Statistical Methods for Food Composition Database Analysis

Subjects: Nutrition & Dietetics | Statistics & Probability
Contributor: Yusentha Balakrishna, Samuel Manda

A food composition database (FCDB) or nutrient database is a compilation of the chemical composition of food and beverage items, obtained from chemical analyses, estimations from published literature, or unpublished laboratory reports. A summary of the statistical methods that have been directly applied to food composition databases and datasets is described here.

Keywords: food composition database ; nutrient database ; statistical methods ; clustering ; dimension reduction ; regression

1. Introduction

Nutrition relates food to an organism's need for growth, metabolism, and repair. To determine optimum nutrition, knowledge of food components is required. Thus, food composition databases (FCDBs) are pivotal in any quantitative nutrition study. A FCDB or nutrient database is a compilation of the chemical composition of food and beverage items, obtained from chemical analyses, estimations from published literature, or unpublished laboratory reports [1].

One of the major applications of FCDBs is providing data for the estimation of nutrient intakes [2]. Nutritional epidemiology has been highlighted during the last few decades and has become an area of public health importance. Due to the growing demand for food consumption data, much research has been done regarding optimal and sound statistical methods for the assessment of dietary patterns within a population [3][4][5]. There also exists literature [6] and studies that describe the application of statistical methods to food composition data (FCD) itself, but most have conducted chemical analyses to obtain the nutrient data before applying statistical methods [6][7][8]. Other studies that have used data from FCDBs have combined nutritional data with that of dietary reference intakes [9], restaurant menus [10], or household income and food price data [11]. However, in the absence of chemical analyses and secondary data, the opportunities that exist for the statistical analysis of FCD alone need to be described.

2. Standard Statistical Methods

These involved the use of descriptive statistics such as frequencies and proportions, means and medians, and confidence intervals to summarize the data. Moreover, standard inferential statistics, including chi-square, Student's *t*-test, Wilcoxon signed-rank, and Kruskal–Wallis tests, were used in an effort to generalize the findings. Eight papers utilised simple methods in their analysis of food composition databases.

Due to the attention on dietary chloride at the time, Yarbrough Al-Bander et al. [12] studied the distribution of chloride occurring naturally in uncooked food items. The results indicated high variability in the chloride and sodium content among the 216 foods analysed and a strong correlation (correlation = 0.84) between the two components. The mean difference between the chloride and sodium content (calculated as chloride content minus sodium content and measured in millimoles to allow comparison) was not statistically different from zero, indicating a high degree of coupling. These results support the need to analyse nutrients both individually and in combination with other nutrients.

Seeing the need for dieticians to be able to quickly identify foods that were either low, medium, or high in nutrient content, Khan [13] attempted to partition food items in this way. Various thresholds using measures of central tendency were tested and the results were compared. The thresholds were considered unsuitable if a ranking category contained zero food items for any nutrient, if more foods were contained in the 'high' category as opposed to the 'medium' category, and if a category contained either an extremely high count or extremely low count. Of the 20 criteria tested, 2 criteria were considered to be suitable. The criteria suggested using less than 0.5 (or 0.75) × mean, 0.5 (or 0.75) × mean to 2.5 (or 2.75) × mean, and more than 2.5 (or 2.75) × mean as cut-offs for low, medium, and high ranks, respectively. Another approach suggested that for each nutrient, the top 5%, middle 47.5%, and last 47.5% of food items be considered as the thresholds for determining high, medium, and low rankings. However, more advanced methods such as clustering algorithms may be better suited to objectively rank food items by nutrient content. By applying clustering algorithms to individual nutrients as done by Nikitina et al. [14], clusters containing foods with a low nutrient content can be separated from clusters containing a moderate nutrient content and clusters containing a high nutrient content. Thus, a natural data-driven ranking of low, medium, and high nutrient content can be found.

Three studies [15][16][17] investigated changes in the food composition of fruits and vegetables between different versions of FCDBs. All three used geometric means of each of the nutrients and compared them using the Student's *t*-test; however, only Davis et al. [15] adjusted for moisture content. While the studies found significant changes in some nutrient components, they note that the results do not account for confounding arising from changes in chemical analyses and sampling methods and the use of mixed sources of composition data. Analysing different versions of FCDBs may be more useful in investigating the impact on consumption studies [18] rather than historical changes [19].

