University of Texas at San Antonio

Nutrient Composition Covariance and Cognitive Health

A Factor Analytic Approach

Joshua Villarreal, Garrett Smith

University of Texas at San Antonio

STA-3013 001: Applied Multivariate Analysis

Dr. Anuradha Roy

Fall 2025

Joshua Villarreal

Dr. Anuradha Roy

STA-3013 001

11/20/2025

Nutrient Composition Covariance and Cognitive Health

INTRODUCTION:

Dietary composition consists of many interrelated nutrients that tend to coincide in patterns rather than in isolation. Understanding how nutrients cluster together can provide insight into broader dietary behaviors that may have implications for health and cognition. The goal of this project was to examine how nutrients relate to one another, identify latent dietary factors through factor analysis, and interpret these factors as meaningful dietary patterns. As an extension of this work, we explored whether these extracted dietary patterns showed any association with cognitive performance in adults from the 2013–2014 National Health and Nutrition Examination Survey (NHANES).

DATA SELECTION:

Nutrient Composition Data

Out of 168 dietary variables present in the data set, we selected 26 variables which consisted of macronutrients, vitamins, minerals, cholesterol, and caffeine. Nutrient components were selected based on their relevance to public nutrition guidelines, and clear interpretability. In doing so, the factor interpretations and conclusions found can be utilized by observers. To ensure the model remained stable, we first screened the data to remove structurally redundant variables.

For example, alpha- and beta-carotene were excluded because they are precursors to Vitamin A; keeping both would cause extreme multicollinearity. By filtering out these known biological overlaps, the Principal Component Analysis (PCA) was better able to capture the underlying variance and identify distinct nutrient groupings that aren't as commonly linked.

Cognitive Function Data

From the cognitive function data, six standardized tests measured memory, processing speed, and verbal fluency. After standardizing each test, we created a cognitive composite score by averaging their standardized test scores.

The variables for both data sets are listed below:

Dietary data (26 variables):

DR1TKCAL - Calorie Count
DR1TPROT - Protein
DR1TCARB - Carbohydrate
DR1TSUGR - Sugar
DR1TFIBE - Fiber
DR1TTFAT - Total Fat
DR1TSFAT - Saturated Fat
DR1TMFAT - Monosaturated Fat
DR1TPFAT - Polyunsaturated Fat
DR1TCHOL - Total Cholesterol
DR1TATOC - Vitamin E
DR1TVARA - Vitamin A
DR1TVB6 - Vitamin B6
DR1TFOLA - Total folate
DR1TCHL - Total choline
DR1TVB12 - Vitamin B12
DR1TVC - Vitamin C
DR1TVD - Vitamin D
DR1TVK - Vitamin K
DR1TCALC - Calcium
DR1TMAGN - Magnesium
DR1TIRON - Iron
DR1TZINC - Zinc
DR1TSODI - Sodium
DR1TPOTA - Potassium
DR1TCAFF - Caffeine

Cognitive function data:

CFDCST1 - Score Trial 1 Recall
CFDCST2 - Score Trial 2 Recall
CFDCST3 - Score Trial 3 Recall
CFDCSR - Score Delayed Recall
CFDAST - Animal Fluency: Score Total
CFDDS - Digit Symbol: Score

METHODS:

