

Market Cluster Analysis: Be Customer Wise or Otherwise

Prepared by

Tong Wen Liang, Samuel

Table of Contents

1.0 Deciding on Clustering Variables, Data Exploration and Preparation	3
1.1 Ensuring Proper Data and Modelling Type	3
1.2 Removing Unnecessary Variables	3
1.3 Reducing Data Dimensionality	5
1.4 Deciding on Clustering Variables	6
2.0 Standardization and Transformation	7
2.1 Transforming Scaled Variables based on the JMP Recommendation Method	12
2.2 Transforming Variables based on Johnson Transformation	14
2.3 Comparing Distribution of Individually Transformed Variables and Johnson Transformed Variables	16
3.0 Performing Cluster Analysis – K Means Cluster	17
4.0 Performing Cluster Analysis – Latent Class Analysis	21
4.1 Interactive Binning	21
4.2 Latent Class Analysis	22
5.0 Validate and Interpret the Clusters	23
5.1 Value ordering of all categorical variables used for LCA	23
5.2 Cluster Summary.....	23
5.2.1 Description of Clusters:.....	25
5.3 MDS Plot	27
5.4 Interpretation of Clusters by Region.....	28
5.5 Interpretation of Clusters by Total Household Income	28
5.6 Interpretation of Clusters by Total Household Income and Region	29

1.0 Deciding on Clustering Variables, Data Exploration and Preparation

1.1 Ensuring Proper Data and Modelling Type

Based on the obtained raw data, we can see that there are several data without the proper data and modelling type. Therefore, we will make changes to the data and modelling type to accurately reflect the actual data and modelling type.

The changes made are as reflected in the table below:

No.	Names of Variables	Changed From	Changed To
1	Agricultural Household Indicator	Numeric/Continuous	Numeric/Nominal
2	Household Head Age		
3	Total Number of Family Members		
4	Members with Age Less Than 5-Year-Old		
5	Members with Age 5 – 17 Years Old		
6	Total Number of Family Members Employed		
7	Number of Bedrooms		
8	Electricity		
9	Number of Television		
10	Number of CD / VCD / DVD		
11	Number of Component / Stereo Set		
12	Number of Refrigerator / Freezer		
13	Number of Washing Machine		
14	Number of Air conditioner		
15	Number of Car, Jeep, Van		
16	Number of Landline / Wireless Telephones		
17	Number of Cellular Phones		
18	Number of Personal Computer		
19	Number of Stove with Oven / Gas Range		
20	Number of Motorized Banca		

1.2 Removing Unnecessary Variables

Prior to the performance of Cluster Analysis, it would be necessary to identify the variables that we need to use for our market segmentation analysis. Since an international retailing giant is planning their market development strategy in the Philippines, it is imperative that the selected variables used for cluster analysis of households are pertinent to the retail business environment.

In this case, we are interested in the expenditure variables that are of relevance to the retail industry. Therefore, unnecessary variables will be excluded from our cluster analysis. To identify which variables are to be excluded, we have referred to the questionnaires included in the “Final Report: 2012 Family Income and Expenditure Survey”¹ to identify what information are encapsulated in each expenditure variable.

The following details in the table below indicate the variables that have been excluded as well as the reason(s) for exclusion:

¹ Final Report: 2012 Family Income and Expenditure Survey –
<https://psa.gov.ph/sites/default/files/2012%20FIES%20%28Final%20Report%29.pdf>

No.	Type of Expenditure	Reason(s) for Exclusion
1	Housing and Water Expenditure	The expense includes the following elements that cannot be sold in a retail store: 1. Actual Rentals for Housing
2	Transportation Expenditure	2. Purchase of Vehicles
3	Communication Expenditure	3. Postal Services
4	Medical Care Expenditure	4. Medical Consultation
5	Education Expenditure	5. Tuition Fees
6	Restaurant and Hotels Expenditure	6. Catering Services 7. Accommodation Services
7	Miscellaneous Goods and Services Expenditure	8. Personal Care Services (i.e. Hairdressing)
8	Crop Farming and Gardening Expenses	9. Wages of Hired Labor 10. Irrigation Fees
9	Special Occasions Expenditure	No field definitions were available in the report. Therefore, the expenditure should not be included to ensure that variables chosen are consistent with the goods and services provided by a retail store.
10	Total Food Expenditure	Please see “Reason for Excluding Total Food Expenditure” below

Reason for Excluding Total Food Expenditure

Based on the survey questionnaire, we noted that the Total Food Expenditure includes consumption and expenses on the following food categories that are not reflected in the data provided:

1. Milk, Cheese and Eggs
2. Oils and Fats
3. Sugar, Jam, Honey, Chocolate and Confectionery
4. Food Products Not Elsewhere Classified

Additionally, to confirm that the amount of “Total Food Expenditure” variable does not commensurate with the total amount of the other food related variables, we created “Total Grocery Expenditure” column with the following formula:

$$\begin{aligned}
 & \text{Bread and Cereals Expenditure} \\
 & + \text{Total Rice Expenditure} \\
 & + \text{Meat Expenditure} \\
 & + \text{Total Fish and marine products Expenditure} \\
 & + \text{Fruit Expenditure} \\
 & + \text{Vegetables Expenditure}
 \end{aligned}$$

Thereafter, we tabulated and compared the total amount in column “Total Food Expenditure” and “Total Grocery Expenditure” and noted that they are not equal, thus confirming that “total Food Expenditure” does not fully capture all other food related variables in the data set.

Total Grocery Expenditure	Sum	2989353903
Total Food Expenditure	Sum	3535359428

We then created a new column to find the differences between “Total Food Expenditure” and “Total Grocery Expenditure” with the following formula:

Total Food Expenditure - *Total Grocery Expenditure*

By tabulating this column to find both Max and Min of the amount, we noted that there were both positive and negative amount. Therefore, this concludes that food related variables might be under- or over-included.

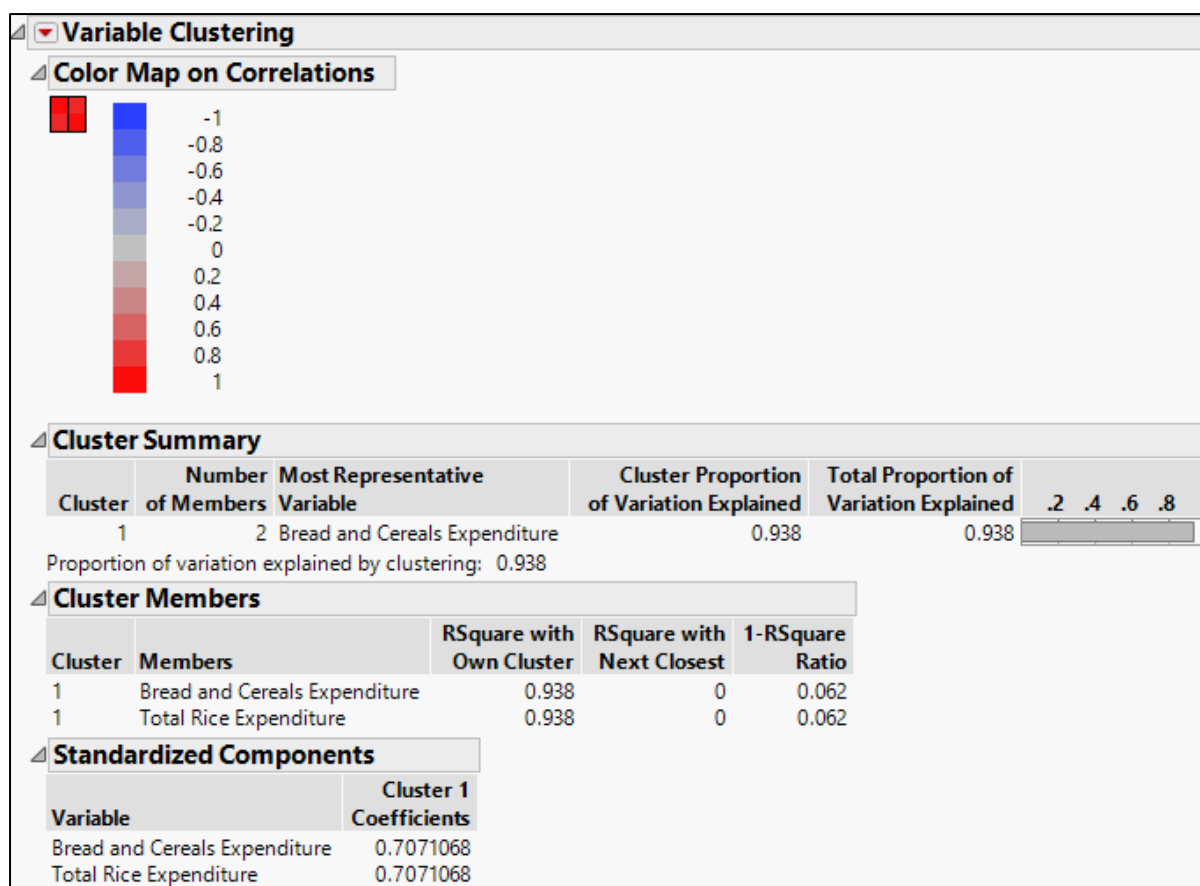
Total Food Expenditure - Total Grocery Expenditure	Max	712364
	Min	-728589

1.3 Reducing Data Dimensionality

Once we are satisfied with the chosen variables to conduct cluster analysis, we will proceed with reducing data dimensionality by removing all but one co-variated variables.

Pairwise Correlations															
Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	-.8	-.6	-.4	-.2	0	.2	.4	.6	.8
Total Rice Expenditure	Bread and Cereals Expenditure	0.8763	41544	0.8740	0.8785	<.0001*									
Vegetables Expenditure	Total Fish and marine products Expenditure	0.4856	41544	0.4782	0.4929	<.0001*									
Clothing, Footwear and Other Wear Expenditure	Meat Expenditure	0.4763	41544	0.4688	0.4837	<.0001*									
Vegetables Expenditure	Meat Expenditure	0.4743	41544	0.4669	0.4818	<.0001*									
Total Fish and marine products Expenditure	Meat Expenditure	0.4654	41544	0.4578	0.4729	<.0001*									
Vegetables Expenditure	Fruit Expenditure	0.4382	41544	0.4304	0.4460	<.0001*									
Fruit Expenditure	Meat Expenditure	0.4194	41544	0.4114	0.4273	<.0001*									
Clothing, Footwear and Other Wear Expenditure	Fruit Expenditure	0.3947	41544	0.3865	0.4028	<.0001*									
Total Fish and marine products Expenditure	Bread and Cereals Expenditure	0.3914	41544	0.3832	0.3995	<.0001*									

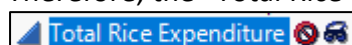
Based on a multivariate analysis of the chosen variables, we can see that the “Total Rice Expenditure” and “Bread and Cereals Expenditure” are highly correlated at 0.8763, which is higher than our set threshold of 0.8. Thereafter, we performed variable clustering to identify the most representative variable to be kept.



Based on the diagram above, we can see that the most representative variable is the “Bread and Cereal Expenditure”.

Additionally, as indicated above under the section “Reason for Excluding Total Food Expenditure”, we have noted from the “Final Report: 2012 Family Income and Expenditure Survey” that amount of expenditure and consumption of rice is already included under “Bread and Cereals Expenditure”.

Therefore, the “Total Rice Expenditure” has been excluded from our Cluster Analysis.



