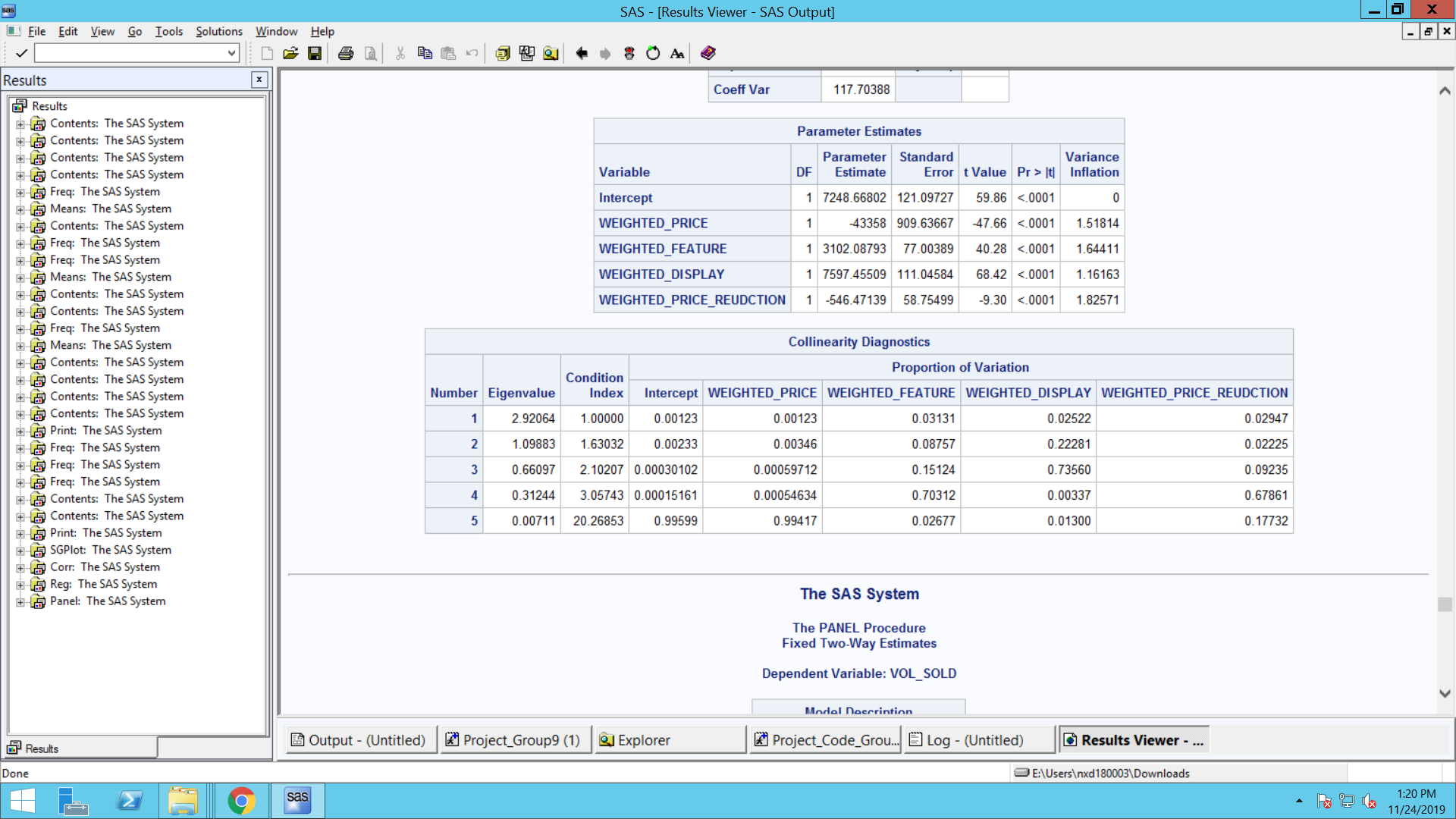
Standardized price per unit for a unit volume equivalent was calculated so that they can be compared with all other brands

