



# Segmentation of Local and International Brand of Chips Based on Macronutrients and Micronutrients Content Using Hierarchical Cluster Analysis



**Bernadine Ruiza G Ang\***

*University of the Philippines Diliman, Philippines*

**\*Corresponding author:** Bernadine Ruiza G Ang, University of the Philippines Diliman, Manila, Philippines

**Submission:** 📅 March 27, 2018; **Published:** 📅 April 11, 2018

## Abstract

Knowing the nutritional value of local and international chips is helpful for avoiding chronic diseases such as obesity, diabetes or cancer. Hierarchical Cluster Analysis will be used to categorize these chips according to their nutritional value. Forty-nine samples of local and international chips and eleven macronutrients and micronutrients will be examined. Multivariate Statistical Analysis of data using hierarchical cluster analysis (HCA) technique will be utilized through SPSS program. California Crunch was the most caloric with 210kcal. Miaow and Pods offered the highest in protein content, with the same value of 4g. In total fat, Loaded! (White-Choco) contains the highest with 13g of fat. Pee Wee (BBQ) has the highest sodium content with 390mg. Tattoos Corn Tube contains the highest fiber with 4g. For carbohydrates, the highest again is California Crunch, evidently portraying the relevance and relationship between carbohydrate and calories. Moby and Pillows both have the highest sugar content with 13g. Terra Sweet Potato contained the highest vitamin A with a value of 50%. The highest calcium content is 15% from Outback. Chip n' Chip and Pringles both have the highest vitamin C with 10%. Pods also contained the highest with 40%. HCA can be used as a tool to educate consumers on the nutritive value of chips and help consumers select the appropriate snack that they can include in their diets.

**Keywords:** Chips; Hierarchical cluster analysis; Nutrition; Metabolism; Carbohydrates; Fat

**Abbreviations:** HCA: Hierarchical Cluster Analysis; SPSS: Statistical Package for the Social Sciences

## Introduction

Research on local and international chips is an excellent model to discover the nutritional content of chips and how they vary depending on type of chips. Little is known that we can improve the nutritional content of chips and enjoy these snacks that are known to be unhealthy. Food fortification can indeed make this a reality and therefore research on the nutritional content of these chips is an effective method to begin with. Macronutrients are defined as protein, fat, carbohydrate and sugar while micronutrients are defined as vitamin A, calcium, vitamin C and iron. Excessive calories, protein, fat, sodium, carbohydrate and sugar are a few of the main sources of chronic diseases such as diabetes, obesity and cancer. Aggravated effects of these diseases can lead to fatality. It is essential to know how much macronutrients and micronutrients we intake.

In this research, the researcher will concentrate on the clustering of calories, protein, fat, sodium, fiber, carbohydrate, sugar, vitamin A, calcium, vitamin C and iron within these chips.

In studying the metabolic pathways of some macronutrients such as fat and carbohydrate, triglyceride coming from fat will split its components-glycerol and three fatty acids. The glycerol portion will undergo gluconeogenesis which is the generation of glucose while the remaining three fatty acids will go through the Krebs cycle which is irreversible. In realizing this, we comprehend how excess fat through glycerol can be converted to glucose which is a form of carbohydrate. For carbohydrate, their composition will continue on to gluconeogenesis and glycolysis which will form or breakdown carbohydrate respectively into glucose entering Krebs cycle as well and consequently then generating ATP, a source of energy that we use as our fuel. From this, we comprehend the interrelationships of nutrients such as fat and protein as a general example.

Cluster Analysis is a statistical technique that groups similar objects together. Hierarchical cluster analysis (HCA) will be used to classify chips according to their macronutrient and micronutrient composition. Hierarchical cluster analysis is a type of cluster

analysis and follows three basic steps: calculate the distances, link the clusters and choose a solution by selecting the right number of clusters [1]. Again, the primary purpose of this study is to investigate the classification of calories, protein, fat, sodium, fiber, carbohydrate, sugar, vitamin A, calcium, vitamin C and iron in local and international brand of chips.

