USDA National Nutrient Analysis

By:

Chinmay Patil, Karan Raj Jonnalagadda, Kriti Srivastava, Pramathesh Shukla, and Unnati Thakkar

---

Executive Summary

We all know that healthy eating is an essential part of a healthy lifestyle. The modern diet is full of processed food which generally lacks the essential nutrients. The habit of eating unhealthy food can lead to inadequate intake of necessary nutrients. Such a diet can lead to causes like diabetes, hypertension, anemia, and obesity. Therefore, it is essential to understand healthy eating , as it is absolutely essential to understand the nutrients that we consume from our daily eating products. Sometimes even if we eat healthy food, our body is deficient in some important nutrients. Therefore, this study is to analyze different food products with their nutrients in it and trying to answer following questions:

- How to gain and retain maximum energy?
- Is it possible to explain the food products based on classification of their nutrients description?
- If yes, is there any relationship between the nutrients ?

Database Content

This study is made on the USDA (U.S. Department of Agriculture) National Nutrient database back. The USDA National Nutrient Database for Standard Reference (SR) is the major source of food composition data in the United States. It provides the foundation for most food composition databases in the public and private sectors. This version, Release 27 (SR27), contains data on 8,618 food items and up to 45 food components. The sample dataset is a flattened version of the USDA National Nutrient Database, from the now outdated version SR27.

Methods:

- The Principal component analysis of the nutrients of the food sample was implemented to explain the food products based on classification of their nutrients description.
- To get the deep knowledge of the USDA National Nutrient Database data confirmatory factor analysis (CFA) was conducted using the psych package using the factanal function applied using R in RStudio.
- Canonical correlation analysis was implemented on the components of PCA analysis to find the correlation between the PCA components to study the correlation between different food groups explained by its nutrients.
- Multiple Regression is implied on the dataset to conclude the results and interpretation between the dependent variable and the response variables.

Results:

- Regression analysis gave us five major nutrients which contribute towards energy and the nutrients are Proteins, Carbohydrates , Fiber , Fats and Magnesium.

- In PCA analysis, the food nutrients can be reduced to three major components as fiber-riched carbohydrates, anti- anemic nutrients that is very rich in vitamins, iron and zinc and finally high protein nutrients,

- By applying factor analysis it can be interpreted that most of the vitamins can be divided into 5 groups: Dietary, Minerals, Aminos, strength, and disaccharides.
- In food items as Fiber- rich Carbohydrates increases, Anti-anemic food nutrients also increase. Similarly, as Fiber- rich Carbohydrates increase in the food items, high protein nutrients are also observed to be increased.

Future Works:

- This study is done on the outdated version SD27 of USDA National Nutrient Database which has only 35 nutrient variables however, the latest version has upto150 nutrient components.
- This study only shows how the food products are rich or deficits in different good components however, it does not cluster the food products or labels the good product into categories on the basis of their nutrient factors.
- This study does not include the relationship between the components produced from the factor analysis which can be done in the future analysis.

Limitations:

- The study does not identify the amount / portion of the nutrients are needed in our diet.
- Absence of data to study the estimation of the risk of disease with nutrient patterns.

Conclusions:

It was found that to stay healthy and energetic, it is important to have a complete diet that includes all five nutrients in our diet: protein , fat, carbohydrates , fiber and magnesium. Depending upon the need of the individual, it is possible to focus on food products that are rich in specific sets of nutrients: Fiber - rich Carbohydrates that are a rich source of fiber, as fiber itself is a form of carbohydrate. Anti-anemic nutrient : includes high content of iron with vitamin C and zinc anti-anemic diet is good for people struggling with anemia. High-Protein includes proteins and good fats that are good for building muscles and are very important especially for athletes.It is also possible to focus on food product that are rich in specific set of nutrient in a different level well as: Dietary: food items rich in dietary nutrients are rich in vitamin b6, iron, thiamin etc. and they are involved in many processes of our body and are necessary for normal cell growth and function. Minerals: mineral rich diet includes vitamin b 12, copper, vitamin A and they can be found in nuts and seeds , oysters and liver. Aminos: They are very important for building muscle, regulating gene expression, cell signaling, and immunity and found in food products like soy sauce. Strength: The strength includes energy and fats. Disaccharides :Include sugar and vitamin C.

