

## Catégorisation des aliments selon leurs propriétés nutritionnelles

Jean Pierre Adiouma NDIAYE, NFEGUE ZOFOA Stewart, MEDEHOUIN Crépin, P. TIENDREBEOGO

*Sous la supervision de Mme Iphygénie SARR*

Ecole nationale de la Statistique et de l'Analyse économique Pierre Ndiaye

Claudel, Dakar, Senegal

<https://www.ensae.sn>

**Résumé** – Les décisions prises lors du Sommet sur la Nutrition pour la croissance (2021), traduites dans les Objectifs de Développement Durable (ODD), visent à assurer la sécurité alimentaire et nutritionnelle de tous. Les défis pour atteindre cet objectif sont au centre des débats même dans les pays développés comme les USA. En supposant que les propriétés nutritionnelles peuvent déterminer les aliments, cet article vise à répondre à la question générale suivante : quelle est la catégorisation des aliments selon leurs propriétés nutritionnelles ? Par une approche basée sur l'Analyse des Composantes Principales, les résultats obtenus ont confirmé les hypothèses émises a priori. En effet, il ressort trois principaux regroupements des aliments : (1) les aliments ayant une forte teneur en Zinc, Carbone, Vitamine C – il s'agit des poissons et viandes blanches ou rouges ; (2) les aliments ayant une faible teneur en protéines et en graisse – ils font référence aux féculents, épices, noix et dans une moindre mesure les casse-croûte et pâtisseries ; (3) et les aliments ayant une forte teneur en Magnésium, Folate et Vitamine E – ce sont les matières grasses. Le présent article suggère que des programmes accrus sont nécessaires pour fournir des aliments apportant différents niveaux de calorie et avec leur bonne combinaison favorisant un bon équilibre alimentaire.

**Mots-clés :** aliments, propriétés nutritionnelles, catégorisation, analyse des composantes principales.

### 1. Introduction

Tout être humain a droit à une alimentation adéquate (Rapport sur la nutrition mondiale, 2022). La concrétisation progressive de ce droit partout dans le monde est indissociable de l'existence de systèmes alimentaires fonctionnels et durables apte à assurer la sécurité alimentaire et nutritionnelle de tous. Par ailleurs la faim demeure un sujet de préoccupation majeur, mais l'excès pondéral et l'obésité progressent à un rythme rapide partout dans le monde, y compris dans les pays à faible revenu et à revenu intermédiaire. Les diverses formes de malnutrition - dénutrition, carences en micronutriments, excès pondéral et obésité - n'épargnent donc désormais aucun pays, qu'il s'agisse de pays à revenu faible, intermédiaire ou élevé. Elles peuvent coexister à l'échelle d'un pays ou d'une communauté, et parfois d'un ménage ou d'un individu, et même, paradoxalement, être liées : il convient donc de les combattre toutes en même temps.

Cependant, cette quête requiert des choix alimentaires à la fois meilleurs pour la santé et plus judicieux du point de vue du développement durable, et qui assurent la sécurité alimentaire et la nutrition de tous, notamment des populations vulnérables qui ont des besoins particuliers en nutriments (comme les jeunes enfants, les adolescentes, les femmes enceintes ou allaitantes, les personnes âgées et les personnes malades) et des populations marginalisées dont l'alimentation est davantage tributaire de facteurs hors de leur contrôle.

Dès lors, trois questions cruciales interpellent les statisticiens. Face à ces quantités de données considérables qui existaient déjà, qui ont été améliorées et étendues à d'autres aspects et à d'autres pays : Comment catégoriser des aliments selon leurs propriétés nutritionnelles ? Sur quelles bases peut-on dire de manière objective qu'il existe des propriétés caloriques et nutritionnelles spécifique à chaque aliment ? Quels outils scientifiques pourraient permettre de déceler ou de tirer des conclusions efficaces sur cette étude.

L'objectif général de cet article est d'étudier les propriétés caloriques et nutritionnelles des aliments et de proposer une catégorisation des aliments selon leurs propriétés. De façon spécifique, il s'agit d'abord de visualiser les liaisons entre les propriétés nutritionnelles et ressemblances entre individus pour ensuite catégoriser les aliments selon leur propriétés nutritionnelles.