### Data description

The study used data collected in local and international chips from UniMart supermarket and SM Center point grocery. Manual data gathering using simple random sampling was observed. 49 chips were evaluated according to the nutrition facts label at the back of each chip. Nutritional Facts for each chip was collected and tabulated in Microsoft Excel. The following were obtained: calories, protein, fat, sodium, fiber, carbohydrate, sugar, vitamin A, calcium, vitamin C and iron. Each value for each chip was recorded in Table 1.

**Table 1:** List of chips.

Chip Brand Name	
California Crunch	Oishi Potato Fries (Tomato Ketchup flavour)
Cassava Crisps	Oishi Ridges (Onion and Garlic Flavor)
Chabi Chips	Oishi Shoops
Cheetos (Jalapeno)	Okeji
Cheez-It	Outback
Chicharron ni Mang Juan	Pee Wee (BBQ)
Chip n' Chip	Pillows
Chippy	Pizza Potato
Chiz Crispers	Pods
Doritos	Potato Chips
Ebisen Prawn Crackers	Pringles
Fishda	Ri-Chee
Garden of Eatin	Ripples
Gourmet Picks	Roller Coaster
Humpy Dumpty	Snacku
Kirei Yummy Flakes	Sun Chips
Labzter	Taquitos
Loaded! (White-Choco)	Tattoos Corn tube
Miaow	Tempura
Moby	Terra Sweet Potato
Mr. Chips	Tomi
Munchies	Tostillas
Nova	Tostitos
Oishi Marty's Baconette Strips	VCut
Oishi Oheya!	Yasai

### Research Methodology

Statistical Package for the Social Sciences (SPSS) was utilized to analyze and further obtain the frequency of each chip. This study

was conducted starting December 2016 in Manila, Philippines. Pub Med was utilized to search similar studies on the said topic. Application of Multivariate Technique especially cluster analysis is appropriate in the given data set. It sought to use to form clusters from the brand chips based on the nutrients content for each chips. A Hierarchical Cluster Analysis will help for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case as a separate cluster, i.e. there are as many clusters as cases, and then combines the clusters sequentially, and reducing the number of clusters at each step until only one cluster is left [2]. The clustering method uses the dissimilarities or distances between objects when forming the clusters. The SPSS programmer's calculates 'distances' between data points in terms of the specified variables. A hierarchical tree diagram, called a dendrogram on SPSS, can be produced to show the linkage points. The clusters are linked at increasing levels of dissimilarity.

### Statistical Analysis

Eleven macronutrients and micronutrients compositions of forty-nine local and international brand of chips were investigated. The observed macronutrients and micronutrients were: calories, protein, fat, sodium, fiber, carbohydrate, sugar, vitamin A, calcium, vitamin C and iron. Prior to HCA, the data was transformed and scaled. Hierarchical Cluster Analysis was then used to define the elements of macronutrients and micronutrients in local and international chips. It was performed to assess the similarities of the chips. Ward's method was used as a cluster method. The interval measure is Squared Euclidean distance. The transform values were standardized according to Z scores by variable. A dendrogram was utilized to visualize the similarity between samples are nested tree-like structures, and usually reflect a development sequence. Each chip is treated as a separate and distinct cluster to begin with. They are merged using an appropriate similarity measure until every object belongs to a large cluster. It may help for "seeing the market structure" in terms of brands.

### Results and Discussion

Table 2 presented below shows the summary of reports. In cluster 1, the mean for calories is 154.170kcal which is the highest compared to cluster 2 which yields 79kcal as the mean while the third cluster has 152kcal. For protein, cluster 3 contains 6.4g which is the highest and consequently the mean for cluster 1 is 1.91g while the mean for cluster 2 is only .85g. For fat, cluster 3 yields the highest with 34.6g while cluster 1 got 8.02g while cluster 2 has 4.35g. For sodium, cluster 3 has the highest value with 297mg while cluster 1 has 191.96mg and cluster 2 only has 93.5mg. For fiber, the mean highest is 3.4g which belongs to cluster 3. The next highest mean is .92g from cluster 1 and .50g from cluster 2. For carbohydrates, the highest mean is 18.57g from cluster 1. Cluster 3 was the next highest with 18.08g and cluster 2 with a mean of 8.5g. For sugar, the highest value is 2.42g from cluster 1. The next highest is 1.10g from cluster 2 while the last is 0 from cluster 3. For vitamin A, the highest mean is 10.55% from cluster 1. The other values from cluster 2 and 3 were both 0%. For calcium, the highest