It was also observed that the nutrients patterns are related to each other as : In food items as Fiber- rich Carbohydrates increases, Anti-anemic food nutrients also increase. Similarly, as Fiber- rich Carbohydrates increase in the food items, high protein nutrients are also observed to be increased.

Technical Summary

Abstract

Nutrition is the preeminent part of everyone's life starting from the new born baby to oldsters. Every person in the age group of 20-40 is aiming to gain and retain a healthy lifestyle. Gym can help shape the physique of people but nutrition also plays a vital role to achieve the goal. In this paper we are introducing the various approaches performed on USDA dataset containing the food and nutrition information. The first technique we used is multiple regression, which tells us about how to gain and retain energy by minimum food consumption and 98% of variance , makes the approach very tight. Performed PCA and Factor analysis to find possible to explain the food products based on classification of their nutrients description.PCA, that is a dimensionality reduction approach shows the food products can be explained in a high level by three nutrient sets as protein rich food , food beneficial for anemic people and fiber rich carbs. Canonical correlation provides evidence of a strong relationship between the nutrient patterns produced from PCA.

Introduction :

We all know that healthy eating is an essential part of a healthy lifestyle. The modern diet is full of processed food which generally lacks the essential nutrients. The habit of eating such food can lead to adequate intake as well as lead to health problems like diabetes, hypertension, anemia, and obesity. Therefore, it is essential to understand healthy eating, it is absolutely essential to understand the nutrients we consume from our daily eating products. Sometimes even if we eat healthy food, our body is deficient in some important nutrients. Therefore, this study is to analyze different food products with their nutrients in it.

The USDA National Nutrient Database for Standard Reference (SR) is the major source of food composition data in the United States and provides the foundation for most food composition databases in the public and private sectors. To develop and update this database, the National Food and Nutrient Analysis Program (NFNAP) was initiated in 1997 which was an InteragencyAgreement between the National Institutes of Health and the US Department ofAgriculture (USDA since then it had become the most important means of accomplishing a comprehensive update to the National Nutrient Databank. This program's objectives were:*(1) evaluation of existing data; (2) identification of Key Foods and nutrients for analysis; (3) development of nationally based sampling plans; (4) analysis of samples; and (5) compilation and calculation of representative food composition data*. The sampling plan that was developed was based on a self-weighting stratified design where first, the U.S. was divided into four regions, then each region was further divided into three implicit strata from which generalized Consolidated Metropolitan Statistical Areas (gCMSAs) were selected. Then the Rural and urban locations were selected within gCMSAs. Commercial supermarket lists were used to select 24 outlets for food pickups; specific brands were selected based on current market share data (pounds cons ethnic and regional foods. Sampling plans have been developed for margarine, folate-fortified foods (e.g. flours, bread, and pasta), and a number of highly consumed mixed dishes **.** (Pehrsson, P. R., Haytowitz, D. B., Holden, J. M., Perry, C. R., & Beckler, D. G. , 2000). With the help of NFNAP and the new database system that was developed at NDL, USDA continues updating its food composition databases to support nutrition-related research in the scientific community and provides accurate and

representative mean estimates of nutrient profiles in generically described foods as well as brand-specific products.

Another study (Stricker, Onland-Moret, Boer, Schouw, Verschuren, May, Beulen , 2013) was aimed to explore differences between dietary patterns derived from principal component analysis (PCA) and k-means cluster analysis (KCA) in relation to their food group composition and ability to predict CHD(Chronic Heart Disease) and stroke risk. Both PCA and KCA extracted a prudent pattern (high intakes of fish, high-fiber products, raw vegetables, wine) and a western pattern (high consumption of French fries, fast food, low-fiber products, other alcoholic drinks, soft drinks with sugar) with small variation between components and clusters. PCA and KCA found similar underlying patterns with comparable associations with Chronic Heart Disease and stroke risk. A prudent pattern reduced the risk of Chronic Heart Disease and stroke.

There was similar study (McCann, S., Marshall, J., Brasure, J., Graham, S., & Freudenheim, J., 2001) *"to assess the effect of different methods of classifying food use on principal components analysis (PCA)-derived dietary patterns, and the subsequent impact on estimation of cancer risk associated with the different patterns."*

The source of data for our study is a flattened version of the USDA National Nutrient Database. This study involves the various approaches to questions related to food and their nutrition value. With the help of regression techniques like Multiple regression and Dimensionality reduction techniques like principle component analysis and factor analysis also canonical correlation analysis gives profound results.