**Total Volume** = Volume Equivalent of a Product \* Number of Units Sold

**Standardized Price** = Total Cost of a Purchase / Total Volume

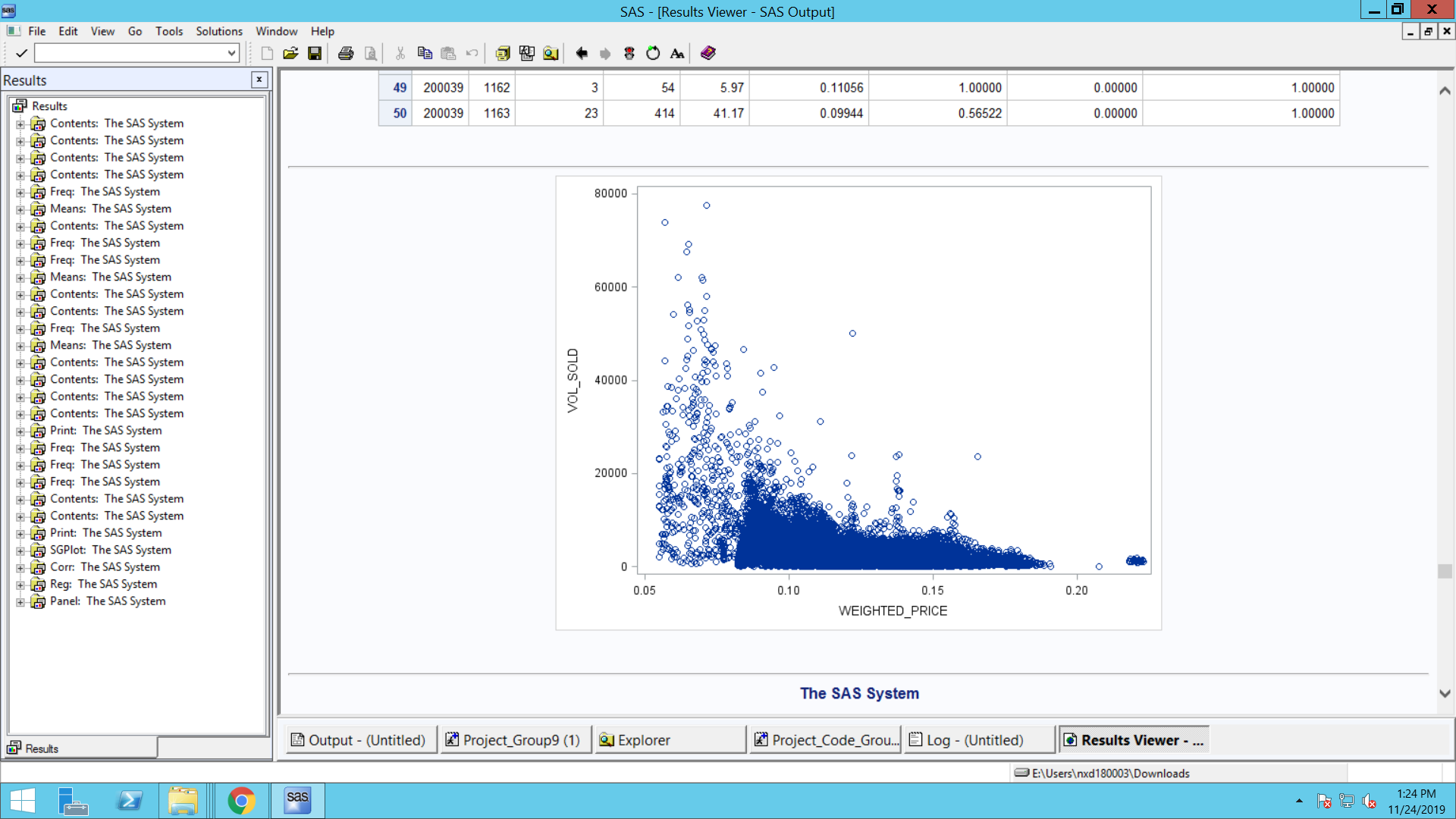
**Multicollinearity Check:**

We observe that the VIF values are well within 10 and the Collinearity Condition indices are much lesser than 100. Therefore, we can rule out multicollinearity in this model.