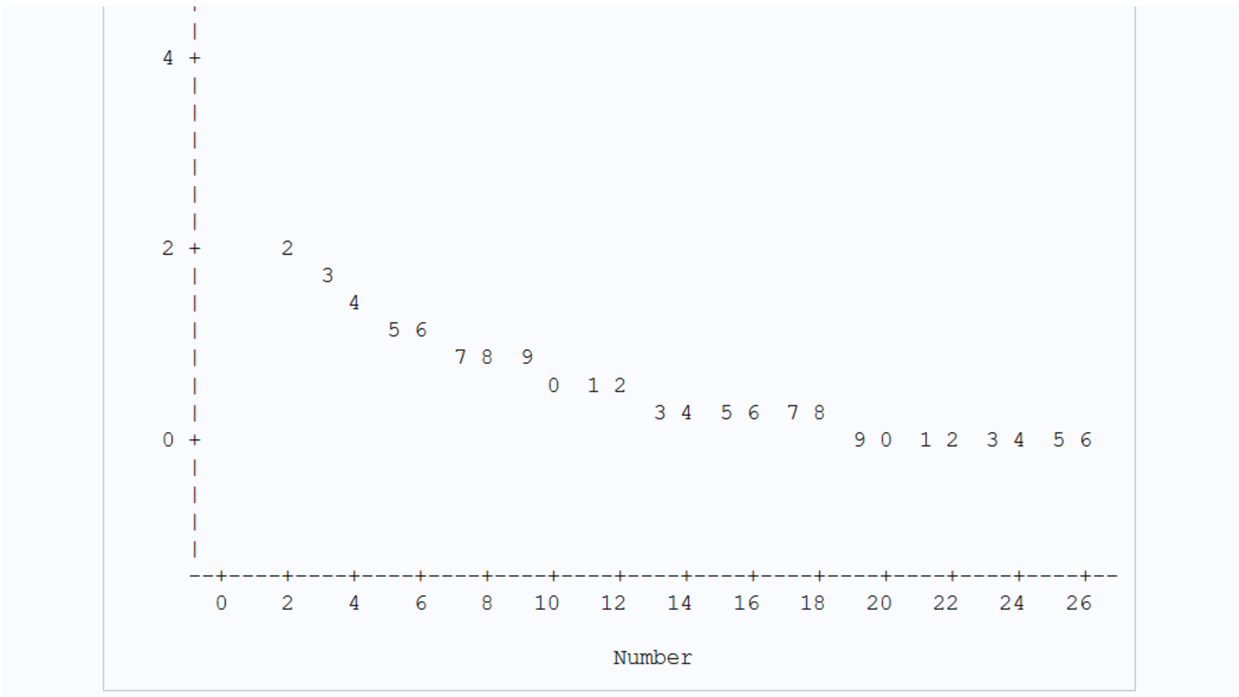
Factor Analysis of Nutrient Composition

The central analytical method of this project was factor analysis, used to uncover underlying patterns in nutrient intake. The extraction method used was Principal Factor and varimax rotation was used to improve interpretability of the factors.

Factor selection

Determining the number of factors (latent variables) that should be used can be done by observing the scree plot, the eigenvalues, and the residuals (error values). We determined that 6 factors were appropriate to reduce dimensionality, increase interpretability of factors, and to keep residuals low.

Scree plot:



The numbers listed on the plot correspond to the eigenvalues of the variables (Eigenvalue 1 excluded for fitting reasons). The number of factors that should be used can be determined by selecting which eigenvalue the “elbow” is, (the point where the slope is steep to the left but at the same time not very steep to the right). You then subtract the number of substantial eigenvalues by 1, giving you the number of factors that should be used. Here, the argument can be made that the eigenvalue is at 7, so 7 -1 = 6 factors. On the contrary, arguments can be made that the “elbow” is at eigenvalues 10 or 13 or even 19. So, we move on to examining the eigenvalues.

Eigenvalues:

Eigenvalues of the Correlation Matrix: Total = 26 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	12.6884261	10.6827279	0.4880	0.4880
2	2.0056983	0.3570499	0.0771	0.5652
3	1.6486484	0.2798241	0.0634	0.6286
4	1.3688242	0.2544662	0.0526	0.6812
5	1.1143581	0.0399413	0.0429	0.7241
6	1.0744168	0.1723590	0.0413	0.7654
7	0.9020577	0.1167888	0.0347	0.8001
8	0.7852689	0.0496277	0.0302	0.8303
9	0.7356412	0.1714534	0.0283	0.8586
10	0.5641878	0.0469245	0.0217	0.8803
11	0.5172633	0.0575812	0.0199	0.9002
12	0.4596821	0.0757668	0.0177	0.9179
13	0.3839152	0.0949408	0.0148	0.9326
14	0.2889745	0.0095014	0.0111	0.9437

Eigenvalues greater than 1 are appropriate for factor number selection, and here we can see the first six eigen values (highlighted in yellow) are greater than 1. Additionally, the first 6 eigenvalues capture 76.54% of the variance of the original data, which is also sufficient.