Methods:

We have used **Linear Regression** to check whether we can gain and retain energy by consuming minimum nutrients from our daily diet.

Stepwise regression was the first approach to find the way towards the answer. Which gave the basic idea of contributing nutrients towards energy, it failed to give satisfied results . Then we applied a lasso regression analysis to gain more knowledge, but again it failed.So we decided to move manually to get the results.

The **principal Component Analysis (PCA)** was used to see if the food products can be explained with less nutrients.

For PCA analysis **,** the data was first cleaned up by omitting zeros as entries with zeros are considered as missing values. On performing the normality test, it was found that most of the feature distributions were skewed therefore, log transformation was applied on all the features which made the distributions of features either symmetric or moderately skewed.

The factorability was tested by performing Kaiser-Meyer-Olkin factor adequacy Test, Bartlett's Test of Sphericity, and Reliability Analysis using Cronbach's Alpha.
After performing the sensitivity test, PCA analysis was made for creating three components.

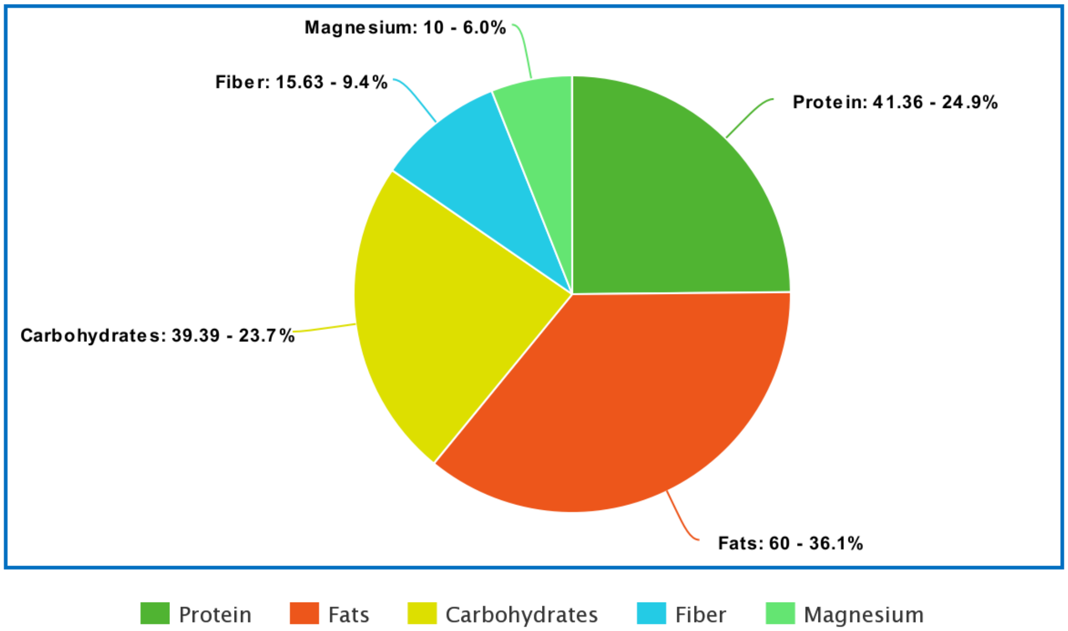
Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. The theory behind factor analytic methods is that the information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.The purpose of factor analysis is to reduce many individual items into a fewer number of dimensions. Factor analysis can be used to simplify data, such as reducing the number of variables in regression models. Most often, factors are rotated after extraction. Factor analysis is also used to verify scale construction. It may help to deal with data sets where there are large numbers of observed variables that are thought to reflect a smaller number of underlying/latent variables. One of the key strengths of CFA is that researchers could identify the fit of the measurement model prior to estimating the model. It may help to deal with data sets where there are large numbers of observed variables that are thought to reflect a smaller number of underlying/latent variables.

A **Canonical Correlation Analysis (CCA)** is conducted to evaluate the multivariate shared relationship between the two variable sets that were obtained from PCA. One of them is done between RC3 and RC2, which is FiberRichedCarbsFood and Anti-AnemicFood respectively. Second one is done between RC3 and RC1, which is FiberRichedCarbsFood and HighProtienFood respectively.