Ces objectifs spécifiques sont couplés aux hypothèses suivantes :

- ✓ Les propriétés nutritionnelles déterminent les aliments.
- ✓ Les aliments peuvent être regroupés selon leurs propriétés nutritionnelles

Pour répondre à la question de l'étude, le travail est structuré de la façon suivante : d'abord, nous présenterons les données et la méthodologie utilisée, ensuite nous ferons les analyses préliminaires avant d'entamer l'analyse factorielle des données.

## 2. Présentation des données et de la méthodologie utilisée

La base de données à notre disposition provient du Ministère de l'Agriculture des Etats-Unis et présente des aliments répartis selon leurs apports calorifiques et nutritionnels. Elle est composée de 8618 individus et de 42 variables. Cependant il est clair que la représentation de toutes ces variables et de tous les individus dans le cercle de corrélation va être floue donc ne reflétant aucune information. C'est pourquoi nous avons procédé à une sélection de variables et d'individus :

### O Pour les individus

A ce niveau nous utiliser une librairie de R dénommée explor. Cette dernière permet de faciliter la visualisation et l'exploration de résultats d'analyses en composantes principales (ACP). C'est ainsi que nous avons sélectionner les 20 individus qui ont les plus grandes contributions en considérant la totalité des données de la base de données.

Figure 1: Individus avec les plus grandes contributions

Individus actifs			
Show 20 entries		Search:	
Name	Coord	Contrib	Cos2
2026	46.904	2.68	0.608
1864	46.249	2.61	0.724
1833	44.979	2.47	0.769
1851	41.601	2.11	0.712
4356	34.967	1.49	0.202
1817	31.547	1.21	0.538
5818	27.285	0.91	0.414
7649	26.282	0.84	0.726
7656	26.181	0.83	0.712
5282	26.059	0.83	0.187
5283	25.559	0.80	0.196
1947	25.164	0.77	0.711
5446	25.102	0.77	0.210
1970	24.527	0.73	0.726
4455	24.047	0.70	0.583
7458	23.822	0.69	0.111
7459	23.688	0.68	0.101
4191	23.619	0.68	0.224
5280	22.968	0.64	0.290
3818	22.739	0.63	0.219

**Source :** Ministère de l'agriculture des Etats-Unis, calculs des auteurs

NB : Le *name* correspond dans le tableau d'en haut au numéro de ligne correspondant aux individus concernés dans la base

## O Pour les variables

Concernant les variables, on a remarqué que les aliments selon les nutriments pouvaient être en trois groupes (Proximates, Minerals and Vitamins). Ainsi nous avons sélectionné 10 variables actives et 5 variables supplémentaires selon leurs apports dans l'étude.

### ❖ Pourquoi l'ACP<sup>1</sup> ?

Étant donné que notre base de travail est un tableau d'individus décrits par des variables quantitatives à priori hétérogènes, alors la méthode d'analyse la plus pertinente est l'ACP. Elle vise globalement à dégager les corrélations entre les variables ainsi que similarités entre les individus qui sont décrits par les variables.

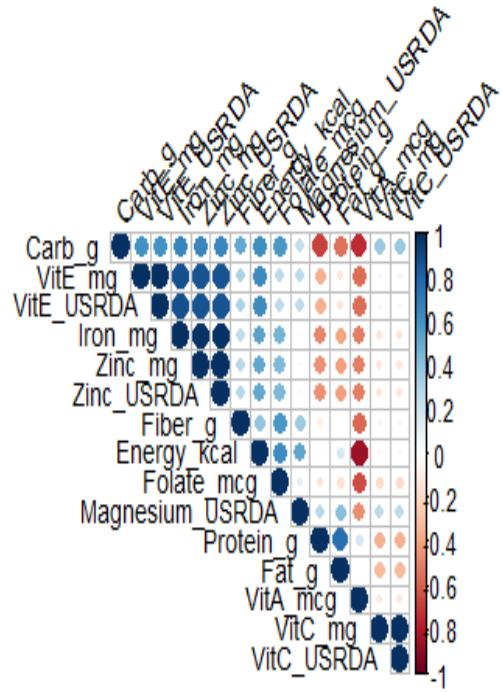
## 3. Analyses préliminaires

Tout d'abord, une étude des corrélations entre les différentes variables de la base de données nous a permis de produire la figure suivante :