Regular audits to identify unlikely nutrient values are necessary to maintain a reliable FCDB. Errors could result from coding mistakes, out of range values, or laboratory analysis mistakes. Automating the auditing process will save time and resources by informing the compiler whether selective revision or further laboratory analysis is needed. Chu et al. [20] calculated and ranked coefficients of variation (CV) within each food subgroup and within each nutrient to detect outliers that may be potential errors. A rank of '1' was assigned to the largest CV, '2' to the second largest CV, and so on. The top two ranked CVs in each subgroup and nutrient were flagged as 'hits' (unlikely nutrient values) and products of the subgroup ranks and nutrient ranks that were less than or equal to 20 (since larger CVs for either subgroups or nutrients will have smaller products). The proportion of hits that were regarded as true errors by a panel of experts ranged from 1.4% to 37.6% for the various food groups. The likelihood of error detection increased 38-fold compared to manual detection and this low-cost process could help improve the accuracy of FCDBs.

Pennington and Fisher [21] determined the means and standard deviations of 24 food components in 10 previously determined subgroups (found in Pennington and Fisher [22]). The subgroups were constructed such that the nutritional composition and classification characteristics (part of a plant, colour, botanical family, etc.) for each food item were similar within each group. The subgroups exhibited unique concentrations (tested using Kruskal–Wallis ANOVA with pairwise multiple comparison procedures) of the food components and could assist the design of food frequency questionnaires (FFQs), the evaluation of dietary intake data, and the dietary guidelines provided to patients.

Nguyen et al. [23] used the Friedman and post-hoc Wilcoxon signed-rank test to determine if 'healthier' versions of common foods contained more sugar. By comparing fat-free, low-fat, and regular versions of the same food, it was found that the amount of sugar was higher for fat-free and low-fat versions as compared to the regular versions, despite having a lower caloric content.

3. Regression Methods

Regression methods are a powerful tool in statistics that are widely used to predict the values of dependent variables using information concerning independent variables. Three papers [24][25][26] reported applying regression methods to the FCDB. Two papers [25][26] related concepts from traditional Chinese medicine to food composition. All foods in traditional Chinese medicine are categorised into the four natures: cold, cool, warm, and hot. The purpose of these papers was to examine the association between the nutrient content of these foods and their cold-hot nature category. Liu et al. [25] utilised multiple logistic regression and found fat, carbohydrates, and selenium to be significantly associated with 'hot' foods while copper and iron were significantly associated with 'cold' foods. Xie et al. [26] additionally considered that multiple food components together may influence the cold-hot characteristics of the food. The authors fitted a multivariate ordinal logistic regression model to predict the probability of a food item being hot-, cold-, or plain-natured. Six components (folate, B6, calcium, vitamin A, and caffeine) were included as predictors in the final model. The developed model can be used to objectively classify foods as per traditional Chinese medicine. The study was limited by the lack of data for five food components, resulting in their exclusion from the analyses. Missing data for the included components was replaced by the mean value of observed similar foods. Thus, the nutrient content of food may be one of the distinguishing factors for the traditional Chinese medicine categorisation of the cold-hot nature of foods. These papers exemplify exploring nutrient differences between generally accepted food categories and informs the metabolic effect of consumption from these categories.

Ispirova et al. [24] aimed to address the issue of missing data in FCDBs by using imputation techniques based on statistical prediction. The four imputation methods applied were non-negative matrix factorization (NMF), multiple imputations by chained equations (MICE), nonparametric missing value imputation using random forest (MissForest), and k-nearest neighbours (KNN). These methods were compared with the traditional approaches of fill-in with the mean and fill-in with the median. Two types of datasets were also considered. The first type explored imputing values for the same nutrient in one food item using different national FCDBs. The second type explored imputing nutrient values from the same FCDB using similar food items. Using five regression metrics to assess the performance of the imputation techniques, the study found that the commonly used approaches always resulted in the largest errors. For all the statistical prediction techniques, the error increased with the percentage of missing data. Overall, the MissForest imputation method performed best by yielding the smallest errors. It is necessary to note that imputation using random forests is distribution-free or non-parametric. This makes it more flexible than the distribution-based methods and is more suitable for use when the data contains non-linearities or interactions.