Results and Discussions:

After performing stepwise regression we got 15 nutrients which contributes in gaining energy and which is unsensible to have 15 nutrients for our daily diet as we are focused towards the healthy lifestyle not obesity.Lasso and Ridge regressions results outclassed the stepwise regression results as it gave us twice the number of nutrients.So we decided to move manually and we got 5 nutrients which highly contributes towards gaining energy.These nutrients are Proteins, Fats , Carbohydrates , Fiber and Magnesium. These nutrients together contribute 99.15%.Which means you get 99.15% energy by consuming these nutrients in your daily diet.

Energy


Magnesium: 10 - 6.0%
Fiber: 15.63 - 9.4%
Protein: 41.36 - 24.9%
Carbohydrates: 39.39 - 23.7%
Fats: 60 - 36.1%
Protein
Fats
Carbohydrates
Fiber
Magnesium

Proteins	Fats	Carbohydrates	Fiber	Magesium
Lean meats and Seafood	**Nuts** (almonds, peanuts, macadamia, hazelnuts, pecans, cashews)	Vegetables, Whole fruits	Beans	Dark chocolate
Poultry & Eggs	Avocados	Legumes, Seeds	Avocados	Avocado
Dairy products	Whole Eggs , Fish	**Nuts**	Whole Grains	Lean meats and Seafood
Nuts (including nut pastes) and seeds	Yogurt & Cheese	Tubers	Potatoes	**Nuts** & Legumes
Legumes and beans	Dark Chocolates and Chia seeds	Whole grains	**Nuts**	Seeds and Wholegrains

Three tests of factorability on the entire data set was performed before performing PCA:

• Kaiser-Meyer-Olkin factor adequacy Test: Overall MSA = 0.89 (Since it was greater than .7 that suggests that sample of dataset is good for performing the PCA analysis).
• Bartlett's Test of Sphericity: p-value < 2.22e-16 which is very small that shows we had enough variance in the data so we can perform factor analysis)
• Reliability Analysis using Cronbach's Alpha: raw_alpha = 0.92 that also suggests that the sample data is good to perform PCA.

PCA summary information (Fig:2) shows, approx. 80% of the cumulative variance is explained by 5 components. Approx 47% of the cumulative variance is explained by the first component itself however 3 components are determined by the Scree plot (Fig: 3) which is greater than 1 Eigenvalue.However, On performing the Knee test, there are 2 components. Since 68% of the cumulative variance is determined by the 3 components Therefore, the decision was taken to choose 3 components for the PCA model.

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Standard deviation	2.9060	1.5500	1.22185	0.99912	0.96728	0.82340	0.69657	0.64301	0.61984	0.58567	0.54954	0.48290
Proportion of Variance	0.4692	0.1335	0.08294	0.05546	0.05198	0.03767	0.02696	0.02297	0.02134	0.01906	0.01678	0.01295
Cumulative Proportion	0.4692	0.6026	0.68558	0.74103	0.79301	0.83068	0.85764	0.88061	0.90195	0.92101	0.93778	0.95074

	PC13	PC14	PC15	PC16	PC17	PC18
Standard deviation	0.46305	0.40805	0.39144	0.36914	0.34138	0.31580
Proportion of Variance	0.01191	0.00925	0.00851	0.00757	0.00647	0.00554
Cumulative Proportion	0.96265	0.97190	0.98041	0.98798	0.99446	1.00000