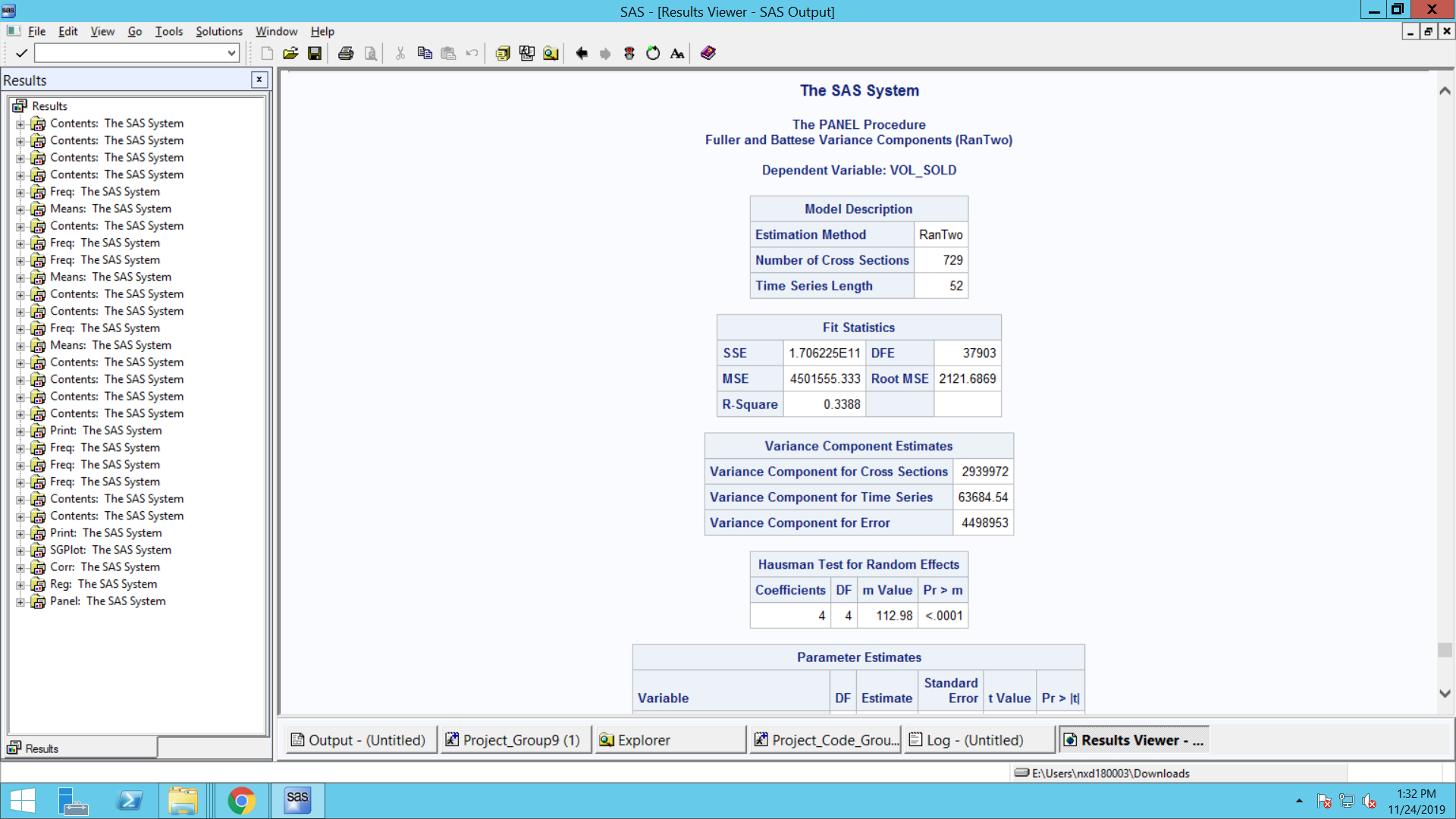
**Non-Linearity Check:**

These clearly shows the Non-Linear characteristics of Price. For this reason, we have also added a squared price term in our model.



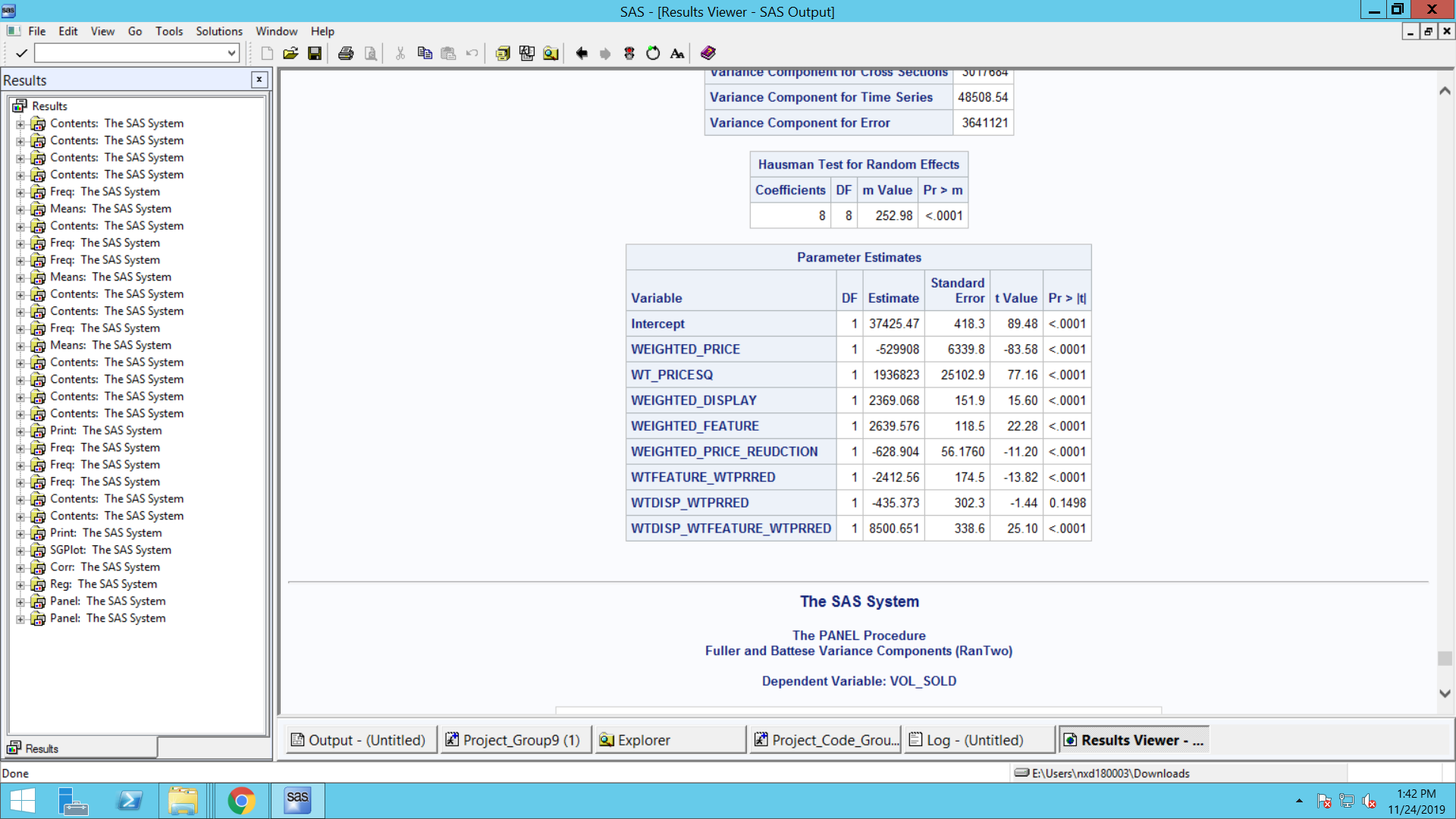
**Panel Regression:**

Based on the Hausman test, we go with Fixed effects model.



**Dependent Variable:** Total Units Sold/Week in a Store

**Independent Variables**: ​Weighted Price, Weighted Display, Weighted Feature, Weighted Promotion, Interaction between Feature and Promotion, Interaction between Display and Promotion, Interaction between Feature and Promotion, Interaction between Display, Feature and Promotion



Effects of Price, Display, Feature and Price Reduction on Sales

Apart from the interaction effect between display and promotion every other variable is significant

**Weighted Price:** Nonlinear relationship between price and number of units sold is seen from the estimates. As price increases the sales decreases to a certain limit and then increases.