---

<sup>1</sup> Une méthode d'analyse des données et plus généralement de la statistique multivariée, qui consiste à transformer des variables liées entre elles (dites « corrélées » en statistique) en nouvelles variables décorrélées les unes des autres

Figure 2: Corrélation entre les variables de la base de données



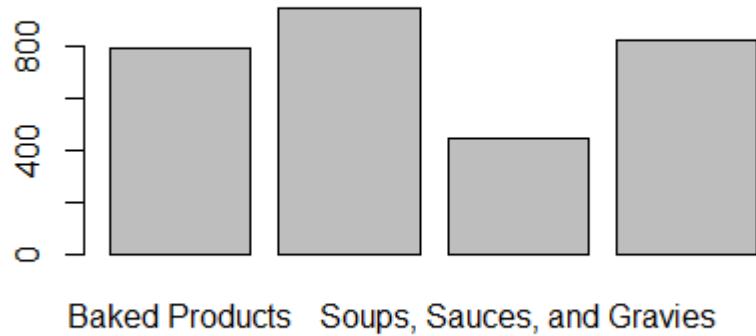
**Source :** Ministère de l'agriculture des Etats-Unis (2020), calculs des auteurs

L'observation de la figure ci-dessus montre que la variable carb\_g est négativement corrélée aux variables VitA\_mcg et Protein\_g. Cela veut dire que les aliments riches en carbone ont une faible teneur en Protéine et en Vitamine A. Cela se confirme avec les épiceries.

Cette figure est intéressante en tenant compte de la liaison entre la variable Energy\_kcal et les autres variables. On remarque que les aliments riches en vitamine A ont une faible teneur en calorie. L'analyse descriptive a également montré que la moyenne des calories rapportées par tous les aliments est de 284.

De surcroit, Les groupes d'aliments qui sont plus fréquentes dans la base de données sont les « Baked Products », « Soups », « Sauces » and « Gravies ».

Figure 3: Fréquences des 4 groupes d'aliments les plus récurrents



**Source :** Ministère de l'agriculture des Etats-Unis (2020), calculs des auteurs

Les données disponibles révèlent que les aliments sont classés suivant 25 groupes : les produits à base de bœuf (les plus représentés, regroupant 10,98% des aliments de la base) ; les produits à base de légumes ; les pâtisseries ; les soupes et sauces ; les produits d'agneau, de veau et de gibier ; les produits de volaille ; les légumineuses ; les fast-foods ; les sucreries ; les fruits et jus de fruits ; les produits à base de porc ; les breuvages ; les produits de poissons et crustacés ; les produits laitiers et œufs ; les saucisses et charcuteries ; les graisses et huiles ; les céréales et pâtes ; les casse-croûte ; les aliments autochtones d'Amérique ; les produits à base de noix et de semences ; les repas, plats principaux et accompagnements ; les aliments de restaurant ; les céréales du petit-déjeuner ; et les épices et herbes (groupe le moins représenté, regroupant 0,74% des aliments). Il est mis en évidence un déséquilibre entre les différents groupes d'aliments, ce qui conforte le fait que l'application de la méthode d'analyse factorielle ne devrait pas inclure le groupe d'aliments comme variable active.

L'analyse descriptive des autres caractéristiques des aliments révèle que leur apport en énergie (exprimée en kilocalories) varie entre 0 kcal et 902 kcal. Selon la littérature, cet apport en énergie dépend des nutriments contenus dans les aliments ; la connaissance des teneurs en nutriments permettrait donc de connaître l'apport calorique. Ainsi, nous pouvons nous focaliser sur l'étude des nutriments afin de catégoriser les aliments. Les nutriments étudiés sont : les protéines, les matières grasses, les hydrates de carbone, le sucre, les fibres, les vitamines A, B6, B12, C, E, B9 (l'acide folique), B3 (la niacine), B2 (la riboflavine), B1 (la thiamine), les minéraux tels que le calcium, le cuivre, le fer, le magnésium, le manganèse, le phosphore, le sélénium et le zinc.