Fig:2 Importance of components in PCA summary information

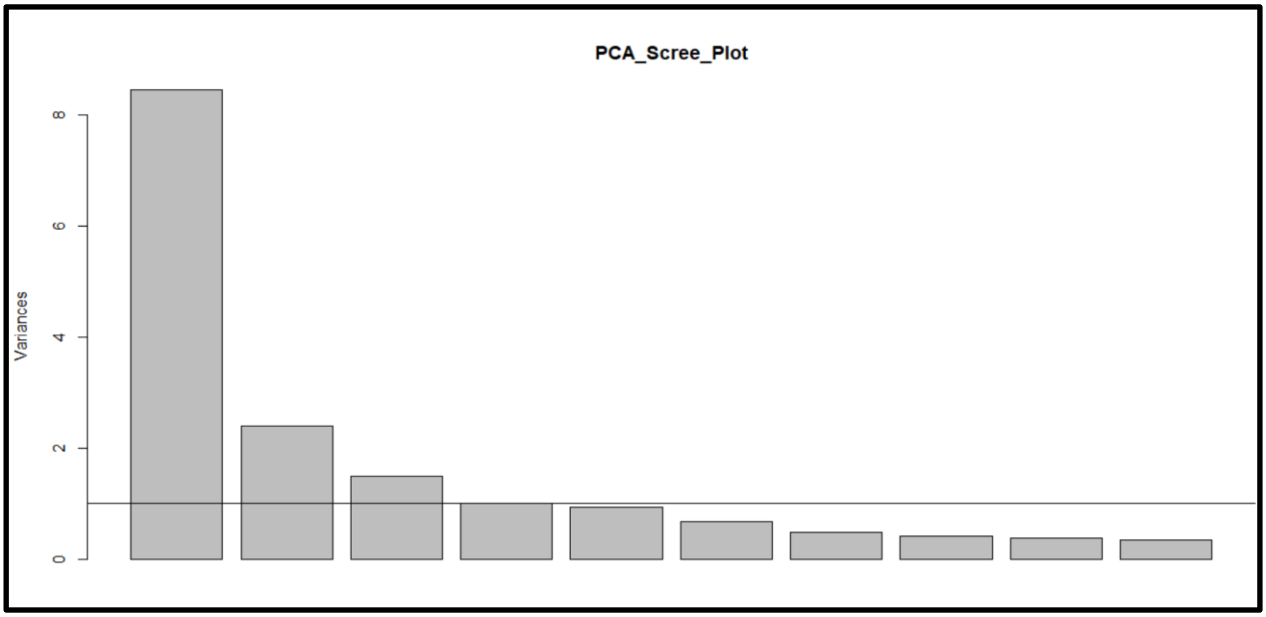


Fig:3 Scree plot of PCA of USDA National Nutrient Database

Loadings:

	RC3	RC2	RC1
Carb_g	0.709		
Fiber_g	0.826		
VitE_mg	0.547		
Copper_mcg	0.756		
Magnesium_mg	0.640		0.547
Manganese_mg	0.832		
VitA_mcg		0.737	
VitB6_mg		0.800	
VitB12_mcg		0.782	
VitC_mg		0.799	
Riboflavin_mg		0.665	
Iron_mg	0.581	0.596	
Zinc_mg		0.603	
Protein_g			0.822
Fat_g			0.641
Calcium_mg			0.686
Phosphorus_mg			0.853
Selenium_mcg			0.672

	RC3	RC2	RC1
SS loadings	4.224	4.155	3.962
Proportion Var	0.235	0.231	0.220
Cumulative Var	0.235	0.465	0.686

Table4: PCA components

Since the variables are independent, the factor rotation used in PCA analysis was “varimax” with nfactors = 3

the components generated from the CPA analysis can be seen in the table 4:

Component RC3, formulated by Carb_g, Fiber_g , VitE_mg, Copper_mcg, Maganese_mg. Since, Magnesium_mg is explained by 64% of the variance in RC3 however only 54% by RC1 therefore, therefore, we choose to keep Magnesium_mg with RC3. Since this group of food is rich in fiber, Vitamin E as well as carbohydrates, therefore, it was better to rename RC3 as FiberRichedCarbsFood.

Component RC2, formulated by VitA_mcg, VitB6_mg, VitB12_mcg, VitC_mg, Riboflavin_mg, Zinc_mg. Since, Iron_mg was explained by 60% of variance in RC2 however only 58% by RC3 therefore, it was better to keep Iron_mg with RC2. Since, this group of food is rich in Vitamins, iron, and zinc and the foods rich in vitamin C helps in the absorption of iron therefore, they are good for anemic people hence, component RC2 renamed as Anti-anemicFood.

Component RC1 is formulated by Protein_g, Fat_g, Calcium_mg, Phosphorus_mg, Selenium_mcg. Since this group of food is rich in protein and fats, therefore, it was better to rename component RC1 as HighProteinFoods.