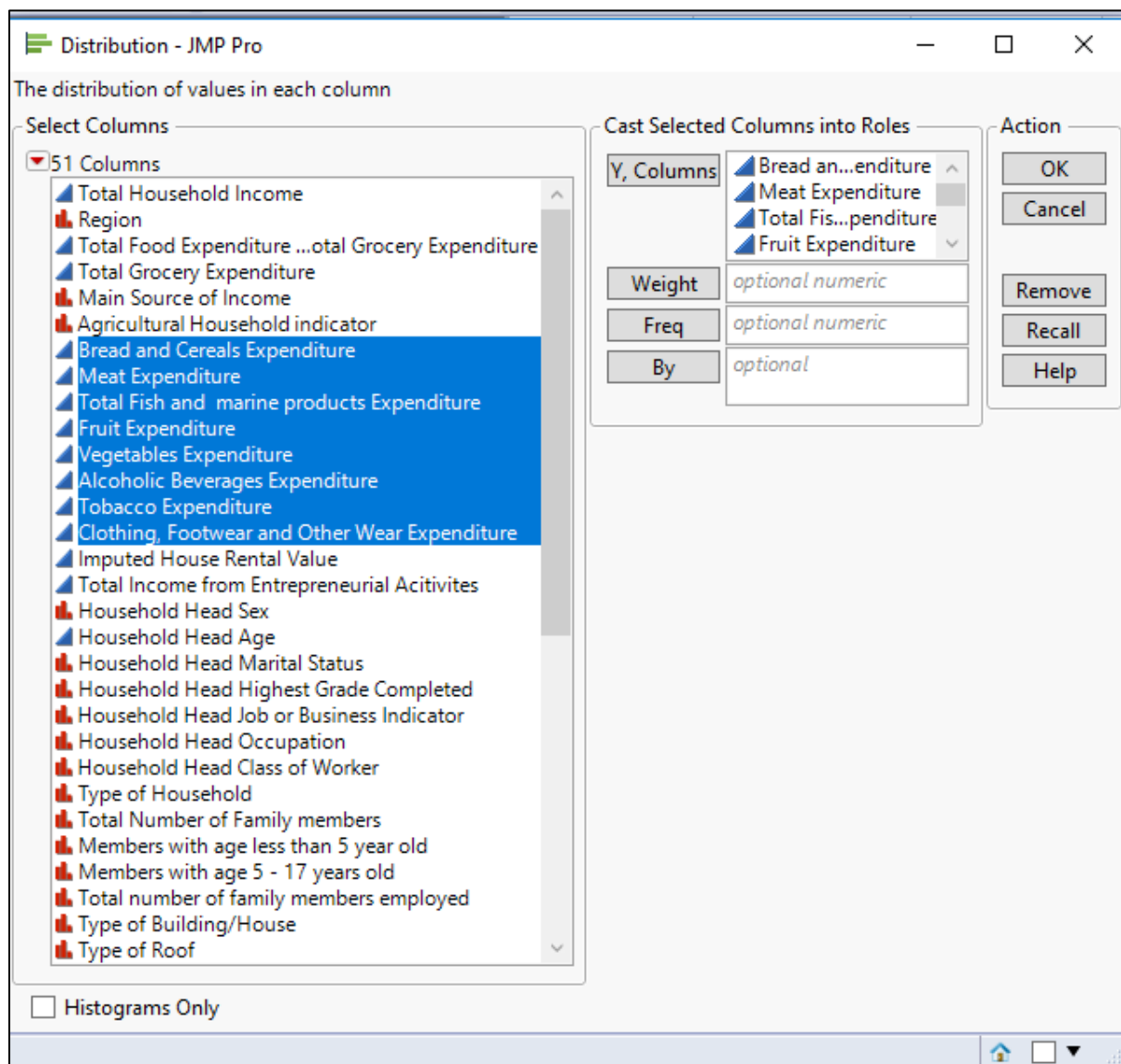
1.4 Deciding on Clustering Variables

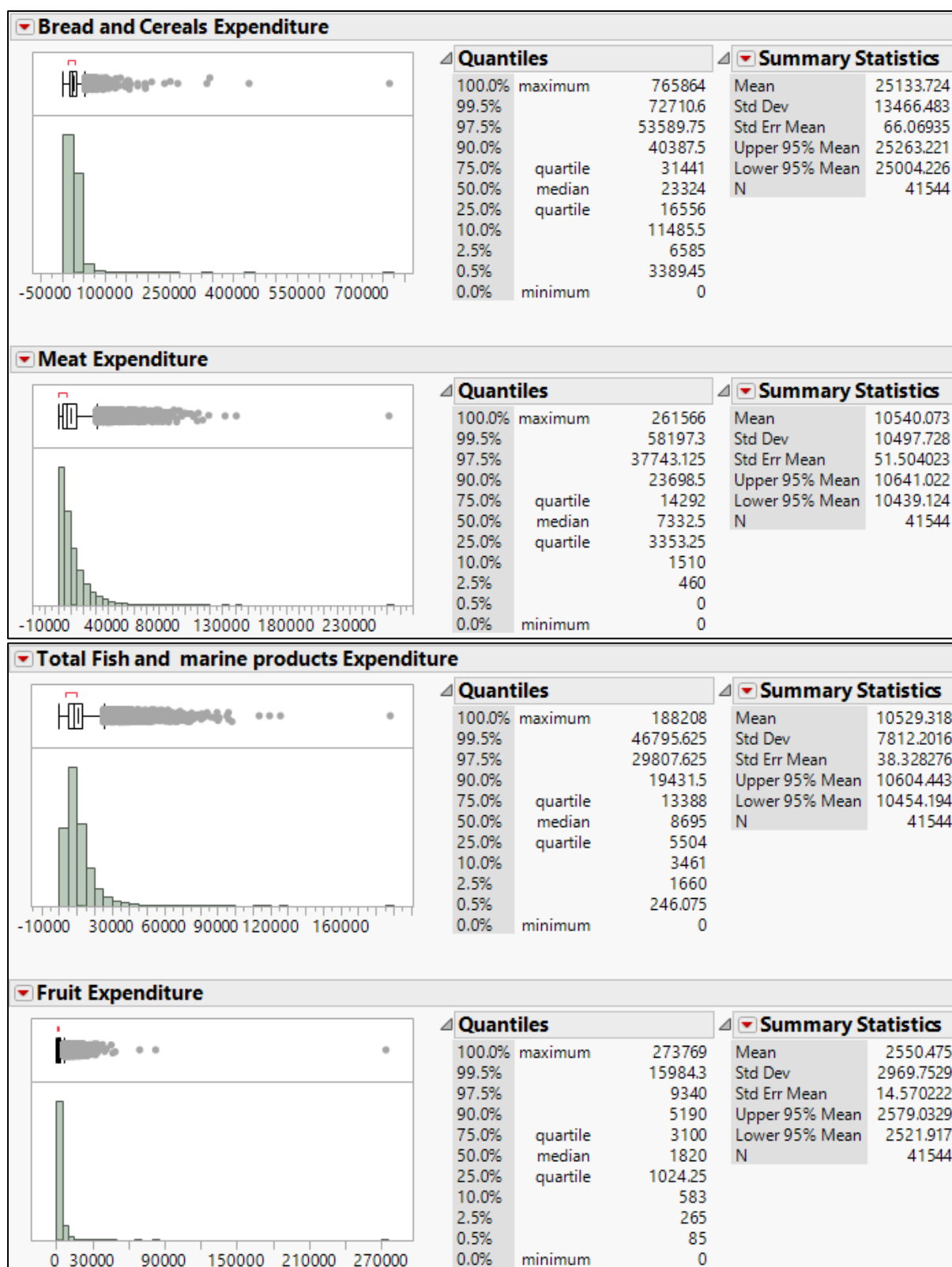
Next, once we have removed all unnecessary variables, we derived the following expenditure variables to be used for our clustering analysis:

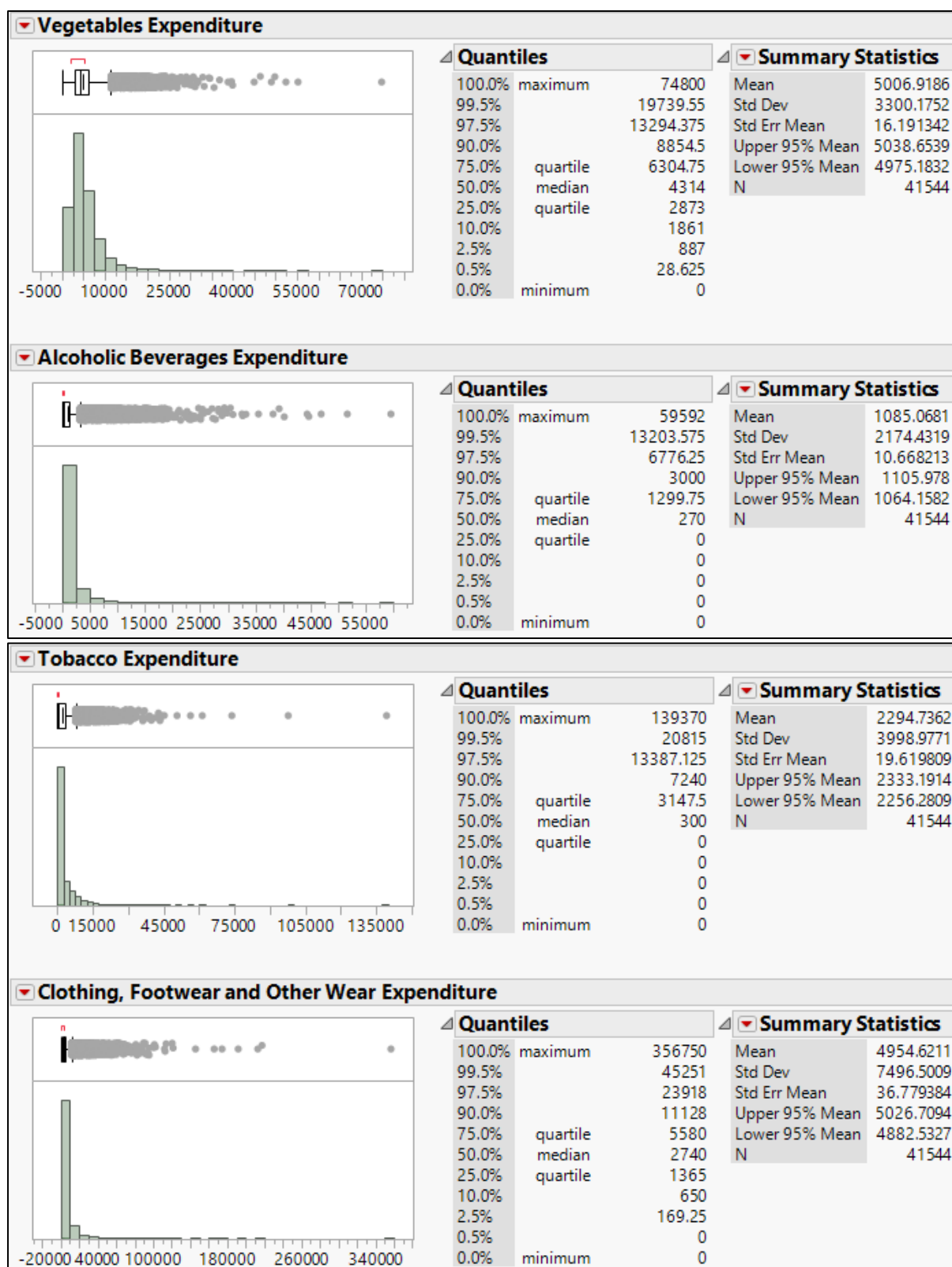
No.	Variables to be Included
1	Bread and Cereals Expenditure
2	Meat Expenditure
3	Total Fish and Marine Products Expenditure
4	Fruit Expenditure
5	Vegetables Expenditure
6	Alcoholic Beverages Expenditure
7	Tobacco Expenditure
8	Clothing, Footwear and Other Wear Expenditure

2.0 Standardization and Transformation

We will now perform distribution analysis of our chosen variables to make inferences about how the variables are distributed.