Residuals:

When running the Proc Factor procedure in SAS, you may select the number of factors you want to use and may include the residuals using the “res” command. A successful factor model will reproduce the correlation matrix of your original variables as closely as possible, while only using the extracted factors. The RMS off-diagonal residuals tell you how well the factor model reproduces the correlation matrix of the variables. In other words, it measures the "leftover correlation" not explained by the factors. An RMS value of less than 0.05 is often deemed appropriate.

Using 6 factors in the proc factor procedure, the output is:

Root Mean Square Off-Diagonal Residuals: Overall = 0.04827641

However, using 5 factors, the output is:

Root Mean Square Off-Diagonal Residuals: Overall = 0.05530477

This suggests 6 factors are appropriate, as the RMS value is under .05 and selecting 5 factors isn't appropriate since the RMS value is over .05. By selecting 6 factors we reduce the dimensionality as much as possible while keeping low error.

FACTOR INTERPRETATIONS:

Using the rotated factor pattern, we identified food groups that represent each factor. Each nutrient component contains a correlation with each factor, indicating how highly it loads on each factor. Values highlighted in yellow represent variables with high loadings on each factor, more specifically loadings greater than .65. Additionally, loadings highlighted in gray represent variables with moderate loadings on a given factor. These correlations are in the 0.45-0.65 range and are important as they further support factor interpretations.

Rotated Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Energy (kcal)	0.74238	0.38859	0.12374	0.46147	0.11283	0.13382
Protein (gm)	0.72035	0.36314	0.34153	-0.01239	0.13103	0.24064
Carbohydrate (gm)	0.44293	0.40449	0.06672	0.71413	0.13214	0.13580
Total sugars (gm)	0.24531	0.09698	0.16186	0.85484	0.10723	0.18508
Dietary fiber (gm)	0.31912	0.65881	-0.08345	0.15202	0.43665	0.01822
Total fat (gm)	0.88207	0.28061	0.06220	0.29365	0.06632	-0.02696
Total saturated fatty acids (gm)	0.78038	0.22183	0.19628	0.36540	-0.05725	-0.09942
Total monounsaturated fatty acids (gm)	0.86376	0.28699	0.03187	0.22787	0.07813	0.00461
Total polyunsaturated fatty acids (gm)	0.77500	0.25644	-0.10853	0.19402	0.19858	0.00222
Cholesterol (mg)	0.75763	-0.05848	0.38957	-0.14191	0.08659	0.23747
Vitamin E as alpha-tocopherol (mg)	0.45218	0.48680	0.01240	0.06779	0.33339	-0.00765
Vitamin A, RAE (mcg)	0.11749	0.28114	0.53916	0.09142	0.40897	-0.10729
Vitamin B6 (mg)	0.20521	0.53300	0.35610	0.02782	0.13525	0.46376
Total folate (mcg)	0.21812	0.79854	0.17120	0.15003	0.18570	0.01296
Total choline (mg)	0.72025	0.13969	0.42697	-0.09920	0.23099	0.30588
Vitamin B12 (mcg)	0.17819	0.40739	0.65923	0.06950	-0.08623	0.25640
Vitamin C (mg)	-0.04925	0.10457	0.19433	0.29731	0.69485	0.01562
Vitamin D (D2 + D3) (mcg)	0.09237	0.04206	0.79484	0.10626	0.08836	-0.11192
Vitamin K (mcg)	0.17919	0.12321	-0.01606	-0.10467	0.73430	0.00193
Calcium (mg)	0.36261	0.37462	0.49459	0.35129	0.07369	-0.21122
Magnesium (mg)	0.48520	0.57408	0.14297	0.14198	0.41182	0.18215
Iron (mg)	0.24781	0.77551	0.26535	0.19324	0.03169	-0.02570
Zinc (mg)	0.45278	0.63332	0.36815	0.05553	0.01546	0.10919
Sodium (mg)	0.71810	0.35842	0.15919	0.15752	0.09585	0.15693
Potassium (mg)	0.49763	0.42087	0.28837	0.22537	0.46984	0.24362
Caffeine (mg)	0.11583	0.02154	-0.11161	0.17432	-0.01184	0.76565