**Weighted Display:** The number of units sold per week in a shop increases by 2369 units when there is a display as opposed to no display.

**Weighted Feature:** The number of units sold per week in a shop increases by 2369 units when there is a feature as opposed to no feature.

**Weighted Promotion:** The number of units sold per week in a shop decreases by 629 when there is a Promotion as opposed to no promotion.

**Interaction between Display and Promotion:** This produces a cancellation effect in number of units sold per week in a shop

**Interaction between Promotion and Feature:** This produces a cancellation effect in number of units sold per week in a shop

**Interaction between Display, Promotion and Feature:** This produces a synergistic effect in number of units sold per week in a shop

**Recommendation:**

Having just a promotion does not have good effect on sales. The effect on sales when promotion, feature and display combined has a far better result as opposed to their individual effects. Our Recommendation is to run promotions along with features and displays.

This insight can improve sales as opposed to sales obtained by having only display or feature.

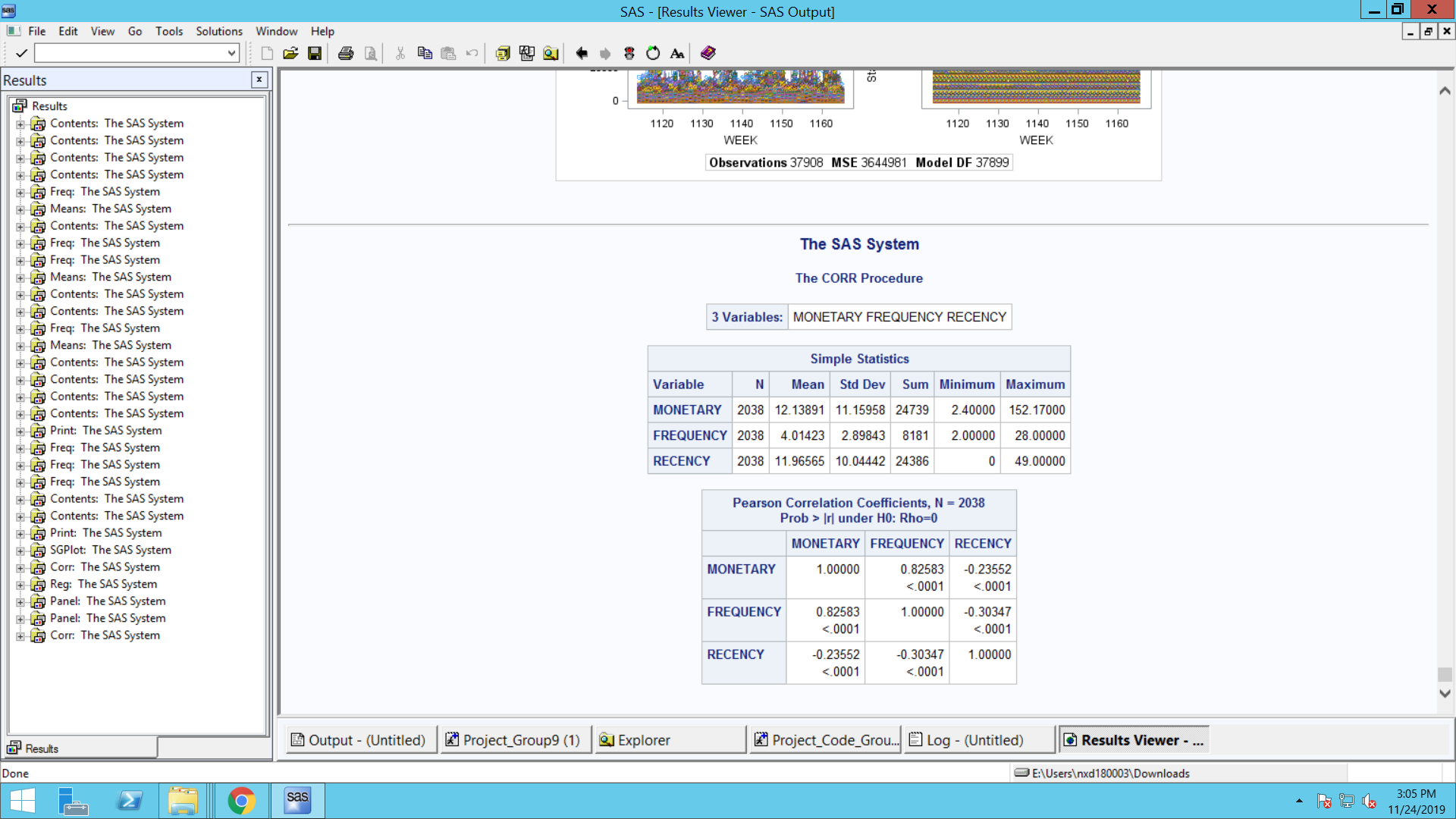
**3. PROBLEM STATEMENT-2**

Analyze the demographic features of the most valuable customers of the Skippy brand.