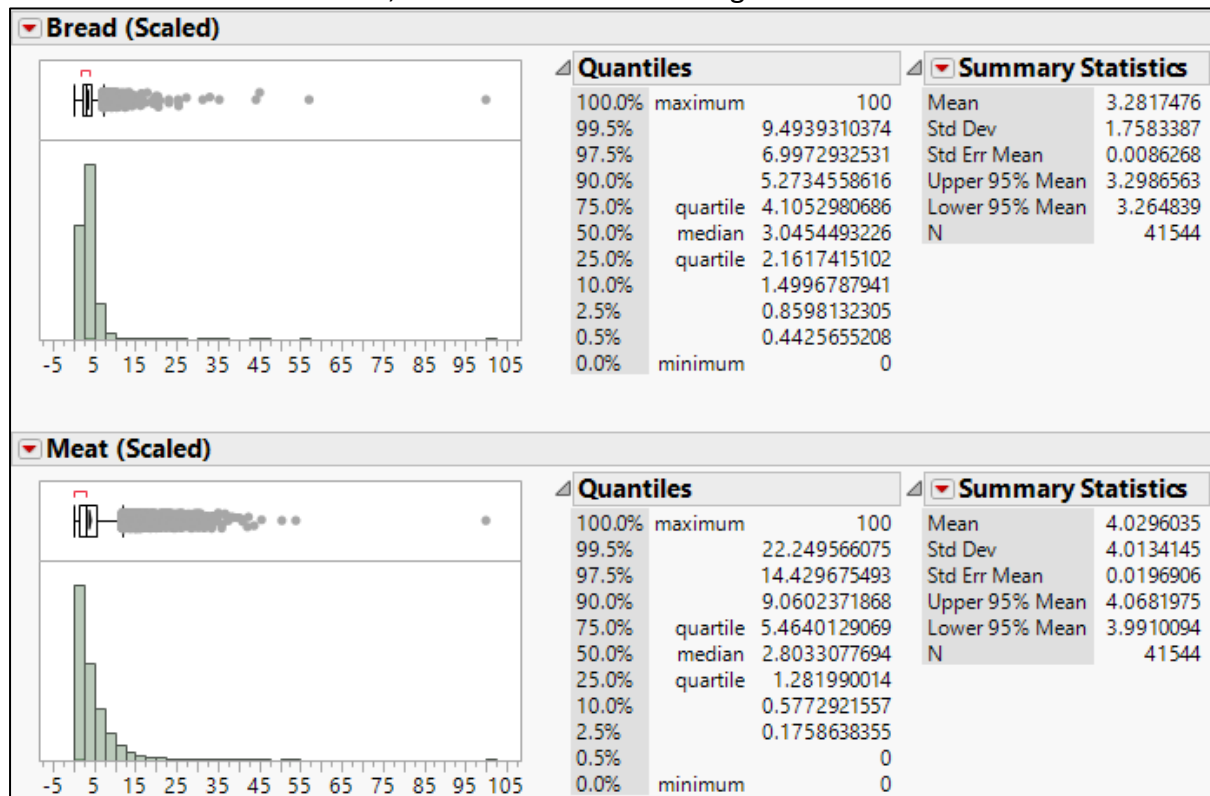


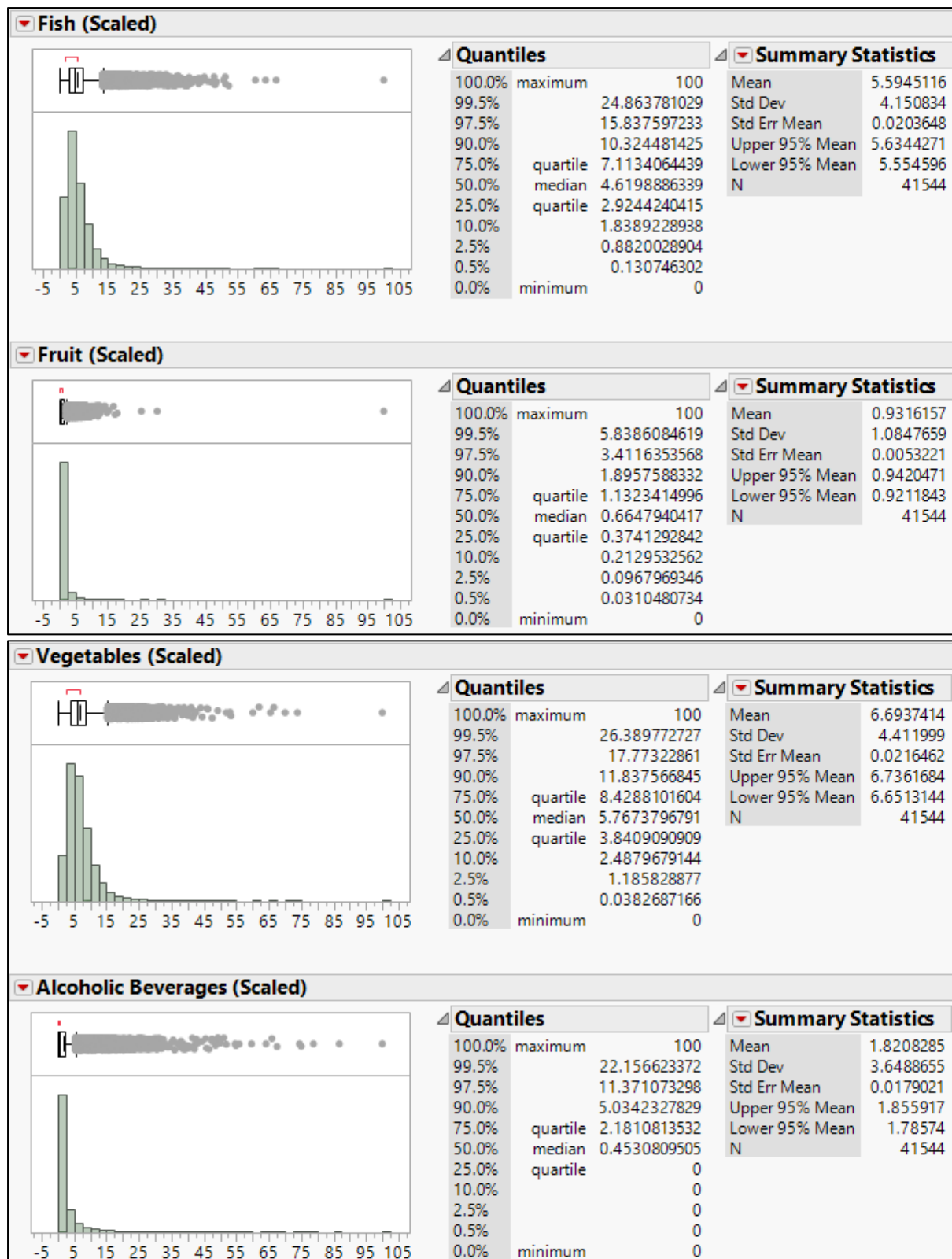


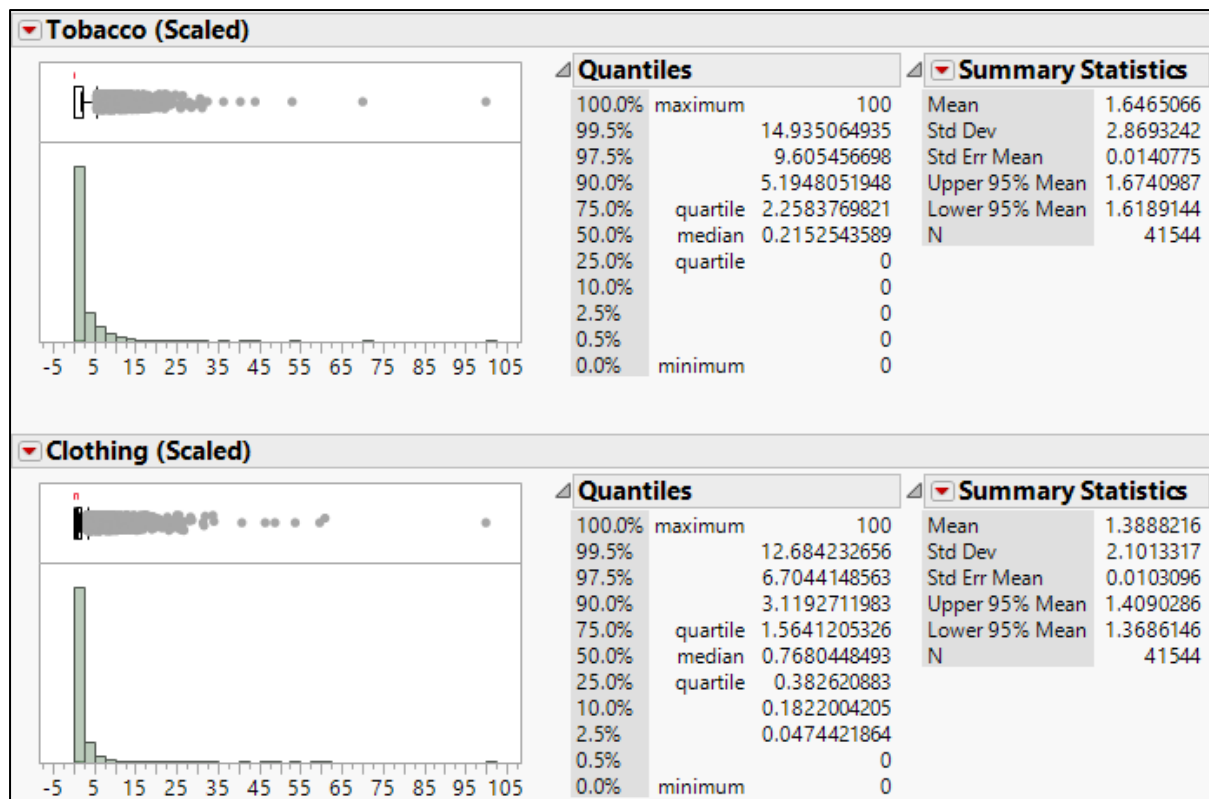
We can see that all the eight variables are not of the same scale. Therefore, the following formula will be applied to scale them:

$$\left(\frac{\left(\text{Bread and Cereals Expenditure} - \text{Col Minimum}(\text{Bread and Cereals Expenditure}_n) \right)}{\left(\text{Col Maximum}(\text{Bread and Cereals Expenditure}_n) - \text{Col Minimum}(\text{Bread and Cereals Expenditure}_n) \right)} \right) \cdot 100$$

Based on the scaled variables, we obtained the following distributions of the variables:





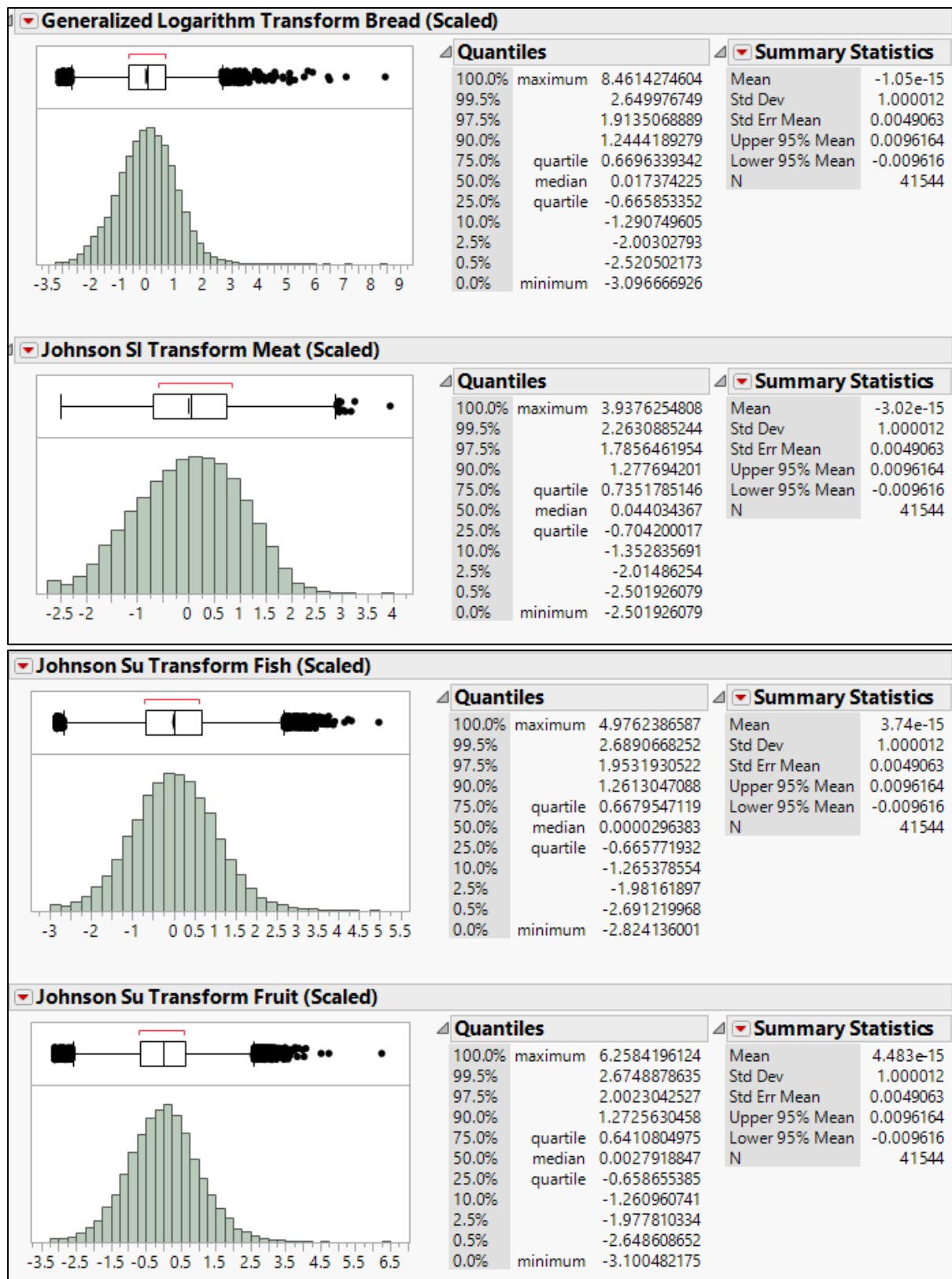


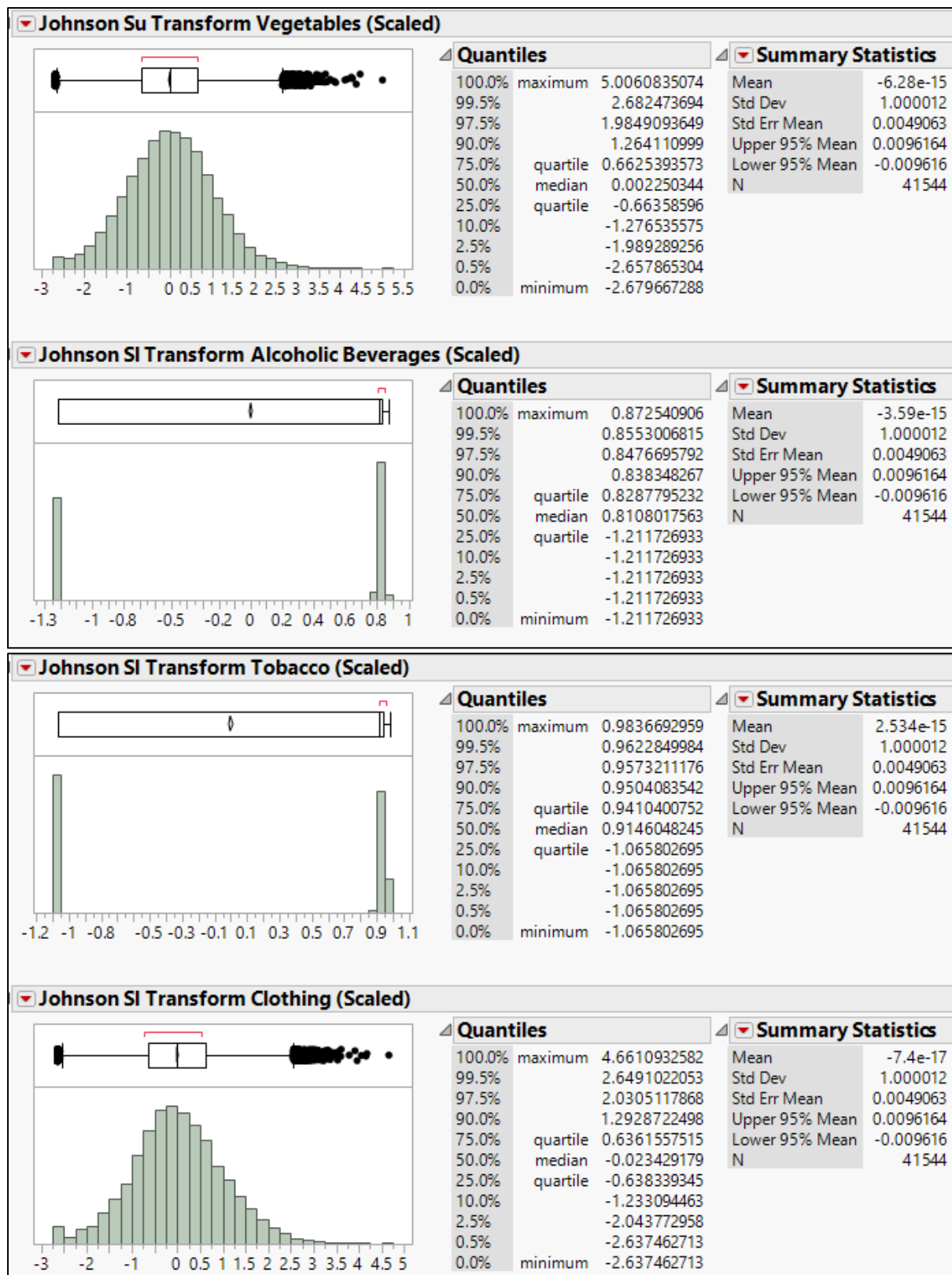
With reference to the distribution analysis of the scaled variables above, we noted that all variables are skewed. Therefore, we will be performing two types of transformation to identify which transformed variables will be the most useful to stabilize the variances to make errors more uniformed.

We will first proceed with transforming the variables based on the recommended transformation methods by JMP.

2.1 Transforming Scaled Variables based on the JMP Recommendation Method

The following depicts the distribution analysis of the individually transformed variables based on the JMP recommendation method:

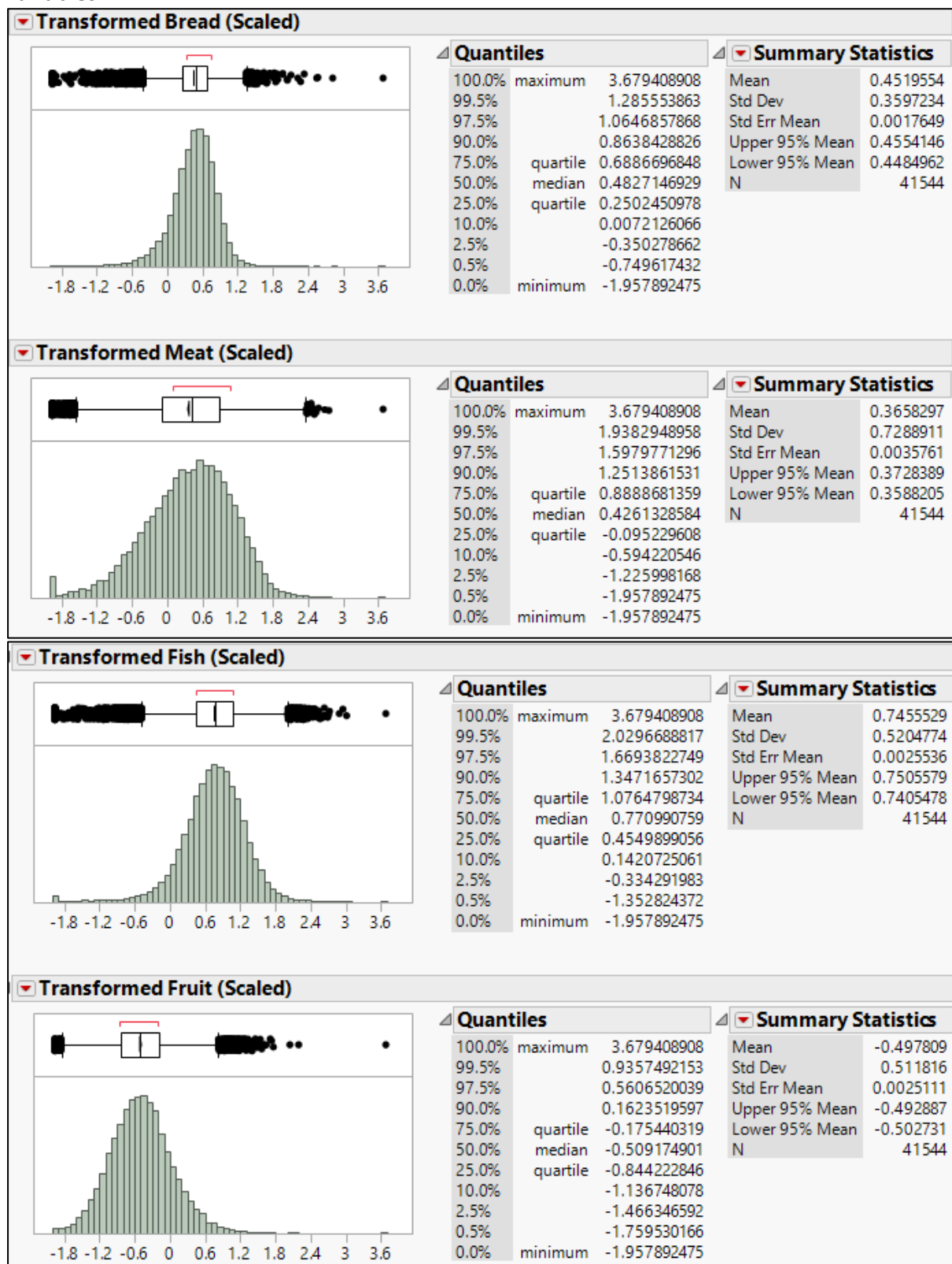


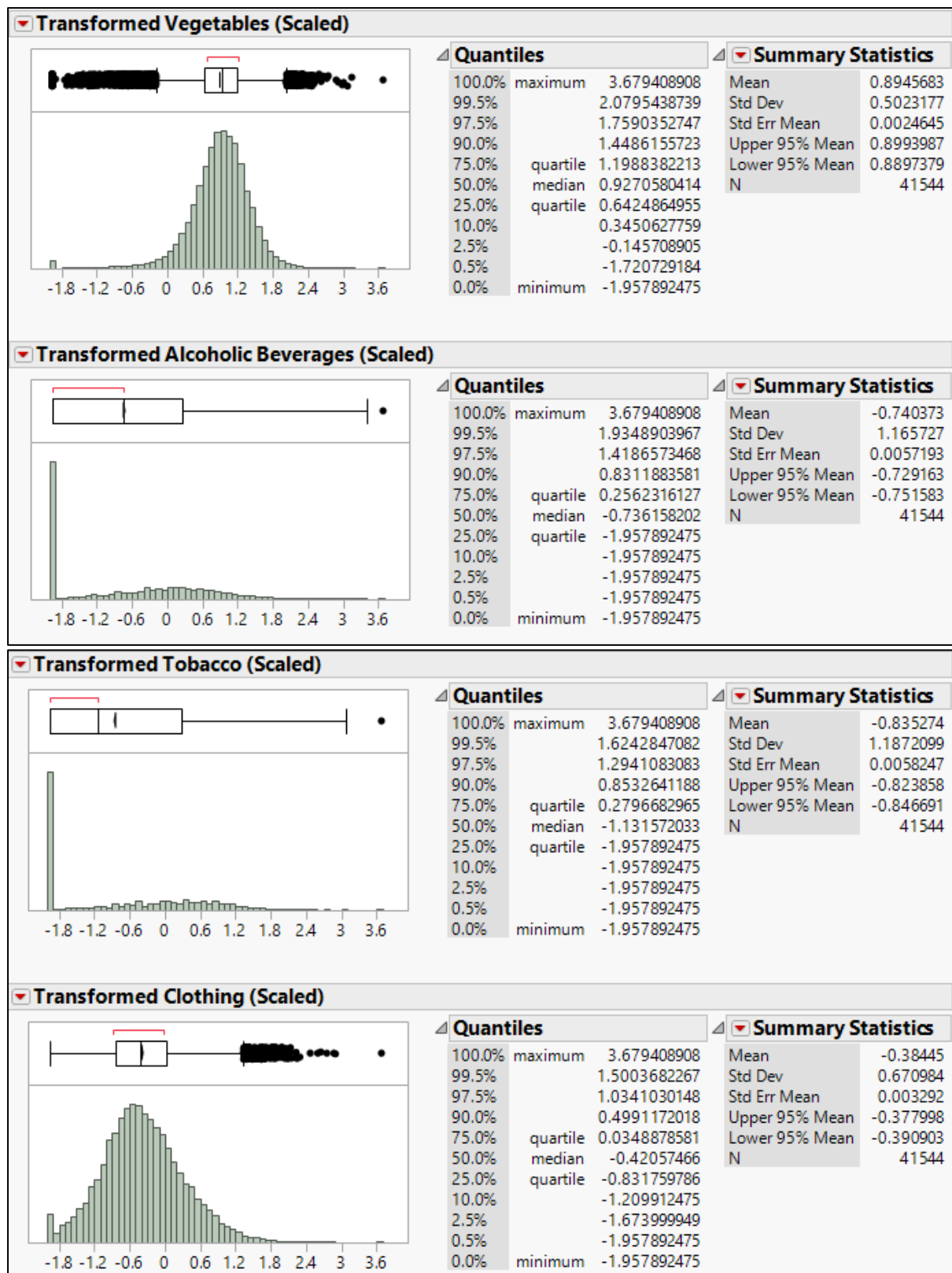


2.2 Transforming Variables based on Johnson Transformation

Next, we will be performing a global transformation of all variables based on the Johnson transformation function under the K-Means Clustering tool, which transform variables using Johnson SU or Johnson SB.

Upon transformation, we will now perform a distribution analysis of the transformed variables.





2.3 Comparing Distribution of Individually Transformed Variables and Johnson Transformed Variables

Based on the results obtained above, we can see that the Johnson transformed Alcoholic Beverages Expenditure and the Johnson transformed Tobacco Expenditure show a better

distribution than the individually transformed Alcoholic Beverages Expenditure and Tobacco Expenditure, based on JMP's recommended transformation.

3.0 Performing Cluster Analysis – K Means Cluster

Therefore, we will now proceed with K Means Clustering based on the Johnson transformed variables that we have saved into our data table from the earlier steps.

Iterative Clustering

Columns have common scale

Control Panel

Outlier cleanup: Declutter

Method: K-Means Clustering

Number of Clusters: 6 Range of Clusters (Optional): 25

Go Help

☐ Single Step
☐ Use within-cluster std deviations
☐ Shift distances using sampling rates

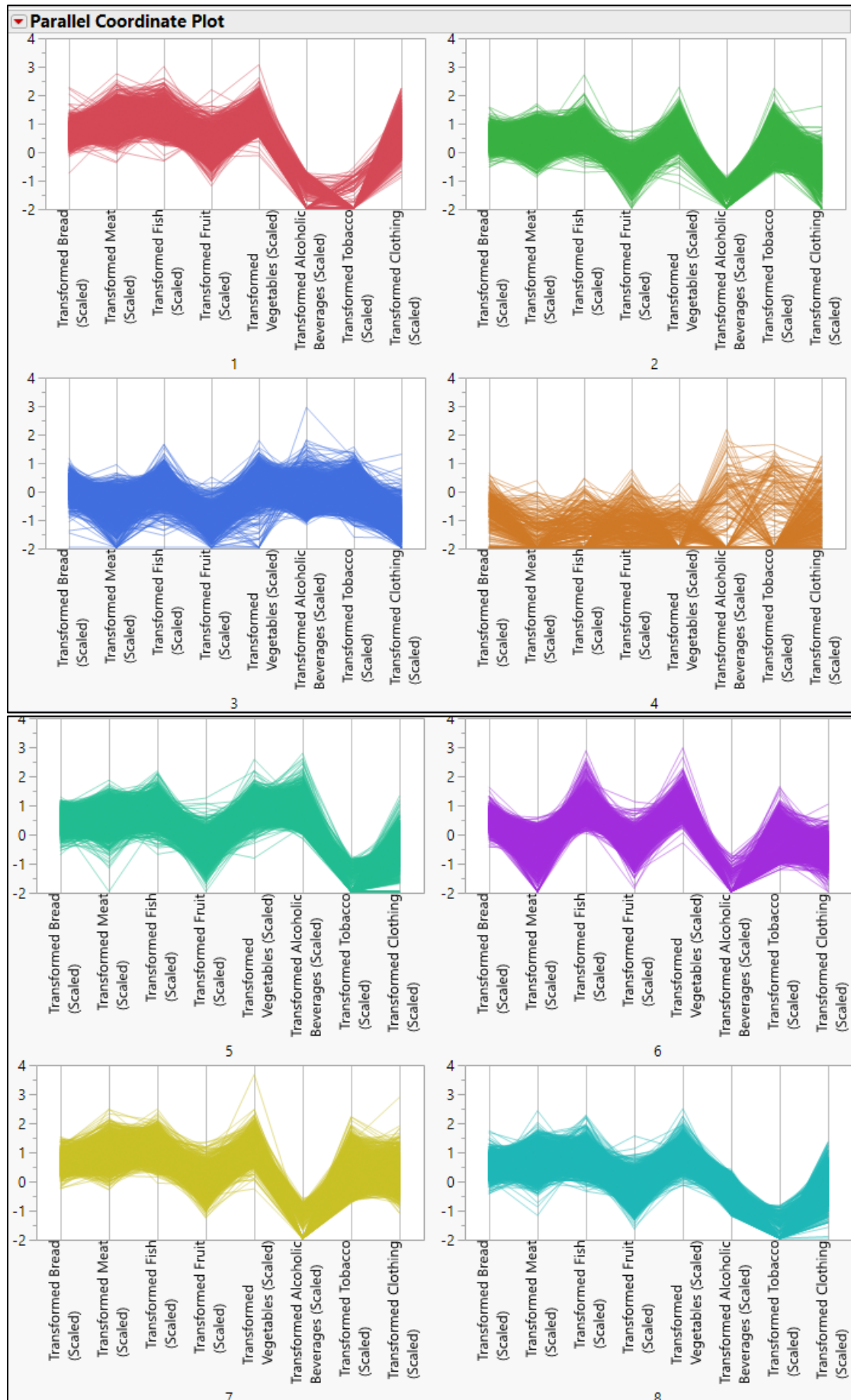
Cluster Comparison

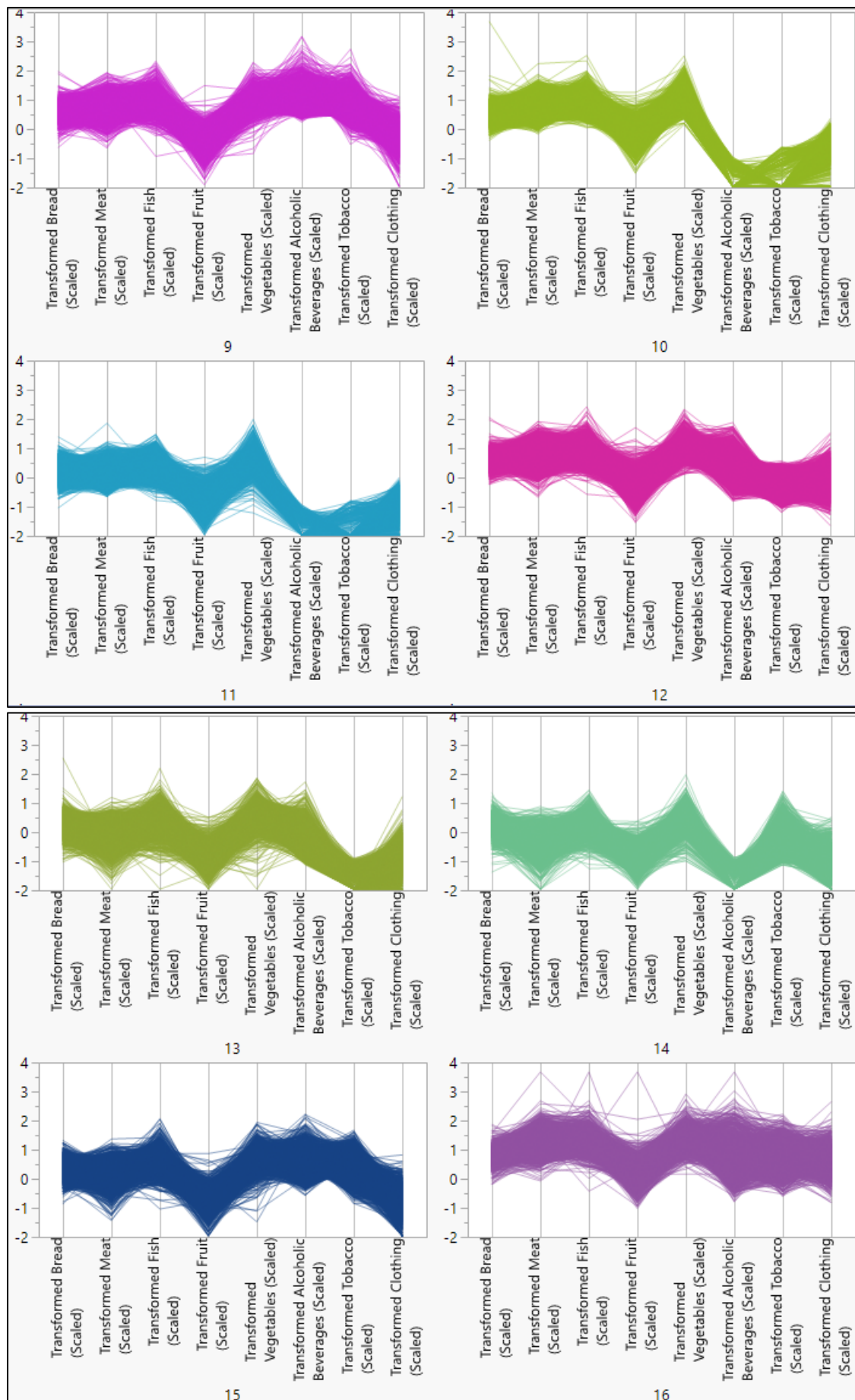
Method	NCluster	CCC	Best
K-Means Clustering	6	23.8146	
K-Means Clustering	7	22.8058	
K-Means Clustering	8	27.5477	
K-Means Clustering	9	30.5666	
K-Means Clustering	10	33.2238	
K-Means Clustering	11	34.5597	
K-Means Clustering	12	32.2161	
K-Means Clustering	13	34.0963	
K-Means Clustering	14	37.0193	
K-Means Clustering	15	34.7556	
K-Means Clustering	16	38.2044	
K-Means Clustering	17	34.863	
K-Means Clustering	18	41.8486	
K-Means Clustering	19	42.9453	
K-Means Clustering	20	38.6358	
K-Means Clustering	21	42.7432	
K-Means Clustering	22	44.2377	Optimal CCC
K-Means Clustering	23	42.423	
K-Means Clustering	24	40.098	
K-Means Clustering	25	43.1174	