L'étude de la corrélation entre les variables relatives aux nutriments révèle que les teneurs en nutriments (en g, mg ou mcg) sont liées positivement et fortement aux pourcentages en apport nutritionnel selon la recommandation des Etats-Unis (variables de la forme nutriment\_USRDA). En effet, pour les variables VitE\_mg et VitE\_URSDA par exemple, correspondant respectivement à la teneur en mg de vitamine E dans 100g d'aliments et au pourcentage de vitamine E qu'apportent les aliments comparativement à la recommandation journalière des Etats-Unis, l'intensité et le sens de la corrélation entre elles sont très élevés.

Ceci est logique puisque le pourcentage en apport nutritionnel d'un aliment en un nutriment donné se calcule comme le rapport entre la teneur en nutriment de cet aliment et l'apport journalier en ce nutriment recommandé par les Etats-Unis. Ainsi, pour catégoriser les aliments suivant leurs propriétés nutritionnelles, nous allons uniquement utiliser les variables portant sur les teneurs en nutriments, d'autant plus qu'elles renseignent sur plus de nutriments (22 au total) que celles portant sur les pourcentages en apport journalier (15 au total).

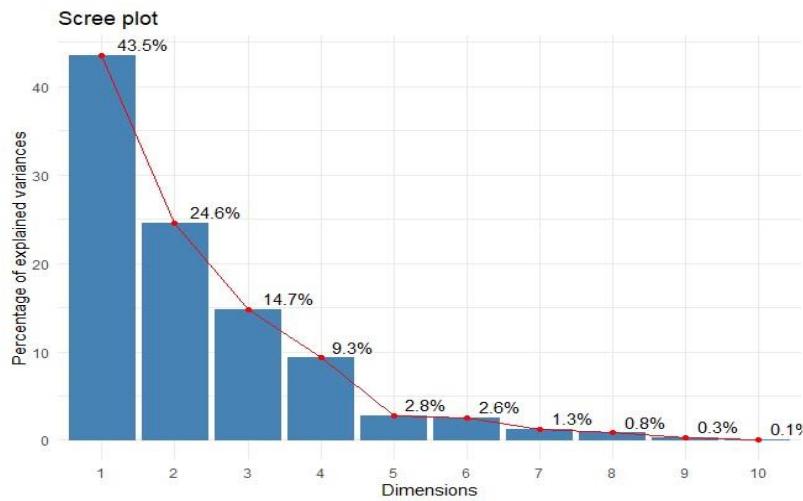
## 4. Analyse factorielle

### 4.1 Choix du nombre d'axes factoriels et des plans factoriels

Une mise en œuvre de l'ACP permet d'avoir l'histogramme des valeurs propres suivantes. Le critère de l'inertie est vérifié pour les deux axes car le premier axe à lui seul contribue à près de 43,5 % de l'inertie totale.

Les 2 premiers axes expriment 68.11% de l'inertie totale ; cela signifie que 68.11% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc convenablement la variabilité contenue dans une grande part des données. Cette valeur est supérieure à la valeur référence de 48.49%, la variabilité expliquée par ce plan est donc significative (cette inertie de référence est le quantile 0.95). Ainsi, sous une base théorique solide nous fixons le nombre d'axes à analyser à 2 pour plus de clarté et pour des raisons pratiques.