4. Clustering

After standard statistical methods, clustering was the next most utilized statistical method. Cluster analysis is a class of multivariate methods that aim to classify observations into homogenous groups that are different from other groups. Seven papers applied clustering methods to FCDBs. Windham et al. [27] applied the fuzzy c-means clustering algorithm within the dairy, grain, and fat commodity groups to determine foods with a similar nutritional content. The authors opted to classify foods both by nutrients that are scarce in terms of food supply and by nutrients that pose health risks when consumed in excess, thus expanding current groupings.

Akbay et al. [28] aimed to divide lamb meat into compositionally similar groups to offer healthier dietary replacements. The composition data of several lamb preparations was extracted from the FCDB and analysed using agglomerative hierarchical cluster analysis with average linkage. The Euclidean distance was used, and the data was standardised. The results found two distinct clusters (with families and subfamilies) differing in fatty acids, cholesterol, and energy. This distinction is beneficial to the population in general in making informed dietary choices.

Pennington and Fisher [22] aimed to empirically group fruits and vegetables based on food components of public health significance and thereafter, relate them to four classification variables: botanic family, colour, part of the plant, and total antioxidant capacity. The proposed classifications may aid researchers and nutritionists in developing FFQs and providing dietary guidance. The data consisted of food composition information for 37 fruits and 67 vegetables. Missing values were obtained from the literature or imputed. An agglomerative hierarchical clustering method using Ward's linkage was applied to 23 food components, which were standardised using the range method. Graphical (dendrograms) and statistical (pseudo t^2 statistic) methods were used to determine the optimal number of clusters and multivariate analysis of variance (MANOVA) was used to compare differences between cluster means for each of the 23 food components. The study mentioned that while cluster membership differed by cluster methodology, certain fruits and vegetables remained grouped, suggesting value in employing mathematical clustering as a guide toward useful groupings.

A more recent study was documented by do Prado et al. [29] in Brazil. Manufactured food products are often reformulated to reduce the content of negative nutrients, such as sodium, or to increase the content of positive nutrients, such as dietary fibre. This study aimed to compare the change in the nutrient composition of specific Brazilian food groups (259 food items) from 2003 and 2013 using percentage change, hierarchical cluster analysis (HCA), and principal component analysis (PCA). The Brazilian FCDB was used for 2003. Analytical reports and updated data from food manufacturers were used for 2013. Percentage change in component data was classified as either negligible change (0–9.99%), moderate reformulation (10–24.99%), or substantial reformulation (≥25%). HCA was used to separate food items within each food group (cereal and cereal products, meat and meat products, milk and milk products, and manufactured foods) based on their compositional similarities. The results showed that some food items did not cluster together despite having similar matrices. This suggests that using pre-established food groups alone may be inaccurate in assessing changes in nutritional composition and should be combined with multivariate statistical analyses. PCA was also used to partition food items within food groups. After comparison with the HCA results, HCA was found to be the most suitable tool to group the food products concerning their composition. Mean nutrient content for 2003 and 2013 was calculated for each HCA-derived cluster and compared using a paired Student's *t*-test or Wilcoxon test. The percentage change for each cluster was also calculated. do Prado et al. [29] concluded that the joint use of percentage change and other statistical techniques allowed efficient identification of changes in the nutritional composition of food items.

Another study that used both PCA and HCA was conducted by Li et al. [30]. Li et al. [30] analysed 268 raw plant foods from 5 food categories. Applying a PCA showed that cereal grains, nuts and seeds, and legumes could be well separated when considering nutritional content, but fruits and vegetables exhibited significant overlap. A follow-up soft independent modelling of class analogies (SIMCA) analysis suggested that nuts and seeds are similar to all other plants. Finally, using agglomerative hierarchical clustering with Ward's distance, the resulting clusters contained foods from different food categories. Better separation was achieved using clusters based on compositional similarity rather than the food categories.