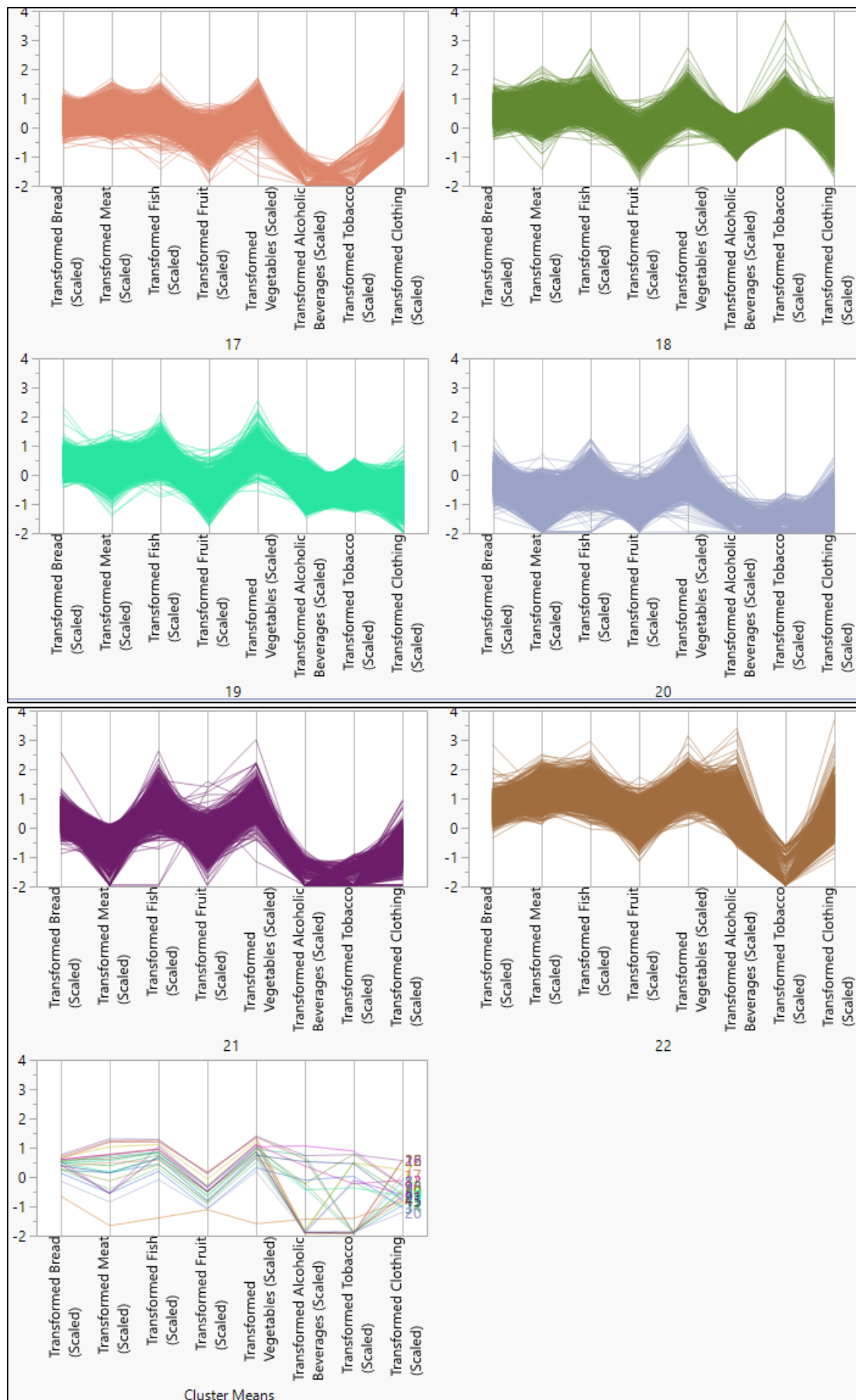
Columns have common scale

Based on the Optimal CCC, we have obtained 22 clusters as the best number of clusters, with the Cluster Summary and Parallel Coordinate Plot shown indicated below:

Cluster Summary			
Cluster	Count	Step	Criterion
1	2136	112	0
2	1947		
3	1301		
4	325		
5	2093		
6	1334		
7	1291		
8	2523		
9	1865		
10	2446		
11	2553		
12	2177		
13	2214		
14	1528		
15	2339		
16	1639		
17	2133		
18	2353		
19	2400		
20	1552		
21	1713		
22	1682		

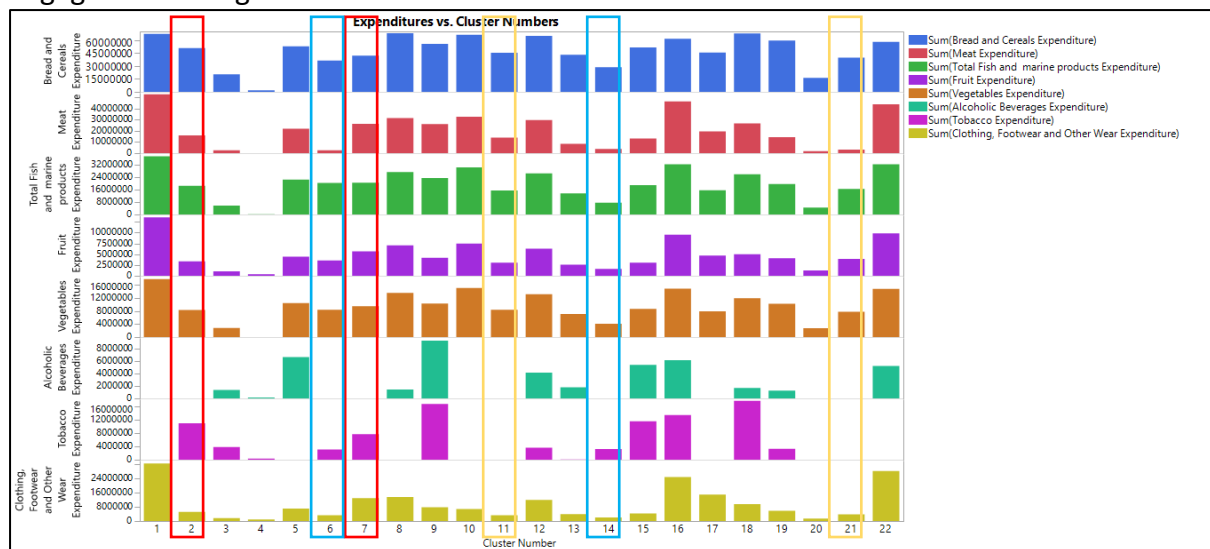






From the results, we can see that Clusters 1, 4, 10 and 17 show an “X” pattern, resulting in a bimodal distribution for variables “Alcoholic Beverages” and “Tobacco”. This indicates that there are households that consumes only either one of the product, both products, or none at all. Therefore, it would be difficult to define a set characteristic for households in these clusters with regards to both “Alcoholic Beverages” and “Tobacco” variables.

Additionally, when households are differentiated into 22 different clusters, it could be problematic when it comes to differentiating specific characteristics due to each cluster. For example, clusters 2 and 7, 6 and 14, and 7 and 21 seems to display similar characteristics as seen in the bar chart below. With numerous clusters displaying only faint differences, it might not be economically feasible to develop 22 different marketing and business strategies to engage all our target audiences.



Therefore, we proceed with conducting Latent Class Analysis to see if we can obtain better clustering results.

4.0 Performing Cluster Analysis – Latent Class Analysis

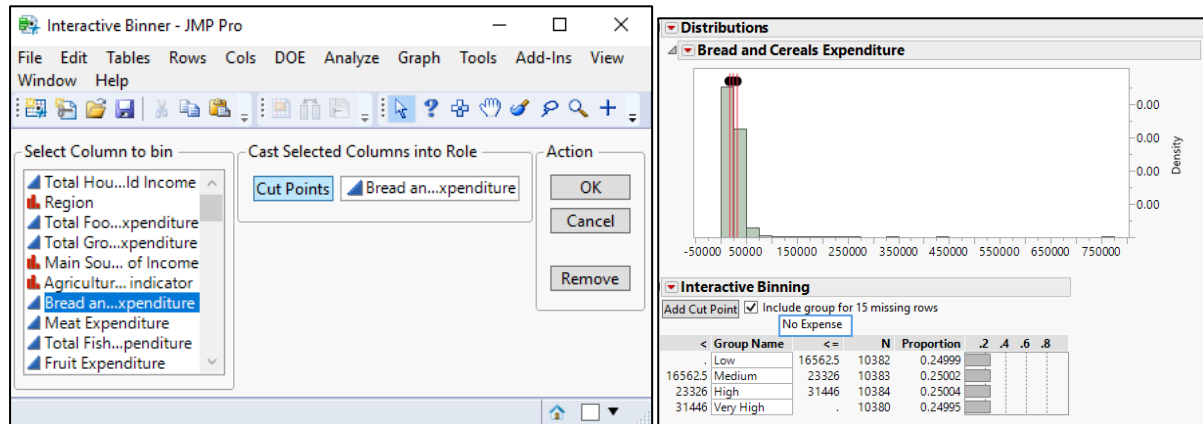
4.1 Interactive Binning

First, we will recode all 0s to “.” for all our chosen variables.

Recode - JMP Pro			
Bread and Cereals Expenditure			
Count	Old Values (26082)	New Values (26082)	
15	0	.	
1	25	25	
1	31	31	
1	32	32	
1	42	42	
1	72	72	

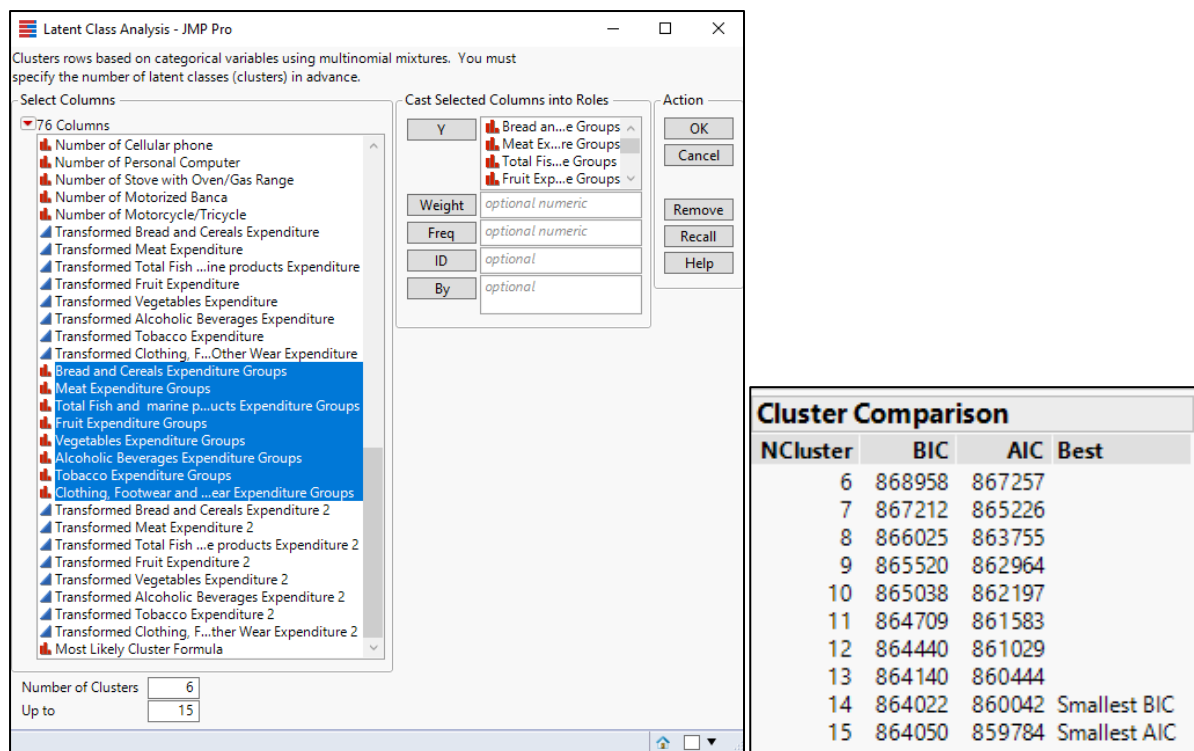
This allowed us to ensure that when we conduct binning, it will not affect the range amount and that we will be able to capture all missing values as “no expenditure” made by the households.