Figure 4: Histogramme des valeurs propres

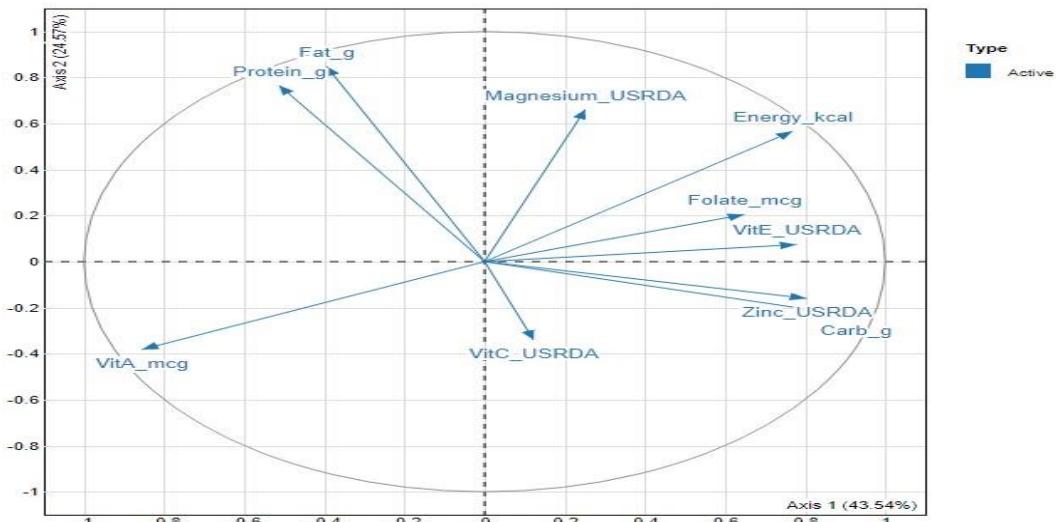


**Source :** Ministère de l'agriculture des Etats-Unis (2020), calculs des auteurs

## 4.2 Analyse du nuage des variables

Le nuage des variables révèle est donné par la figure suivante.

Figure 5: Nuage des variables



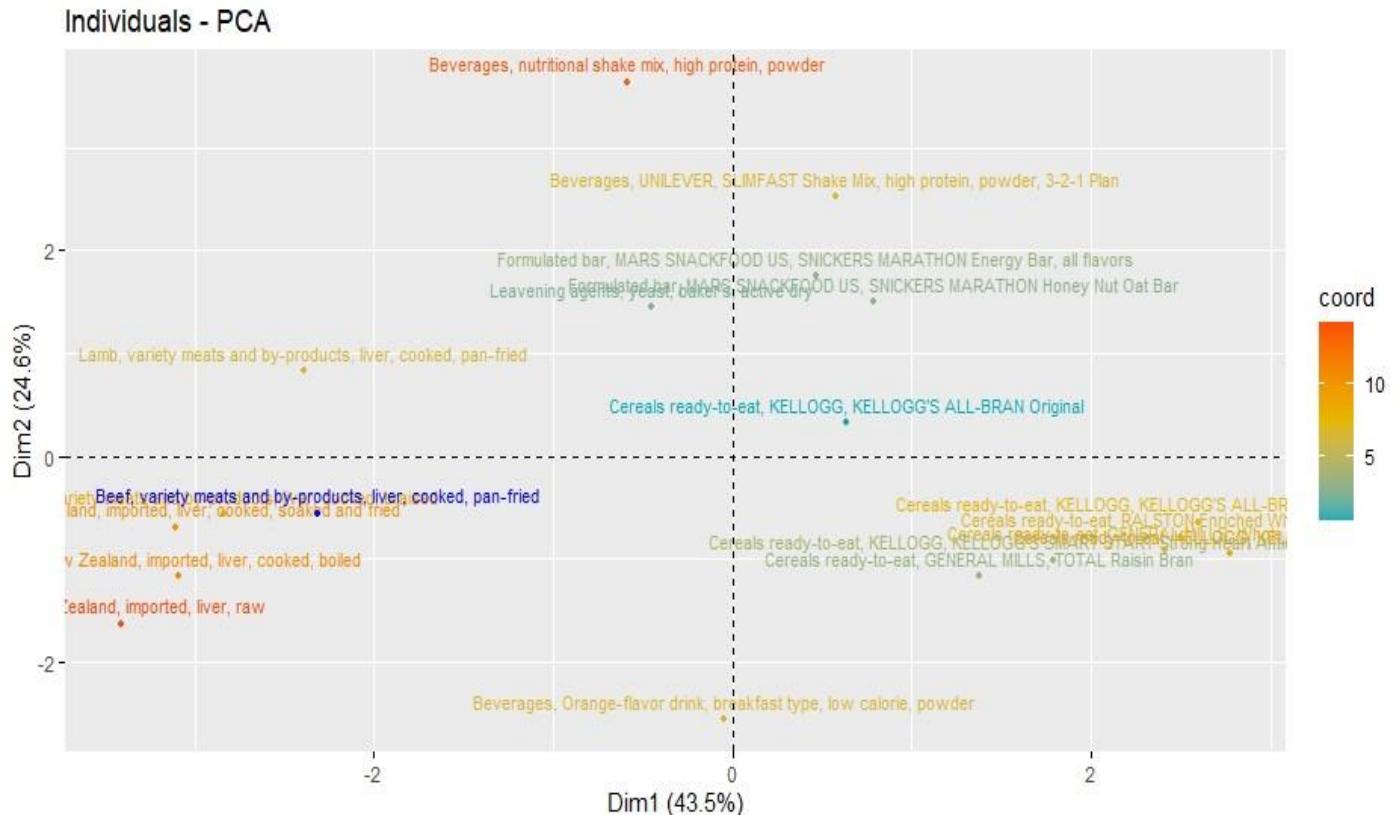
Source : Ministère de l'agriculture des Etats-Unis(2020), calculs des auteurs