**Data Preparation & Analysis:**

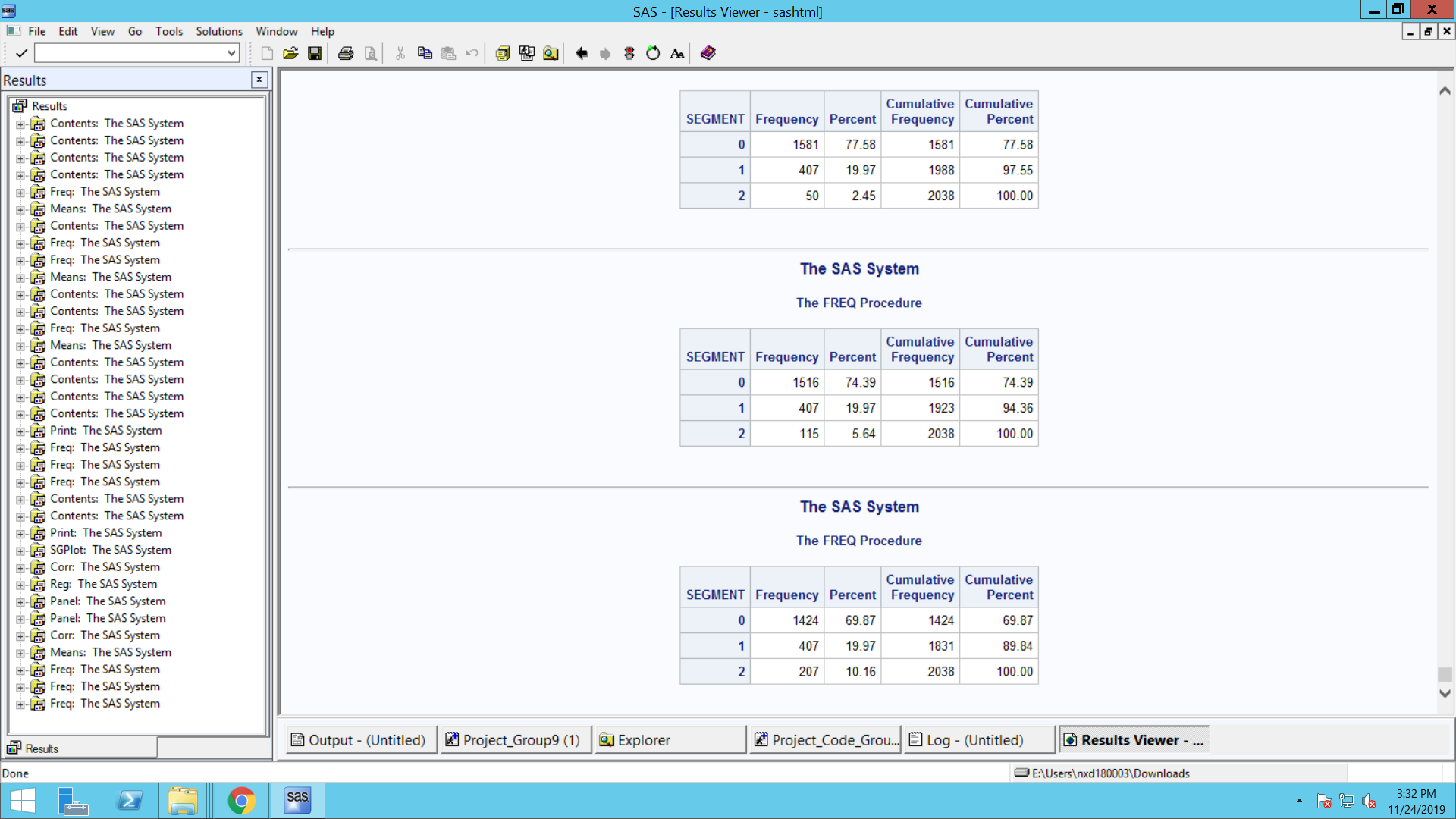
Panel Data has purchase information for all peanut butter products purchased in all stores. We filter out for Skippy and perform RFM analysis to segment customers who are most valuable with respect to the following 3 metrics:

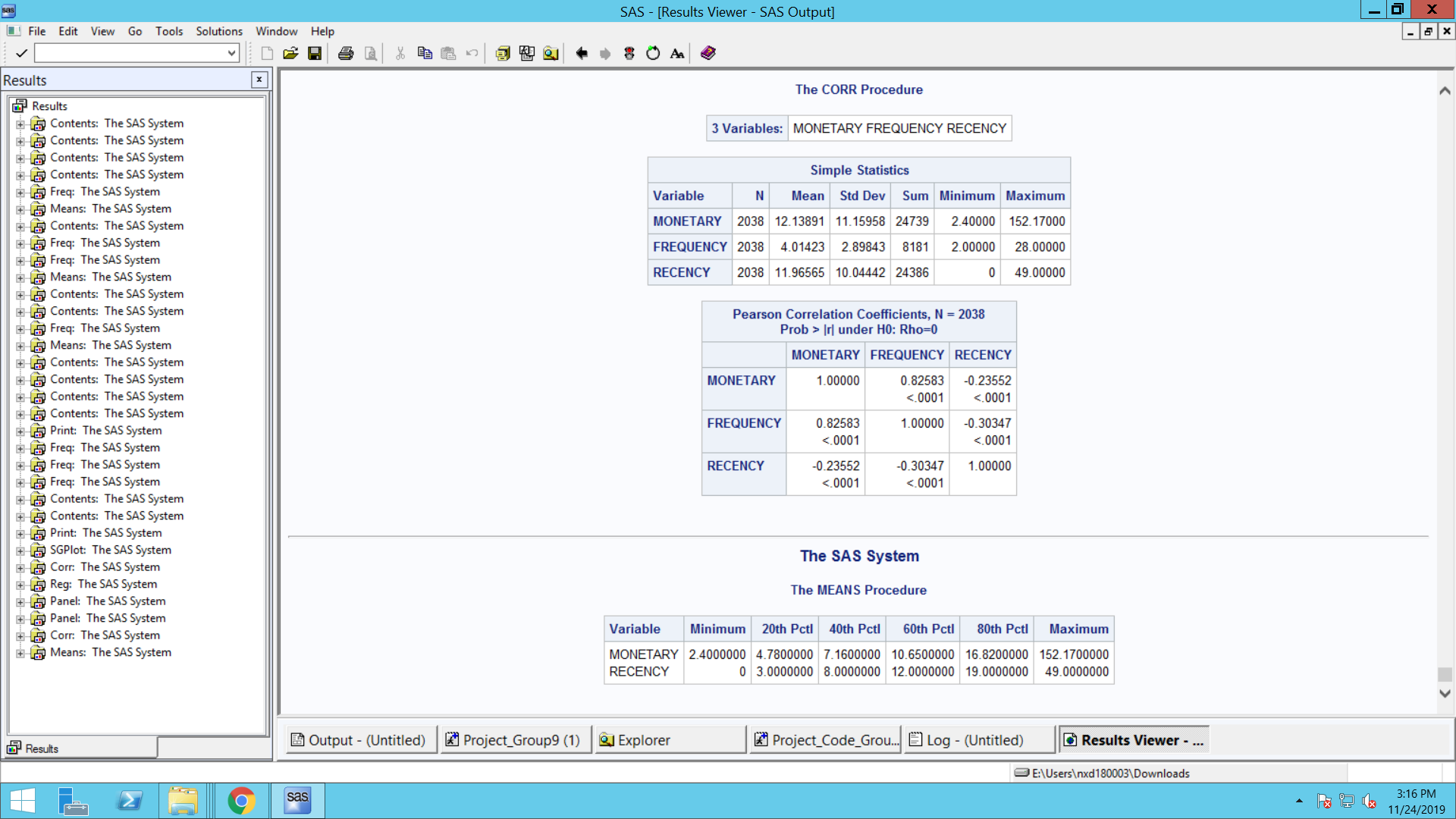
* Recency: To find out how recently a purchase has been made.
* Frequency: Number of weeks a customer has made a purchase
* Monetary: Amount spent by customers



From our initial analysis, we understand that a customer on average spent around $12 by making 4 visits every 12 weeks

Since Monetary and Frequency are highly correlated, we will use Monetary and Recency as a criterion to split our customers into 3 segments.





Using the Pareto principle, we assume that 20% of the customers contribute more revenue to the company than the rest. And this is confirmed by analysis where we find that the top 20% have the highest monetary return. So, we split the customers into 3 segments.

Segment 1 - Monetary > $16.82

Segment 2 - (Monetary > $4.78 AND Monetary<$16.82) AND (Recency>19)

Segment 0 - Rest

**For Segment 1** we mapped it to the demographics and found the average for each characteristics of a customer using dummy variables.

|  |  |  |
| --- | --- | --- |
| **Customer Demographics** | **Levels** | **Average** |
| Large Family Size | 4,5,6 | 44% |
| Normal Family Size | 1,2,3 | 55% |
| Low Income | 1,2,3,4 | 5% |
| Medium Income | 5,6,7,8 | 42% |
| High Income | 9,10,11,12 | 52% |
| Age of Male - Young | 1 | 0% |
| Age of Male - Mid | 2,3,4 | 48% |
| Age of Male - Old | 5,6 | 51% |
| Age of Female - Young | 1 | 0% |
| Age of Female - Mid | 2,3,4 | 54% |
| Age of Female - Old | 5,6 | 46% |
| Education - Male -School | 1,2,3 | 5% |
| Education - Male - College | 4,5,6 | 64% |
| Education - Male - Graduate | 7,8 | 31% |
| Education - Female -School | 1,2,3 | 3.5% |
| Education - Female - College | 4,5,6 | 38% |
| Education - Female - Graduate | 7,8 | 28% |
| One Child | 1,2,3 | 24% |
| Two Children | 4,5,6 | 10.5% |
| Three Children | 7 | 0% |
| No Children | 8 | 65% |
| Male - White Collar High Post | 1,2,3 | 42% |
| Male - White Collar Low Post | 4,5 | 4% |
| Male Blue Collar | 6,7,8,9 | 23% |
| Male No Occupation | 10,13 | 31% |
| Female - White Collar High Post | 1,2,3 | 37% |
| Female - White Collar Low Post | 4,5 | 16% |
| Female Blue Collar | 6,7,8,9 | 11% |
| Female No Occupation | 10,13 | 35% |
| Pets in House (Cats + Dogs) |  | 45% |

The following characteristics of top 20% Customers are as follows:

* 55% of the Normal family size households belong to high value customers
* Middle age women and Old Men prefer Skippy
* Men and women with white collar jobs prefer this brand
* Households with no children and have at least one pet (cat or dog) will more preferably buy this product

**Recommendation:**

Concentrating on the Top 20% and understanding their purchase history enables us to attract similar customers in the future. New customers with similar demographic features can be targeted to improve sales in long run. Special promotional strategies should be designed to target the other 80%, of which the primary share is below the 20% mark. Since that is the highest number of customers, we can increase sales by marketing to that subgroup of customers.