Factor 1: Dense, Processed Foods & Animal Protein

This factor captures a pattern high in energy-dense, fat-rich, animal-based foods. It is defined by very strong correlations with multiple forms of dietary fat and associated nutrients:

- Total fat (0.88207)
- Monounsaturated fat (0.86376)
- Polyunsaturated fat (0.77500)
- Saturated fat (0.78038)
- Cholesterol (0.75763)
- Energy (0.74238)
- Protein (0.72035)
- Choline (0.72025)
- Sodium (0.71810)

These nutrients commonly appear together in animal products and processed foods. A supportive loading from potassium (0.49763) aligns with foods such as meats and prepared meals that include potassium-containing additives.

<u>Factor 2</u>: Fiber-Rich Grains and Legumes

This factor reflects plant-based, whole-grain, and legume-rich dietary patterns. It is driven by strong correlations with:

- Folate (0.79854)
- Iron (0.77551)
- Dietary fiber (0.65881)

Supportive nutrients include zinc (0.63332), magnesium (0.57408), vitamin B6 (0.53300), and vitamin E (0.48680), which often co-occur in fortified grains, legumes, nuts, and seeds. This profile represents nutrient-dense plant sources.

<u>Factor 3</u>: Dairy

This factor aligns with a classic dairy nutrient signature:

- Vitamin D (0.79484)
- Vitamin B12 (0.65923)

Both nutrients are concentrated in milk, yogurt, and fortified dairy products. Supportive loadings on calcium (0.49459) and vitamin A (0.53961) further reinforce this interpretation, as these nutrients also commonly originate from dairy.

Factor 4: Simple Carbohydrates & Sugars

This factor isolates refined carbohydrate intake:

- Sugars (0.85484)
- Carbohydrates (0.71413)

The very strong sugar loading indicates that this factor reflects sweetened foods, desserts, refined grains, and other rapidly absorbed carbohydrates.

Factor 5: Fruits & Vegetables

This factor represents a nutrient profile found in fresh produce:

- Vitamin K (0.73430)
- Vitamin C (0.69485)

A supportive loading for potassium (0.46984) is consistent with fruits and vegetables such as leafy greens, citrus, and bananas, which are major potassium sources.

Factor 6: Caffeinated Beverages

This factor is dominated by:

- Caffeine (0.76565)

A supportive loading from vitamin B6 (00.46376) suggests co-consumption patterns such as fortified teas, energy drinks, or breakfast products that pair caffeine with B-vitamins. Overall, this reflects intake of coffee, tea, sodas, and energy beverages.

ASSOCIATIONS WITH COGNITIVE FUNCTION

Parameter Estimates					
Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.01024	0.02130	0.48	0.6309
Dense/Processed Foods & Animal Protein	1	0.09778	0.02093	4.67	<.0001
Fiber-Rich Grains & Legumes	1	0.06649	0.02100	3.17	0.0016
Dairy	1	-0.04340	0.01950	-2.23	0.0262
Simple Carbs & Sugars	1	0.02830	0.02318	1.22	0.2223
Fruits & Vegetables	1	0.08227	0.01977	4.16	<.0001
Caffeinated Foods/Beverages	1	0.03664	0.02221	1.65	0.0992

The regression model explained 3.41% of the variance in cognitive performance ($R^2 = 0.0341$). Several dietary patterns displayed statistically significant associations:

- Factor 1 (Processed/Animal Protein): Positive
- Factor 2 (Fiber-Rich Grains & Legumes): Positive
- Factor 5 (Fruits & Vegetables): Positive
- Factor 3 (Dairy): Slight negative association

Patterns based primarily on sugars and caffeine were not significantly associated with cognition. These associations are modest but illustrate how the extracted factors can be applied to health-related outcomes.

Works Cited:

"2013-2014 Dietary Data - Continuous NHANES." *Cdc.gov*, 2025, wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Dietary&Cycle=2013-2014. Accessed 25 Nov. 2025.

"2013-2014 Questionnaire Data - Continuous NHANES." *Cdc.gov*, 2025, wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&CycleBeginYear=2013. Accessed 25 Nov. 2025.