A travers la figure ci-dessus, on peut déjà voir se dégager quelques tendances et avoir une idée sur la corrélation des variables. En effet, notre ACP étant normée, le coefficient de corrélation entre deux variables est donné par le cosinus de l'angle formé entre ces variables. On note ainsi une corrélation positive entre des variables telles que Magnesium\_USRDA, Energy\_kcal, Folate\_mcg, VitE\_USRDA. Donc nous pouvons dire que les groupes d'aliments qui ont une forte teneur en Magnesium\_USRDA, Energy\_kcal, Folate\_mcg ont tendance à avoir aussi une forte teneur en Vitamine E et vice-versa. Toutes ces variables sont à leur tour corrélées négativement à des variables telles que VitA\_mcg. C'est-à-dire que les aliments les contenant ont une faible teneur en Vitamine A. D'autre part, les variables Zinc\_USRDA, carb\_g et VitC\_USRDA sont positivement corrélées, c'est-à-dire que les aliments contenant du carbone et du zinc ont une forte teneur en vitamine C. On constate enfin que les variables protein\_g et fat\_g sont corrélées positivement. Nous formons ainsi trois groupes de propriétés nutritionnelles selon leur corrélation.

### 4.3 Analyse du nuage des individus

Considérons la figure suivante qui présente le nuage des individus issu de la mise en œuvre de l'ACP.

Figure 6: Nuage des individus



**Source :** Ministère de l'agriculture des Etats-Unis (2020), calculs des auteurs

Le nuage des individus met en exergue que la majorité des groupes d'aliments sont bien représentés dans le plan 1-2. Les groupes constitués des produits de volailles, de bœufs, de porcs, de poissons et crustacés, d'agneaux, de veaux et gibiers sont proches. Ceci voudrait dire qu'ils ont approximativement les mêmes propriétés nutritionnelles. Ainsi, la dimension 1 oppose des aliments tels que Cereals ready-to-eat, KELLOGG, KELLOGG'S ALL-BRAN COMPLETE Wheat Flakes, Cereals ready-to-eat, KELLOGG, KELLOGG'S PRODUCT 19 et Cereals ready-to-eat à des aliments comme Lamb, New Zealand, imported, liver, cooked, soaked and fried, Beef, New Zealand, imported, liver, cooked, boiled, Beef, New Zealand, imported, liver, raw, Veal, variety meats and by-products, liver, cooked, braised et Lamb, variety meats and by-products, liver, cooked, pan-fried.