**4. PROBLEM STATEMENT-3**

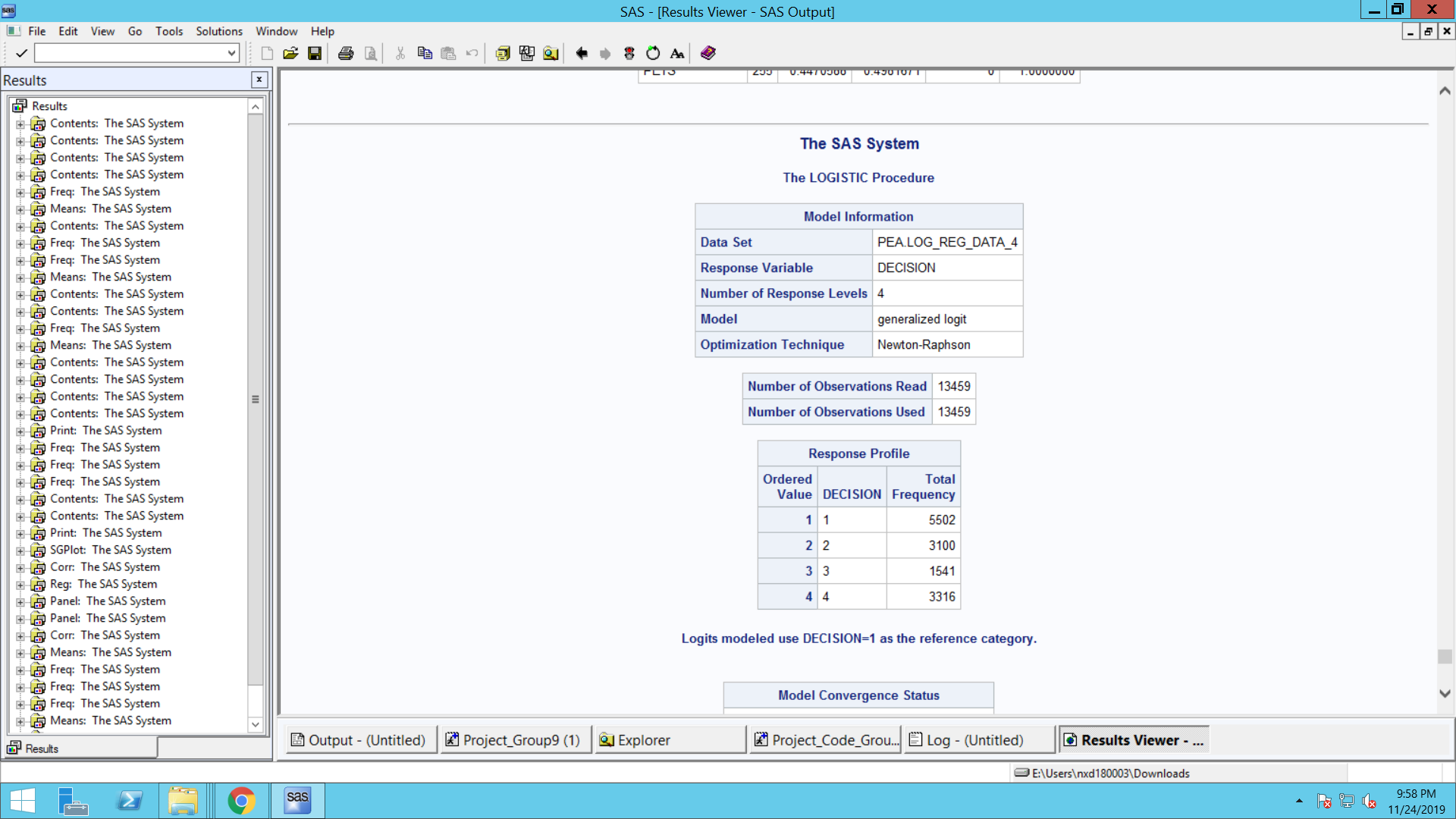
Evaluate the Brand preferences among different customers