Factor analysis on data gave some interesting results. The first step is to get how many factors to be utilized for that two methods were applied on the data were eigenvalue and the scree plot. According to the screen plot, there are a total 5 components or factors we are applying. After that we will see what factors are perfect for the further analysis. Firstly, using the varimax rotation total 5 factors were rotated with the cutoff of 0.55.

```
Loadings:
              Factor1 Factor2 Factor3 Factor4 Factor5
VitB6_mg       0.736
Folate_mcg     0.681
Niacin_mg      0.825
Riboflavin_mg  0.783
Thiamin_mg     0.674
Iron_mg        0.564
VitA_mcg               0.779
VitB12_mcg             0.733
Copper_mcg             0.724
Protein_g                      0.975
Energy_kcal                            0.834
Fat_g                                  0.987
Carb_g                                         0.959
Sugar_g                                        0.596
VitC_mg
VitE_mg
Calcium_mg
Magnesium_mg
Manganese_mg
Phosphorus_mg
Selenium_mcg
Zinc_mg

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      3.631   1.958   1.912   1.844   1.808
Proportion Var   0.165   0.089   0.087   0.084   0.082
Cumulative Var   0.165   0.254   0.341   0.425   0.507
```

We can see most of the variables are contributing to the first factor. The second factor is consisting of 3 variables. The third factor consists of 1 variable. Factor forth contains the 2 number of variables. Factor 5 has 2 variables. Factor 4 and Factor 5 have the same number of variables. As all the factors sum of squares loadings are greater than 1 we will keep all the factors for the analysis. Proportion of variance by each factor individually and the first actor contributes a total of 16.5% variance, the 2nd factor contributes a total of 8.9% variance, 3rd factor contributes 8.7% variance, 4th factor and 5th factor contributes 8.4% and 8.2% of total variance. From the analysis, factors are named as follows:

Factor 1 : Dietary   Factor 2: Minerals   Factor 3 : Aminos   Factor 4: Strength

Factor 5 : Disaccharides

Canonical correlation Analysis between RC3(FiberRichedCarbsFood) and RC1(HighProtienFood).

The variables from the three components in PCA analysis have correlation among themselves and thus a canonical correlation analysis is conducted using the results or components obtained from the PCA analysis. CC analysis for the components RC3 and RC1 is conducted by considering 3 components which are most significant according to the wilks lambda test.

Table2: Canonical Correlations between RC3(FiberRichedCarbsFood) and RC1(HighProtienFood):

CV 1	CV 2	CV 3	CV 4	CV 5
0.79629917	0.37740203	0.36969654	0.25968228	0.04818121

79% of overlapping variance is explained by the first canonical variate CV1. Similarly 37% of the overlapping variance is explained by the second canonical variate CV2.

Canonical correlation analysis between RC3(FiberRichedCarbsFood) and RC2(Anti-Anemic Food):.

As mentioned earlier, the three components are obtained from results of PCA analysis. CCA is performed on the following two components, RC3(FiberRichedCarbsFood) and RC2(Anti-Anemic Food).

Table3: Canonical Correlations between RC3(FiberRichedCarbsFood) and RC2(Anti-Anemic Food):

Canonical Correlations:

CV 1	CV 2	CV 3	CV 4	CV 5	CV 6
0.84323078	0.55744165	0.42056426	0.25938473	0.08425045	0.02286372

There is medium positive correlation between both the groups which is being explained in the CV1 and CV2.
Based on the CV1: 84% of the overlapping variance is explained between the canonical variate pairs of FiberRichedCarbsFood and Anti-Anemic Food. There is 16% of the variance unexplained. **Based on the CV2:** 55% of the overlapping variance is explained

between the canonical variate pairs of FiberRichedCarbsFood and Anti-Anemic Food. There is 45% of the variance unexplained.