La dimension 2 distingue particulièrement des individus tels que *Beverages, nutritional shake mix, high protein, powder* qui ont une forte teneur en *Magnesium\_USRDA* et *Protein\_g*

#### 4.4 Analyse des nuages

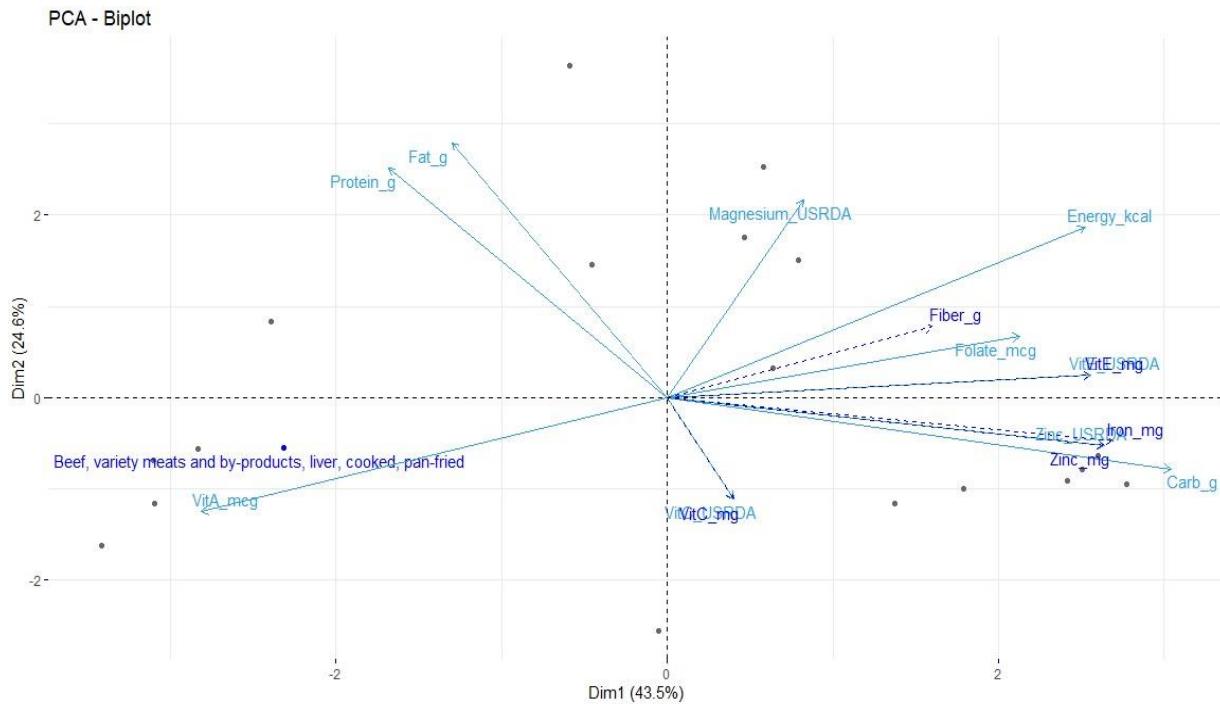
Nous avons enlevé les noms des individus dans ce dernier graphe pour la rendre plus claire. Cependant étant donné qu'on a représenté le nuage des individus uniquement avec leurs étiquettes, on pourra facilement se retrouver. Les aliments tels que Cereals ready-to-eat, KELLOGG, KELLOGG'S ALL-BRAN COMPLETE Wheat Flakes, Cereals ready-to-eat, KELLOGG, KELLOGG'S PRODUCT 19 et Cereals ready-to-eat, GENERAL MILLS, Whole Grain TOTAL appartiennent (caractérisés par une coordonnée positive sur l'axe) ont:

- ❖ une forte teneur en Carb\_g, Zinc\_mg, Zinc\_USRDA, Iron\_mg, VitE\_USRDA, VitE\_mg, Folate\_mcg, Fiber\_g et Energy\_kcal (de la plus extrême à la moins extrême).
- ❖ une faible teneur en Protein\_g, Fat\_g et VitA\_mcg.