**Analysis:**

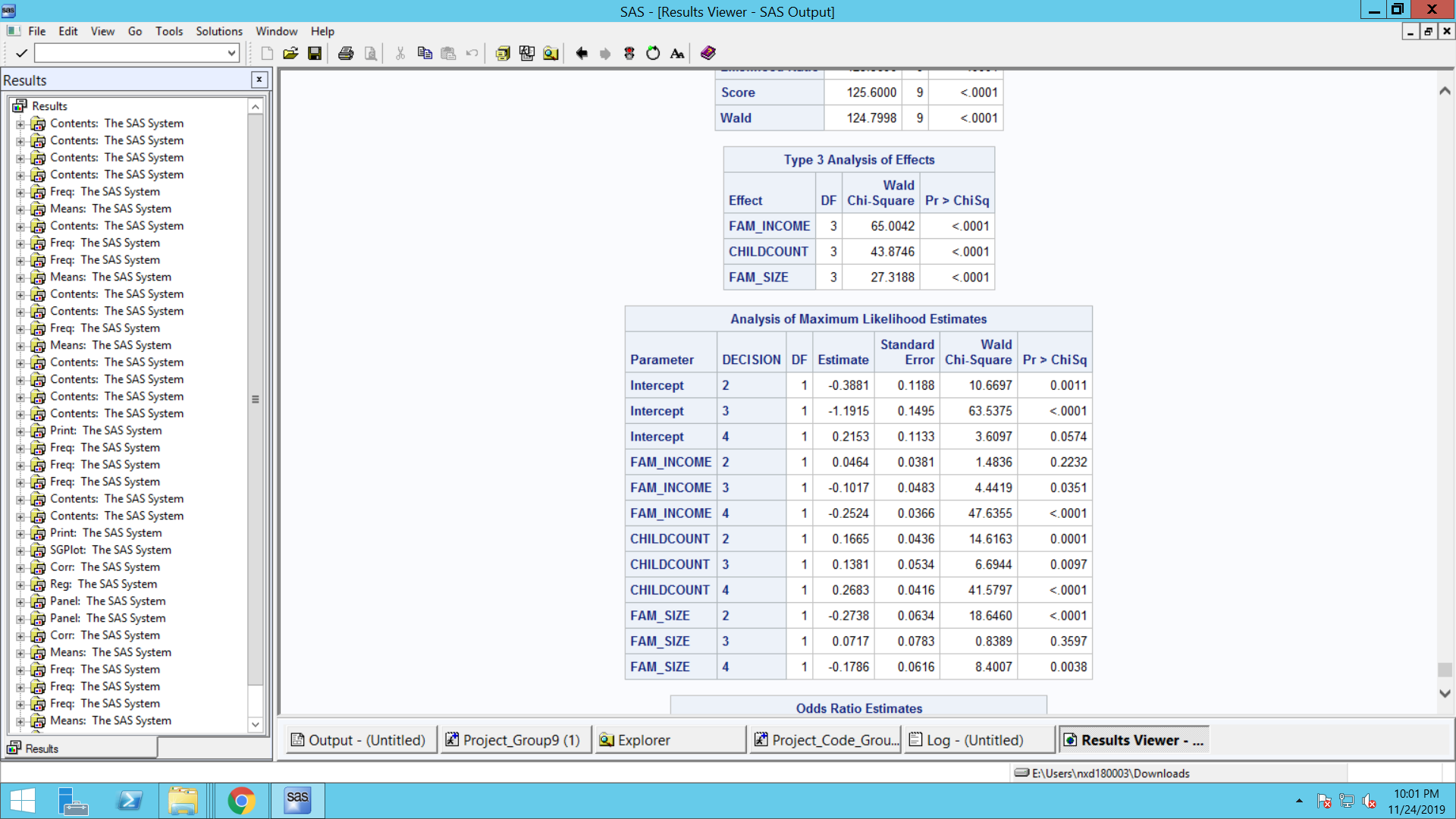
We build a Multinomial Logistic Regression for our analysis. The outcome

measure in this analysis is the preferred Brand – Skippy, Jif, Peter or Private.

Skippy is chosen as the reference category and is coded as 1, Jif as 2, Peter as 3 and Private as 4.



The DECISION number indicates to which model an estimate, standard error, chi-square, and p-value refer.



**MODEL 1: JIF RELATIVE TO SKIPPY**

**Intercept** – This is the multinomial logit estimate for JIF relative to Skippy when the predictor variables in the model are evaluated at zero. The logit for preferring JIF relative to Skippy is -0.38.

**FAM\_SIZE**: This is the multinomial logit estimate for when FAMILY SIZE

increases from regular (1,2,3) to large family size (4,5,6) for JIF relative to Skippy, given the other variables in the model are held constant.

If a customer were to move from regular (1,2,3) to large family size (4,5,6) the

multinomial log-odds for preferring JIF relative to Skippy would be expected to decrease by 0.27 unit while holding all other variables in the model constant.

**CHILDCOUNT**: If the number of children in a house increase by one, the multinomial log-odds for preferring JIF relative to Skippy would be expected to increase by 0.16 unit while holding all other variables in the model constant.

**MODEL 2: Peter relative to Skippy**

**Intercept:** This is the multinomial logit estimate for Peter Pan relative to

Skippy when the predictor variables in the model are evaluated at zero. The logit for preferring Peter Pan relative to Skippy is -1.19.

**FAM\_INCOME**: This is the multinomial logit estimate for when FAMILY INCOME bracket increases from low to medium or medium to high Peter relative to Skippy, given the other variables in the model are held constant.

If a customer were to move from lower income bracket to higher income bracket the multinomial log-odds for preferring Peter relative to Skippy, would be expected to decrease by 0.103 unit while holding all other variables in the model constant.

**CHILDCOUNT**: If the number of children in a house increase by one the multinomial log-odds for preferring Peter relative to Skippy would be expected to increase by 0.14 unit while holding all other variables in the model constant.

**MODEL 3: Private relative to Skippy**

**Intercept:** This is the multinomial logit estimate for Private Label relative to

Skippy when the predictor variables in the model are evaluated at zero. The logit for preferring Private Label to Skippy is 0.21.

**FAM\_INCOME**: This is the multinomial logit estimate for when FAMILY INCOME bracket increases from low to medium or medium to high for Private relative to Skippy, given the other variables in the model are held constant.

If a customer were to move from lower income bracket to higher income bracket the multinomial log-odds for preferring Private to Skippy would be expected to decrease by 0.25 unit while holding all other variables in the model constant.

**FAM\_SIZE**: This is the multinomial logit estimate for when FAMILY SIZE increases from regular (1,2,3) to large family size (4,5,6) for Private relative to Skippy, given the other variables in the model are held constant.

If a customer were to move from regular (1,2,3) to large family size (4,5,6) the multinomial log-odds for preferring Private relative to Skippy would be expected to decrease by 0.17 unit while holding all other variables in the model constant.

**CHILD COUNT**: If the number of children in a house increase by one the multinomial log-odds for preferring Private relative to Skippy would be expected to increase by 0.26 unit while holding all other variables in the model constant.

**Recommendations:**

With an increase in income and family size, Skippy is preferred. But to sustain increase in consumption from Families, Skippy needs to target children as it is clearly seen that other brands are preferred in the houses with children.

Skippy can experiment with children-friendly product lines or try offering free goodies with the product or have a celebrity endorse the product.