mean is 1.4894% from cluster 1. Cluster 2 has 1% while cluster 3 has 0%. For vitamin C, the highest mean is from cluster 3 which yield 25%. The second highest is from cluster 2 with 5% mean. The lowest would be .2979% from cluster 1. Lastly, for iron, the highest is 5.4% from cluster 1. The next highest is 4% from cluster 3 then 1% from cluster 2. In general, cluster 1 is high in calories, carbohydrates, sugar, vitamin A, calcium and iron. Cluster 3 is high in protein, fat, sodium, fiber, and vitamin C. It is advisable to follow or intake brands under cluster 2 since they are not high in fat, carbohydrate, sodium or sugar. It is also evidently projected that there is almost an insignificant portion of micronutrients such as

vitamin A, calcium, vitamin C and iron with only 25% as the highest from vitamin C, 10.55% from vitamin A, 5.4% from iron and only 1.489% for calcium. Using HCA was efficient in determining the macronutrients and micronutrients for each cluster. The result of the cluster analysis is summarized in the agglomeration schedule. For instance, at stage 1 (row 1), we see that case 2 and case 32 are combined. The squared Euclidean distance between these two is displayed in the coefficient column. The cluster created by their joining next appears in stage 27. In stage 27, the observations 2 and 3 were joined. The resulting cluster next appears in stage 34.

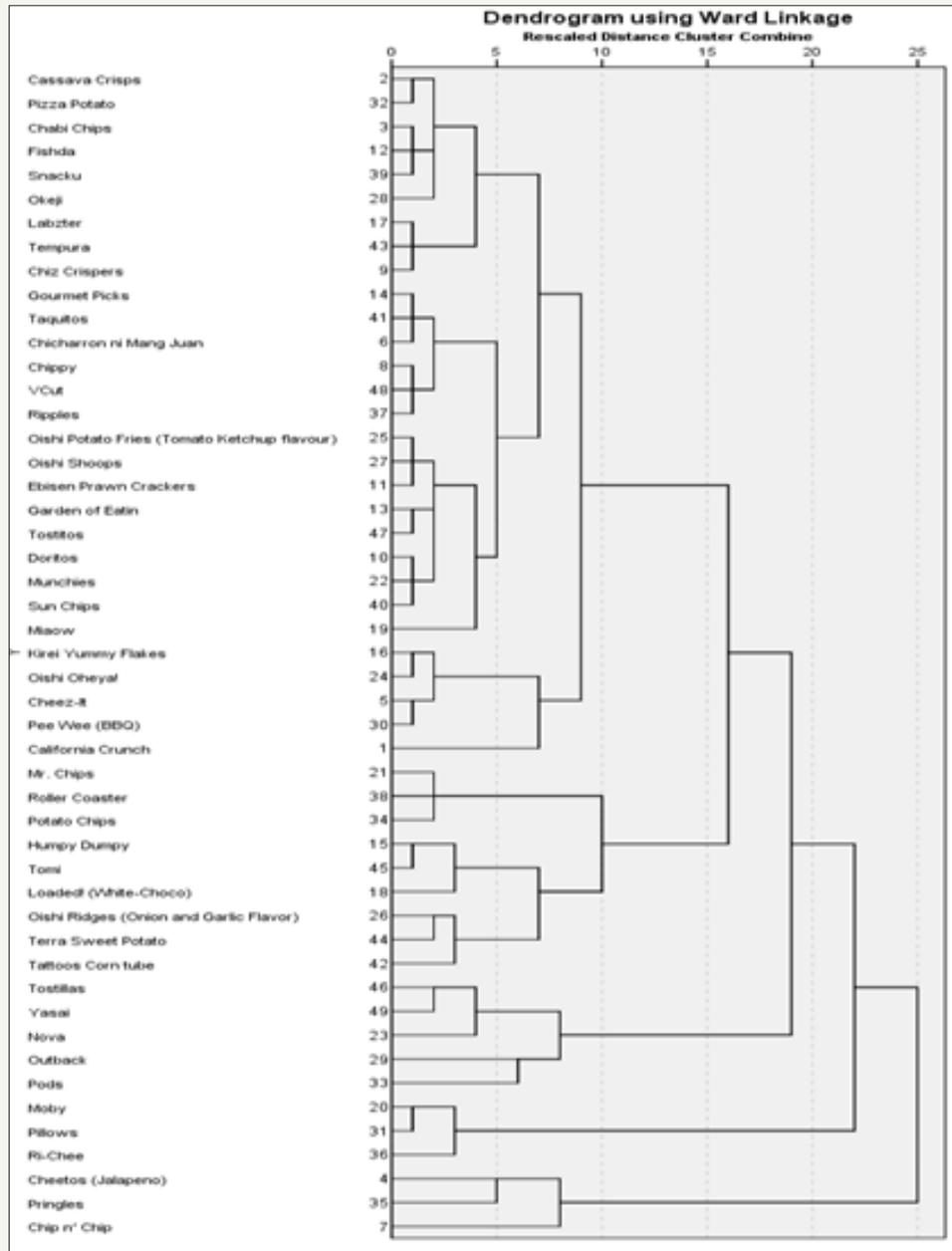
**Table 2:** Summary of statistics of clusters.