Les Lamb, New Zealand, imported, liver, cooked, soaked and fried, Beef, New Zealand, imported, liver, cooked, boiled, Beef, New Zealand, imported, liver, raw, Veal, variety meats and by-products, liver, cooked, braised et Lamb, variety meats and by-products, liver, cooked, pan-fried ont:

- ❖ une forte teneur en VitA\_mcg.
- ❖ une faible teneur en Energy\_kcal, Folate\_mcg, Carb\_g, VitE\_USRDA, VitE\_mg, Iron\_mg, Zinc\_mg, Zinc\_USRDA, Fiber\_g et VitC\_USRDA

Figure 7: Nuage des individus et des variables



**Source :** Ministère de l'agriculture des Etats-Unis (2020), calculs des auteurs

De façon particulière, nous notons à travers cette figure que les aliments tels que les Beef New Zealand imported sont très riches en Vitamine A mais ont une carence en énergie calorifique (energy Kcal), en VitE USRDA et en

Zinc USRDA. D'autre part les aliments tels que les Beverages nutritional shake mix high protein sont très riches en graisse(Fat g) et en protein g. En outre il faut aussi remarquer que le magnésium USRDA est présent dans les aliments tels que Beverages UNILEVER SLIMFAST shack Mix.

**Interprétation des résultats :** L'objectif général de l'étude est de catégoriser. Les analyses préliminaire et factorielles ont permis de montrer que les hypothèses émises à priori sont vérifiées. Les propriétés nutritionnelles déterminent les aliments et Il est bien possible de catégoriser les aliments selon leurs propriétés nutritionnelles. Trois groupes d'aliments sont ainsi formés grâce à l'analyse des composantes principales : (1) les aliments ayant une forte teneur en Zinc, Carbone, Vitamine C ; (2) les aliments ayant une faible teneur en protéines et en graisse (3) et les aliments ayant une forte teneur en Magnésium, Energie calorifique, Folate et Vitamine E. Le troisième groupe attire l'attention car les aliments riches en Magnésium, Folate et Vitamine E sont plus susceptible d'apporter des calories aux personnes. Comme recommandation, il serait important pour les décideurs politiques d'assurer la fourniture d'aliments riches en Magnésium, Folate et Vitamine E comme les haricots et la farine. Une politique favorisant cet accès aurait un impact positif majeur surtout chez les jeunes qui souffrent de malnutrition. L'analyse a également montré une certaine opposition entre les céréales et les viandes, foie. Les institutions sanitaires devraient promouvoir la consommation des aliments qui s'opposent d'après le nuage des individus. En effet cela permettrait de favoriser l'équilibre alimentaire.

## 5. Conclusion

Notre étude, faite au moyen de l'Analyse en Composantes Principales (ACP) et portant sur la Catégorisation des aliments selon leurs propriétés nutritionnelles, a révélé un certain nombre de résultats. L'ACP nous a permis de regrouper d'abord de déceler les corrélations qui existent entre certaines variables mais aussi en second temps nous avons pu atteindre l'objectif principal de cette étude qui consistait à regrouper ou catégoriser les aliments selon leur propriété nutritionnelle. Il ressort trois principaux regroupements des aliments : (1) les aliments ayant une forte teneur en Zinc, Carbone, Vitamine C – il s'agit des poissons et viandes blanches ou rouges ; (2) les aliments ayant une faible teneur en protéines et en graisse – ils font référence aux féculents, épices, noix et dans une moindre mesure les casse-croûte et pâtisseries ; (3) et les aliments ayant une forte teneur en Magnésium, Folate et Vitamine E – ce sont les matières grasses. Toutefois il est important de souligner que la bonne combinaison d'aliments aux propriétés différentes doit favoriser un bon équilibre alimentaire.

## Références bibliographiques

- [1] Benzécri J. P. (1982), Histoire et préhistoire de l'analyse des données, Dunod
- [2] Benzécri J. P. (1973), L'Analyse des données. Tome 1 : La Taxinomie, Dunod
- [3] Benzécri J. P. (1973), L'Analyse des données. Tome 2 : L'Analyse des correspondances, Dunod
- [4] Escofier B. & Pagès J. (2008), Analyses factorielles simples et multiples : Objectifs, méthodes et interprétation, Dunod.
- [5] Volle M. (1997), Analyse des données, Economica

## Annexe

Annexe 1: Les 4 groupes d'aliments les plus fréquents de la base de données..... 12

Annexe 1: Les 4 groupes d'aliments les plus fréquents de la base de données

FoodGroup	Effectif
Vegetables and Vegetable Products	828
Soups, Sauces, and Gravies	452
Beef Products	946
Baked Products	797

**Source :** Ministère de l'agriculture des Etats-Unis (2020), calculs des auteurs