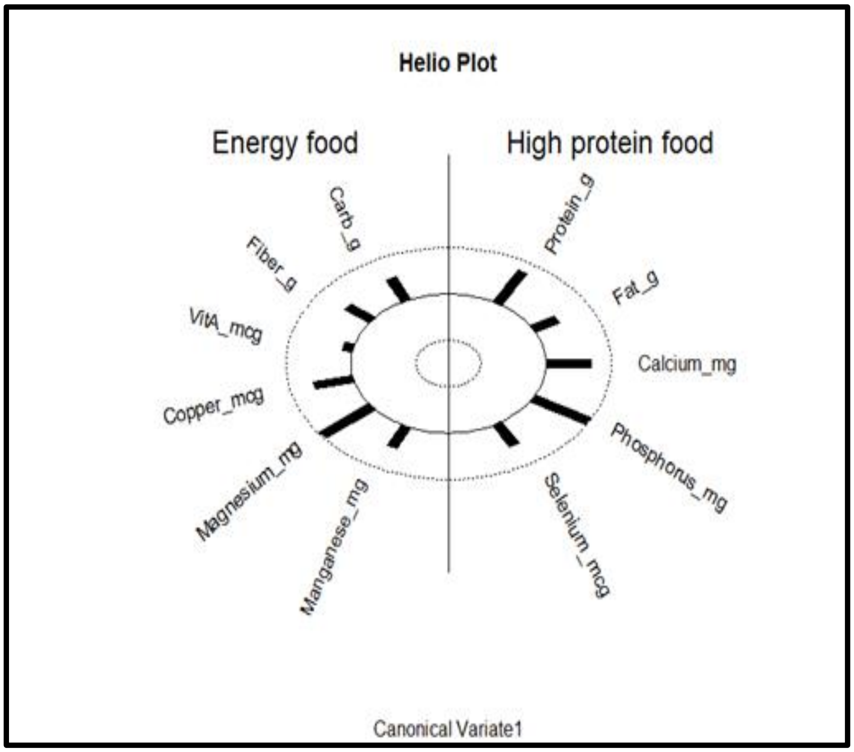


Fig 5 : Helio Plot of FiberRichedCarbsFood and HighProteinFoods

From the helio plot of the first variate we see that the foods with high Magnesium content are also more likely to be rich in phosphorus. Moreover, the variables in Energy food sources which are also called FiberRichedCarbsFood have positive relation among themselves in the first variate.

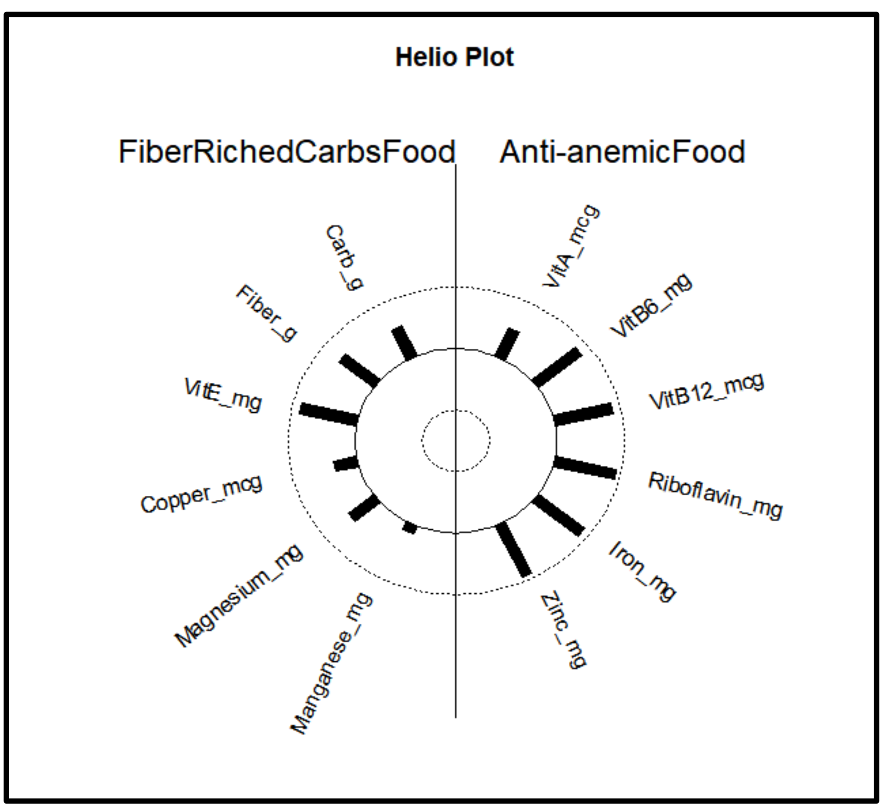


Fig 6 : Helio Plot of FiberRichedCarbsFood and Anti-anemic Food

This makes VitE_mg and Fiber_g as the most important and influential variables for its covariates which is the predictor variate FiberRichedCarbsFood. VitB6_mcg, VitB12_mcg,Riboflavin_mg, Iron_mg , and Zinc_mg are the most important and influential variables for its covariates which is the predictor variate Anti-Anemic Food. Also, the variables seem to have positive correlation among themselves.