Clusters		Calories	Protein (g)	Fat (g)	Sodium (mg)	Fiber (g)	Carbohydrate (g)	Sugar (g)	Vitamin A	Calcium	Vitamin C	Iron
1	Mean	154.17	1.91	8.02	191.96	0.92	18.57	2.42	10.55%	1.49%	0.30%	5.43%
	N	47	47	47	47	47	47	47	47	47	47	47
	Std. Deviation	16.7154	0.612	2.519	87.741	0.975	2.482	2.779	16.32%	2.87%	1.50%	11.05%
2	Mean	79	0.85	4.35	93.5	0.5	8.5	1.1	0.00%	1.00%	5.00%	1.00%
	N	2	2	2	2	2	2	2	2	2	2	2
	Std. Deviation	15.5563	0.212	0.919	23.335	0	0.707	1.273	0.00%	1.41%	7.07%	1.41%
3	Mean	152	6.4	34.6	297	3.4	14	0	0.00%	0.00%	25.00%	4.00%
	N	1	1	1	1	1	1	1	1	1	1	1
	Std. Deviation	.	.	.	.	.	.	.	.	.	.	.
Total	Mean	151.12	1.96	8.4	190.12	0.95	18.08	2.32	9.92%	1.44%	0.98%	5.22%
	N	50	50	50	50	50	50	50	50	50	50	50
	Std. Deviation	22.1015	0.899	4.56	88.632	1.012	3.179	2.732	16.01%	2.79%	4.00%	10.75%

**Table 3:** Agglomeration schedule.

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	32	0.156	0	0	27
2	25	27	0.438	0	0	17
3	20	31	0.773	0	0	29
4	10	22	1.147	0	0	12
5	8	48	1.633	0	0	16
6	17	43	2.146	0	0	18
7	3	12	2.753	0	0	15
8	15	45	3.372	0	0	30
9	14	41	4.118	0	0	14
10	16	24	4.979	0	0	26
11	13	47	5.85	0	0	20
12	10	40	7.08	4	0	20
13	5	30	8.318	0	0	26
14	6	14	10.092	0	9	21
15	3	39	11.903	7	0	23
16	8	37	13.802	5	0	21
17	11	25	15.714	0	2	28

18	9	17	18.507	0	6	34
19	21	38	21.52	0	0	25
20	10	13	24.733	12	11	28
21	6	8	27.959	14	16	35
22	46	49	31.464	0	0	33



Figures 1 : Dendrogram.