Once that has been completed, we performed interactive binning for all expenses as shown below:



4.2 Latent Class Analysis

First, we set the range from 6 to 15 and included all newly binned variables under “Y” and click “OK”.



The resulting table shows that 14 clusters have the smallest BIC and 15 clusters have the smallest AIC. Therefore, we will be choosing 14 clusters as the optimal number of clusters to perform our cluster analysis. The choice of 15 clusters will be rejected as it would be more economically feasible to choose a smaller number of clusters to focus our marketing efforts.

Since Latent Class Analysis shows a much smaller cluster size of 14 clusters, and K Means Clustering shows a bigger 22 clusters size, we recommend using the Latent Class Analysis to optimize our marketing efforts while ensuring the economical-feasibility of our strategies.

5.0 Validate and Interpret the Clusters

5.1 Value ordering of all categorical variables used for LCA

Prior to our interpretation of the LCA clustering results, we will perform value ordering to ensure that our values are arranged in a sensible ordinal manner.

Value Ordering

Specify data in the order in which you want the data to appear in the reports.

Low Expense
Medium Expense
High Expense
Very High Expense
No Expense
optional item

Remove

Move Up

Move Down

Reverse

Add

Value Ordering

Specify data in the order in which you want the data to appear in the reports.

Very High Expense
High Expense
Medium Expense
Low Expense
No Expense
optional item

Remove

Move Up

Move Down

Reverse

Add

Before Value Ordering

After Value Ordering

5.2 Cluster Summary

To further explore the cluster characteristics, we will plot all variables against the number of clusters to investigate distinct characteristics that belong to each cluster number.

The figure is a heatmap titled "Expenditure vs. Cluster Number". The y-axis lists expenditure categories: Clothing, Footwear and Other, Tobacco, Alcoholic Beverages, Vegetables, Fruit, Total Fish and products, Meat, and Bread and Cereals. Each category has a corresponding "Expenditure Groups" label. The x-axis lists 13 clusters: Cluster 1, Cluster 2, Cluster 3, Cluster 5, Cluster 4, Cluster 6, Cluster 7, Cluster 9, Cluster 8, Cluster 11, Cluster 10, and Cluster 13. The heatmap cells are colored based on the expenditure level: Very High Expense (dark blue), High Expense (medium blue), Medium Expense (light blue), Low Expense (very light blue), and No Expense (white). The legend on the right side of the heatmap defines these color codes.

23 | Page

Variable / Cluster Number	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14
Consumer Type	Mid to High	Low to Mid	Low to Mid	Mid to High	Mid to High	Low	Low to Mid	Mid to High	Mid to High	Low	Low	Low to Mid	Low to Mid	Low
Cluster Name	Potential High Flyers	Not Health-Conscious	Fond of Carbs	High Flyers	Health-Conscious	Vice-Free	Veg, Fruit and Bread Lovers	Fashionist as	Vice-Free	Vice-Loving	Vice-Free	Fashionist as	No Meat, No Vice	Outlier
Bread	Very High (52.7%)	Very High (14.9%)	Very High (26.4%)	Very High (67.3%)	Very High (13.2%)	Low (72.0%)	Very High (26.3%)	Very High (17.5%)	Very High (39.6%)	Low (72.8%)	Low (86.5%)	Low (62.5%)	Very High (24.1%)	NA
Meat	Very High (37.2%)	Low (7.9%)	Low (61.4%)	Very High (87.1%)	Very High (23.1%)	Low (44.1%)	Low (12.6%)	Very High (53.7%)	Very High (75.6%)	Low (72.1%)	Low (85.8%)	Low (13.2%)	Low (89.3%)	NA
Fish	Very High (56.7%)	Low (14.1%)	Low (35.9%)	Very High (83.0%)	Very High (10.4%)	Low (45.6%)	Low (20.0%)	Low (8.8%)	Very High (66.0%)	Low (76.9%)	Low (92.7%)	Low (84.8%)	Very High (61.6%)	NA
Fruit	Very High (15.4%)	Low (43.7%)	Low (47.2%)	Very High (82.2%)	Very High (24.6%)	Low (33.4%)	Very High (33.0%)	Very High (39.1%)	Very High (82.6%)	Low (75.4%)	Low (76.5%)	Low (29.5%)	Very High (32.4%)	NA
Vegetables	Very High (39.9%)	Low (28.1%)	Low (45.9%)	Very High (82.6%)	Very High (13.8%)	Low (25.8%)	Very High (31.2%)	Very High (16.2%)	Very High (70.6%)	Low (75.8%)	Low (94.4%)	Low (78.6%)	Very High (51.3%)	NA
Alcoholic Beverages	Very High (29.0%)	Very High (29.3%)	No Expense (46.9%)	Very High (46.0%)	No Expense (85.6%)	No Expense (76.4%)	No Expense (15.1%)	Very High (27.0%)	No Expense (81.7%)	Very High (13.2%)	No Expense (84.8%)	No Expense (51.9%)	No Expense (97.4%)	NA
Tobacco	Very High (29.3%)	Very High (25.8%)	No Expense (33.3%)	Very High (30.3%)	No Expense (99.9%)	No Expense (81.2%)	No Expense (21.0%)	Very High (21.9%)	No Expense (96.1%)	Very High (13.1%)	No Expense (78.9%)	No Expense (70.8%)	No Expense (25.7%)	NA
Clothing	Very High (10.4%)	Low (28.5%)	Low (34.4%)	Very High (80.8%)	Very High (28.4%)	Low (60.7%)	Low (20.0%)	Very High (73.9%)	Very High (70.4%)	Low (66.3%)	Low (67.1%)	Very High (42.0%)	Low (35.2%)	NA
No. of Households / Total No. of Households (%)	12.1%	12.1%	10.2%	9.0%	9.6%	7.2%	7.1%	6.8%	7.0%	5.6%	6.1%	3.5%	3.3%	0.5%
Cluster Income / Total Income (%)	12.0%	8.0%	5.1%	21.3%	10.2%	3.3%	4.9%	10.2%	15.4%	2.1%	2.0%	3.2%	1.8%	0.4%
Cluster Expenditure / Total Expenditure (%)	15.7%	10.4%	7.9%	17.7%	8.7%	3.8%	6.4%	8.2%	10.9%	2.8%	2.0%	2.3%	3.1%	0.1%

5.2.1 Description of Clusters:

Cluster 1 – Mid to High-Range Consumers: Potential High Flyers

The first cluster is a group of moderately-high consuming individuals. This cluster represents our largest group of households, at 5,027 households out of the total 41,544 households. Based on their consumption habits, we can see that they represent one of the highest consuming cluster for all products. Therefore, due to their large number, coupled with their moderately large consumption habit, this cluster would therefore represent a significantly large target audience of our marketing efforts.

Cluster 2 – Low to Mid-Range Consumers: Not Health-Conscious

This cluster refers to households that have low to moderate consumption of all products. However, they do show a high consumption of alcohol beverages, tobacco and meat. Since this group contains the second largest number of household, it would be wise to market these specific products to this cluster of households.

Cluster 3 – Low to Mid-Range Consumers: Fond of Carbs

This cluster contains households that are fond of consuming carbohydrates (Bread and Cereals). This cluster also contains the third largest number of households based on the entire population. Therefore, it is possible to investigate the potential of marketing carbohydrates to this cluster. Since food products have lesser shelf life, it would also be necessary to focus on the inventory and warehousing management of these carbohydrate food products to be sold to the households in this cluster.

Cluster 4 – Mid to High-Range Consumers: High Flyers

This cluster represents the 4th largest number of households. At the same time, they also represent one of the highest consumer of all goods. As a result, it is possible to focus higher marketing efforts on selling premium products to this cluster.

Cluster 5 – Mid to High-Range Consumers: The Health-Conscious

This cluster shows a high percentage of households that do not consume alcoholic beverages and tobacco. Therefore, the Company should consider keeping a smaller inventory amount of such products when catering to this cluster. However, since households in this cluster consume a moderate number of products, with a higher propensity to purchase clothing with 28.4% of the cluster indicating very high expenditure of clothing, it is possible to identify ways to increase expenditure on this product.

Cluster 6 – Low-End Consumers: Vice-Free

Like cluster 5, this cluster does not have a high propensity to consume alcoholic beverages and tobacco. However, the difference lies in the fact that consumers of this cluster do not consume as much as the consumers in cluster 5 in total. Cluster 5 consumes 8.7% of total consumption amount for all 14 clusters, while cluster 6 only consumes at a percentage of 3.8%. Therefore, marketing priority should be given to cluster 5 instead of cluster 6.

Cluster 7 – Low to Mid-Range Consumers: Veg, Fruit and Bread Lovers

Consumers in cluster 7 shows a higher preference in consuming vegetables, fruits and carbohydrates. Therefore, it is possible to focus more of such product sales to this cluster.

Similarly, due to the lower shelf life of such food product, inventory management of such product should be optimized to prevent wastage.

Cluster 8 – Mid to High-Range Consumers: Fashionistas

Like cluster 4, cluster 8 contains households that are relatively high consumers, albeit containing a lesser number of households (9.0% versus 6.8% for cluster 4 and cluster 8 respectively). However, a distinct difference is that cluster 8 households show a significant preference for the purchase of clothing. Therefore, the Company should consider undertaking greater marketing effort in pushing fashion products to this cluster.

Cluster 9 – Mid to High-Range Consumers: The Health-Conscious

Like cluster 4, cluster 9 shows a very similar consumption pattern where there is an overall high consumption of all products. However, a pronounced difference is that households in cluster 9 do not show much preferences for consumption of alcoholic beverages and tobacco, similar to cluster 5. A total of 81.7% of the cluster do not consume alcohol and 96.1% of the cluster do not consume tobacco. Therefore, it is possible to market premium products to households in this cluster, while at the same time, it is essential that the Company manages the inventory of alcoholic beverages and tobacco to households in this cluster such that it is still possible to cater to the minimal number of consumers while keeping storage cost low.

Cluster 10 – Low-End Consumers: Vice-Loving

This cluster shows a significant proportion of households that do not consume as many products as compared to the other clusters. However, the relatively large preference of consuming alcoholic beverages and tobacco in this cluster is noteworthy. As such, it is possible to continue to cater to such preferences held by this cluster, without having to maintain marketing efforts for the other products.

Cluster 11 – Low-End Consumers: Vice-Free

Compared to cluster 10, cluster 11 shows an even lower consumption pattern. Additionally, a large majority of cluster 11 do not consume both alcoholic beverages and tobacco. Additionally, cluster 11 only represents a small proportion of total number of households, at only 6.1% of the total survey sample size. Therefore, it is possible to focus less marketing effort on this cluster to reduce marketing cost.

Cluster 12 – Low to Mid-Range Consumers: Fashionistas

Cluster 12 shows a low to middle consumption pattern for all products except for alcoholic beverages, tobacco and clothing. Cluster 12 shows little to no preference for both alcoholic beverages and tobacco. However, it shows a large interest in the purchase of clothing.