Also, on looking at the table 2: we can see that 79% of overlapping variance is explained by the first canonical variate CV1

From table3: 84% of the overlapping variance is explained between the canonical variate pairs of FiberRichedCarbsFood and Anti-Anemic Food. There is 16% of the variance unexplained.

Future Works:

- This study is done on the outdated version SD27 of USDA National Nutrient Database which has only 35 nutrient variables however, the latest version has upto150 nutrient components.
- This study only shows how the food products are rich or deficits in different good components however, it does not cluster the food products or labels the good product into categories on the basis of their nutrient factors.
- This study does not include the relationship between the components produced from the factor analysis which can be done in the future analysis.

Limitations:

- The study does not identify the amount / portion of the nutrients are needed in our diet.
- Absence of data to study the estimation of the risk of disease with nutrient patterns.

Conclusion:

It was found that to stay healthy and energetic minimalistically, it is important to have a complete diet that includes all five nutrients in our diet: Protein , Fat , Carbohydrates , Fiber and Magnesium. Including high protein food like lean meats ,Poultry,Seafood,Dairy Products ,nuts etc. In combination with the protein the diet should include good fat sources like avocados,Cheese,Coconut and The good carb sources like Quinoa ,Oats , Bananas , Sweet Potatoes, Oranges etc
Depending upon the need of the individual, it is possible to focus on food products that are rich in specific sets of nutrients: Fiber - riched Carbohydrates that are a rich source of fiber, as fiber itself is a form of carbohydrate. A high-fiber diet may protect against conditions like tips and heart disease. Anti-anemic nutrient : includes high content of iron with vitamin C and zinc anti-anemic diet is good for people struggling with anemia. High-Protein includes proteins and good fats that are good for building muscles and are very important especially for athletes.

It is also possible to focus on food product that are rich in specific set of nutrient in a different level well as: Dietary: food items rich in dietary nutrients are rich in vitamin b6, iron, thiamin etc. and they are involved in many processes of our body and are necessary for normal cell growth and function. It can be found in certain foods such as dairy, meat, eggs, nuts, enriched flour, and green. vegetables. Minerals: mineral riched diet includes vitamin b 12, copper, vitamin A and they can be found in nuts and seeds , oysters and liver. They are involved in the form of red blood cells, bone, connective tissue and some important enzymes. Aminos: They are very important for building muscle, regulating gene expression, cell signaling, and immunity and found in food products like soy sauce. Strength: The strength includes energy and fats. Disaccharides :Includes sugar and vitamin C.

It was also observed that the nutrients patterns are related to each other as : In food items as Fiber- rich Carbohydrates increases, Anti-anemic food nutrients also increase. Similarly, as Fiber- rich Carbohydrates increase in the food items, high protein nutrients are also observed to be increased.

References:

[1] Pehrsson, P. R., Haytowitz, D. B., Holden, J. M., Perry, C. R., & Beckler, D. G. (2000). USDA's National Food and Nutrient Analysis Program: Food Sampling. *Journal of Food Composition and Analysis*, *13*(4), 379–389. https://doi.org/10.1006/jfca.1999.0867

[2] US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Release 28 (Slightly revised). Version Current: May 2016. Internet: http://www.ars.usda.gov/ba/bhnrc/ndl

[3] McCann, S., Marshall, J., Brasure, J., Graham, S., & Freudenheim, J. (2001). Analysis of patterns of food intake in nutritional epidemiology: Food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutrition, 4(5), 989-997. doi:10.1079/PHN2001168

[4] Stricker, M., Onland-Moret, N., Boer, J., Schouw, Y. V., Verschuren, W., May, A., . . . Beulens, J. (2013). Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. *Nutrition, Metabolism and Cardiovascular Diseases, 23*(3), 250-256. doi:10.1016/j.numecd.2012.02.006

[5] Uusitalo, L., Nevalainen, J., Salminen, I., Ovaskainen, M., Kronberg-Kippilä, C., Ahonen, S., . . . Virtanen, S. M. (2011). Fatty acids in serum and diet - a canonical correlation analysis among toddlers. *Maternal & Child Nutrition, 9*(3), 381-395. doi:10.1111/j.1740-8709.2011.00374.x https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-8709.2011.00374.x