Based on matrix of proximity, the diagram for dendrogram can be used. There are several ways of grouping these points in multidimensional space to form hierarchical clusters. Two samples close to each other have similar values. Hence, the greater proximity between measures has the greater the similarity. The Agglomeration Schedule (Table 2) follows the proximity matrix in the output. The agglomeration schedule displays how the

hierarchical cluster analysis progressively clusters the cases or observations. Each row in the schedule shows a stage at which two cases are combined to form a cluster, using an algorithm dictated by the distance and linkage selections. The agglomeration schedule lists all of the stages in which the clusters are combined until there is only one cluster remaining after the last stage. The number of stages in the agglomeration schedule is one less than the number

of cases in the data being clustered. In this example, there are 48 stages because the sample consists of 47. The coefficients at each stage represent the distance of the two clusters being combined. Inconclusive result observed in the third cluster having single chip (Table 3).

As mentioned previously, a hierarchical cluster analysis is best illustrated using a dendrogram, a visual display of the clustering process (Figure 1). It appears at the very end of the SPSS output. Examining the dendrogram from left to right, clusters that are more similar to each other are grouped together earlier. The vertical lines in the dendrogram represent the grouping of clusters or the stages of the agglomeration schedule [3]. They also indicate the distance between two joining clusters (as represented by the x-axis, located above the plot). As the clusters being merged become more heterogeneous, the vertical lines will be located farther to the right side of the plot, as they represent larger distance values. While, the vertical lines are indicative of the distance between clusters, the horizontal lines. Cluster analysis results are shown by a dendrogram (Table 4), which lists all samples and indicates the level of similarity of any two clusters joined (Figure 1). In the dendrogram below, sample 2-32 is examples of most similar among a few such as sample 3-12-39 and 17-43-9. The last two clusters to form are sample 2-32-3-12-39-28 and 4-35-7.

**Table 4:** Cluster analysis results are shown by a dendrogram.

23	3	28	35.102	15	0	27
24	26	44	39.03	0	0	31
25	21	34	43.159	19	0	44
26	5	16	47.458	13	10	40
27	2	3	52.306	1	23	34
28	10	11	57.75	20	17	32
29	20	36	63.578	3	0	47
30	15	18	70.087	8	0	39
31	26	42	77.93	24	0	39
32	10	19	86.789	28	0	35

33	23	46	96.031	0	22	41
34	2	9	107.437	27	18	38
35	6	10	120.619	21	32	38
36	4	35	134.178	0	0	42
37	29	33	150.445	0	0	41
38	2	6	168.193	34	35	43
39	15	26	187.283	30	31	44
40	1	5	206.48	0	26	43
41	23	29	226.844	33	37	46
42	4	7	248.238	36	0	48
43	1	2	272.852	40	38	45
44	15	21	299.557	39	25	45
45	1	15	342.258	43	44	46
46	1	23	396.151	45	41	47
47	1	20	457.087	46	29	48
48	1	4	528	47	42	0

## Conclusion

Using Cluster Analysis, the data observed of 3 cluster of brand of chips. This paper has shown that Cluster analysis determines how many 'natural' groups there are in the sample. It also allows you to determine who in your sample belongs to which group. Cluster analysis is not as much a typical statistical test as it is a 'collection' of different algorithms that put objects into clusters according to well-defined similarity rules. Hierarchical cluster analysis is suggested as a practical method in identifying meaningful clusters within samples that may superficially appear homogeneous.

## References

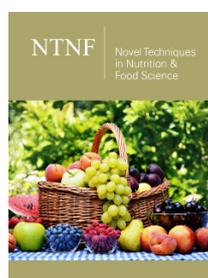
1. Statistics Solutions (2015) Conduct and interpret a cluster analysis. Chicago, USA.
2. IBM. Hierarchical Cluster Analysis.
3. Odilia Yim, Kylee TR (2015) Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. The Quantitative Methods For Psychology, pp. 8-21.



Creative Commons Attribution 4.0 International License

For possible submissions Click Here

[Submit Article](#)



## Novel Techniques in Nutrition and Food Science

### Benefits of Publishing with us

- High-level peer review and editorial services
- Freely accessible online immediately upon publication
- Authors retain the copyright to their work
- Licensing it under a Creative Commons license
- Visibility through different online platforms