Cluster 13 – Low to Mid-Range Consumers: No Meat, No Vice

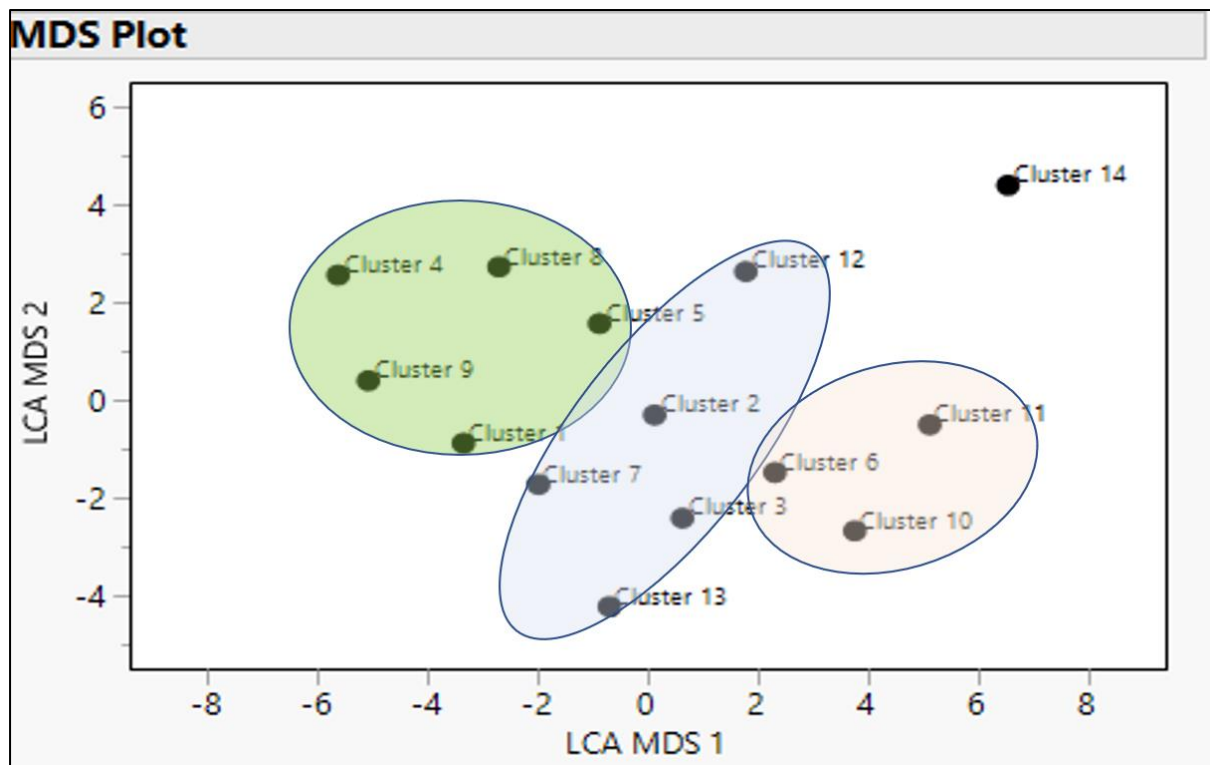
Cluster 13 displays a large preference on consumption of all food except for meat. Additionally, cluster 13 differs from the other clusters as it is the only cluster that shows a higher number of consumption of tobacco with almost no consumption of alcohol. 25.7% of the households do not consume tobacco, while 97.4% do not consume alcohol.

Cluster 14 – Low-End Consumers: Outlier

Cluster 14 shows an anomalous result as compared to the other clusters as they show almost no consumption of food items. Additionally, this cluster only represents 0.5% of the total survey sample size. Therefore, any marketing efforts targeted at this cluster would not be value-adding.

Based on the description above, we will now refer to the MDS plot to identify the distance between each cluster.

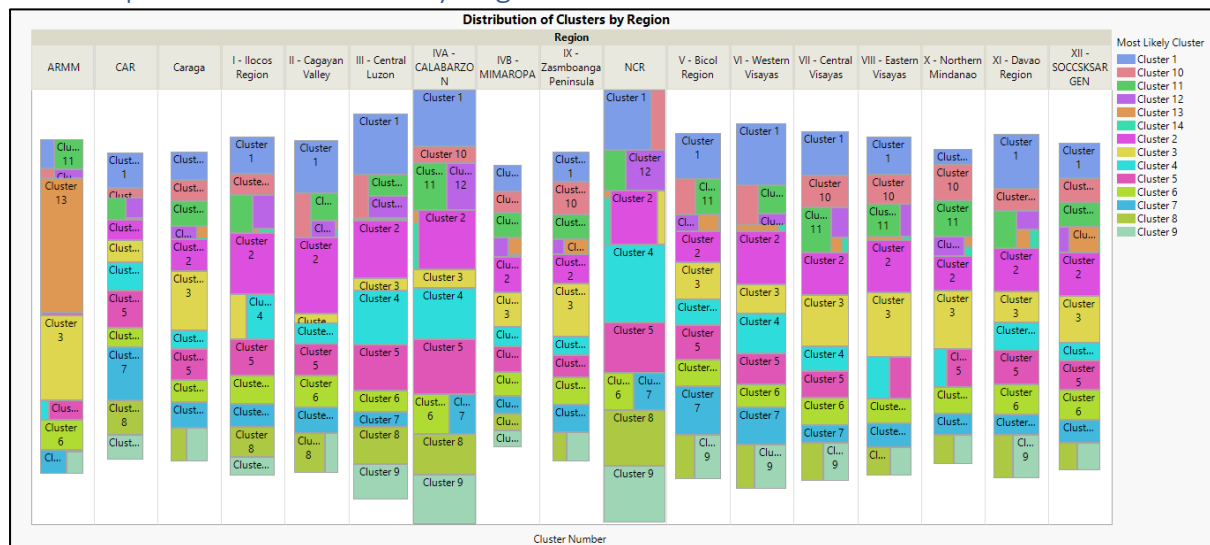
5.3 MDS Plot



Based on the MDS plot above, we can see that Cluster 1, 4, 5, 8 and 9 belongs to the Mid to High range of consumers. Next, cluster 2, 3, 7, 12 and 13 belongs to the Low to Mid-range of consumers. The last group of clusters, cluster 6, 10 and 11 belongs to the low-end consumer group. Cluster 14 is an outlier with too small of a population to consider implementing marketing strategies to.

Therefore, we will be implementing different marketing strategies based on each of the three distinct groups of clusters.

5.4 Interpretation of Clusters by Region

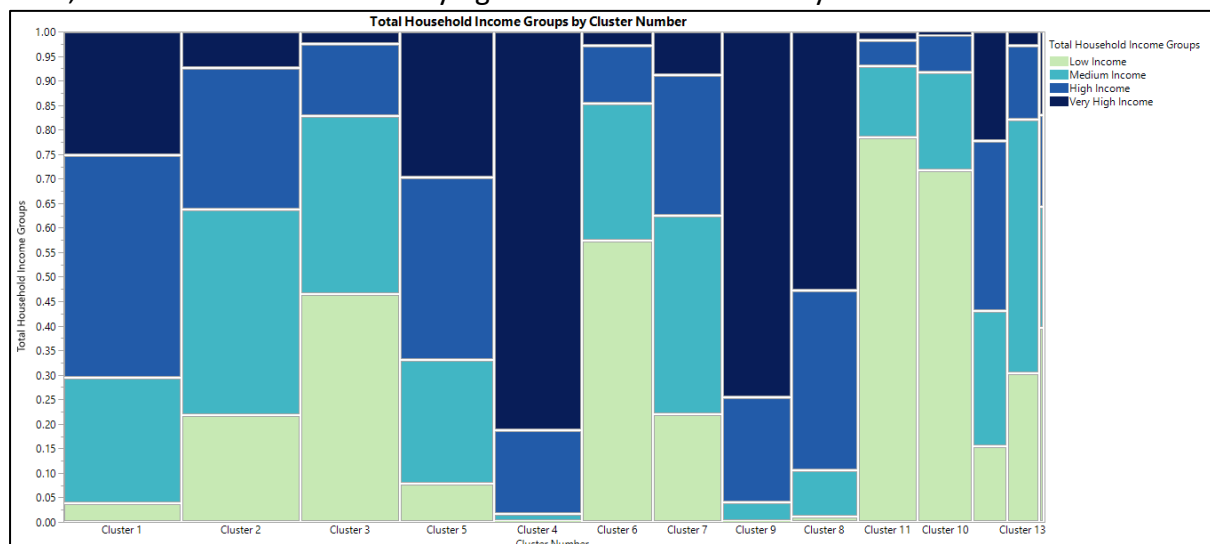


The Treemap above shows that the distributions of all clusters are generally equally distributed throughout all regions, except for cluster 13 in ARMM, which has an abnormally high percentage (at 40.17%) of cluster 13 “No Meat, No Vice”. It is possible that the reason for the lower consumption of meat and no consumption of alcohol and tobacco is that ARMM, also called the “Autonomous Region in Muslim Mindano”², has a Muslim majority population.

Since the distribution of the clusters are generally equally distributed throughout all regions, the extent of the marketing campaigns can too be equally distributed throughout all regions, apart from the ARMM.

5.5 Interpretation of Clusters by Total Household Income

Next, we are interested in identifying the household income by the Cluster Number.



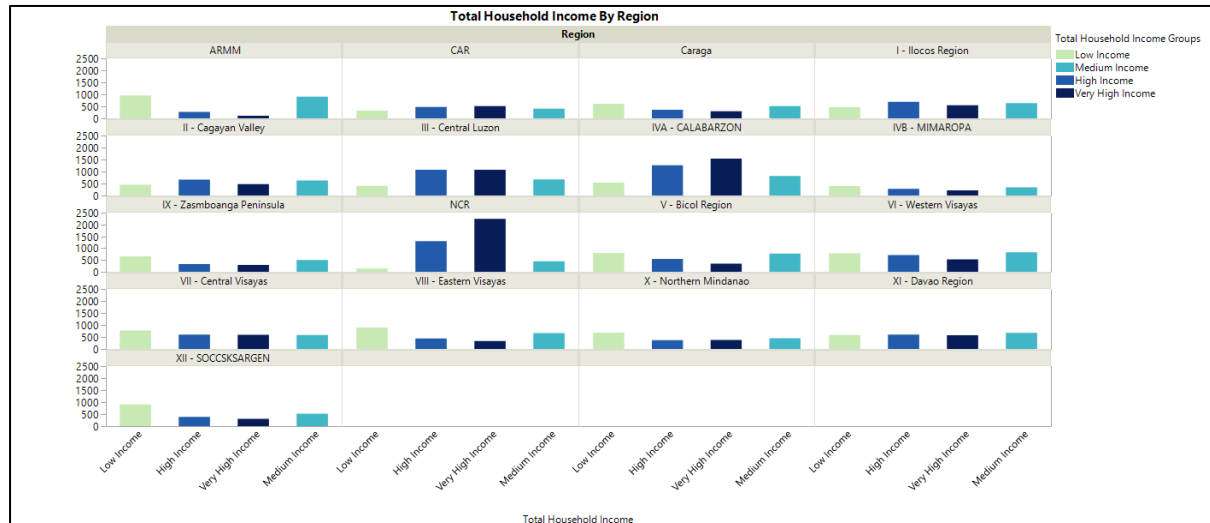
Based on the mosaic plot above, clusters 1, 4, 5, 8 and 9 shows households with a higher proportion of income. Additionally, clusters 6, 10 and 11 contain households with the lowest

² Official Website of the Autonomous Region in Muslim Mindano - <https://armm.gov.ph/>

income out of all the 14 clusters. These observations commensurate with our cluster analysis above, which shows that clusters 1, 4, 5, 8 and 9 are also the clusters with the highest expenditure, while clusters 6, 10 and 11 are the clusters with the lowest expenditure. The mosaic plot above also shows that cluster 3 have a relatively low income, right after clusters 6, 10 and 11.

5.6 Interpretation of Clusters by Total Household Income and Region

Next, we are interested in identifying the total amount of household income by region.



Based on the Bar Chart above, we can see that the highest number of households with very high income are from the regions NCR and IVA – Calabarzon. Next, we will compare the Bar Chart against the Treemap that describes the distribution of Clusters by Region. We can see that the regions NCR and IVA – Calabarzon contain the largest number of households, with clusters 1, 4, 5, 8 and 9 (Mid to High Range Consumers) representing the majority of each region. Therefore, this too, commensurate with our conjecture that high spenders tend to be high earners as well.