Atsa'am et al. [31] applied k-means clustering with Euclidean distance to food items within the 'cereals' category of the West African Food Composition Table. Thirteen nutrients were analysed, and the data utilised a min-max normalisation. The within-group and between-group sum of squares was used to validate the clusters found. The extracted clusters separated the food items by the type of grain and preparation method. For example, all millet items occupied its own cluster and boiled grains were distinct from their raw counterparts.

Nikitina et al. [14] examined 330 cottage cheese products and confectionaries and aimed to cluster them by their carbohydrate content using k-means clustering. The algorithm identified clusters that grouped foods as having either a low, medium, or high carbohydrate content. These clusters are useful for diet planning for diabetics.

5. Dimension Reduction Techniques

PCA [32] and factor analysis [33] are data reduction techniques that combine correlated variables into a smaller number of components. Two studies [34][35] investigated the nutrient co-occurrence patterns in foods. Similä et al. [35] used the food composition database of Finland, which consisted of 530 food items, used as ingredients (fresh, uncooked, and edible), and 106 nutrients and non-nutrients (for instance, phytosterols and heavy metals). The analysis method was factor analysis with varimax rotation and was applied to two approaches of the data. The first approach used nutrient values per 100 g of foods whilst the second approach used nutrient values in a portion of 1 MJ of each food (nutrient densities). The nutrient content patterns identified for both approaches were easily interpretable and consistent with prior knowledge of nutrient composition.

Balakrishna et al. [34] also investigated nutrient co-occurrence patterns. PCA with varimax rotation was applied to 971 food items and 28 nutrients from the South African FCDB. PCA was applied to the nutrients, to establish nutrient patterns, and to the food items themselves for validation. The nutrient patterns obtained mirrored the South African food-based dietary guidelines (FBDGs). One of the patterns identified foods that had a high sodium content, and this corresponded with foods identified under the country's national salt regulations. FBDGs recommend the consumption of foods instead of nutrients and changes in food consumption changes the intake of several nutrients, not just one. Thus, information on nutrient co-occurrence patterns within foods is needed.

6. Other Methods

One paper [36] described applying linear programming (LP) and quadratic programming (QP) methods to FCDBs. LP and QP are processes that find optimal solutions to an objective function subject to one or more constraints. FCDBs often contain information for many nutrients but have values for only a small number since chemical analyses are costly. With the many products being introduced each year, obtaining these missing values via chemical analyses is impractical. Calculation of these missing values can be achieved if a food is a composite of other foods in the database and one knows the proportions of the ingredients and the method of preparation. Other methods using food labels are available but can be time-consuming. Thus, Westrich et al. [36] aimed to develop a mathematical optimisation software that estimates unknown nutrient estimates in products. The optimisation software was able to estimate nutrient values four times faster than conventional (trial-and-error) methods with the same degree of accuracy, although the QP method was slightly slower than LP. The LP version of the software was adopted to help maintain the database.

Ispirova et al. [37] also investigated the missing data problem. Traditionally, missing nutrient values are borrowed from other countries. However, we must be cognizant of the different geographical elements between countries and differing qualities of FCDBs. In practice, borrowing is achieved using the FCDB of a neighbouring country or from that which has a large span of data. The authors developed a method to objectively, rather than subjectively, borrow nutrient values using null hypothesis testing. After ensuring the similarity in the method type (analytical values were used) and food items between the FCDBs of 10 countries, the value for a specific nutrient and food item was compared across the FCDBs using the appropriate statistical tests and post hoc procedures, e.g., Friedman test with Nemenyi and Holm post hoc procedures. This methodology (Missing Nutrient Value Imputation UsinG Null Hypothesis Testing—MIGHT) gave more accurate results for imputation when compared to currently used techniques. MIGHT supports the premise that proper statistical analysis can improve the missing data problem of FCDBs while minimising the error of current nutrient imputation practices.

Phanich et al. [38] applied the Self-Organising Map (SOM) and k-means clustering to Thai food composition data to develop a food recommendation system for diabetic patients. This system enabled diabetic patients to find suitable substitutions, similar in the nutrient composition and characteristics, for food items. The SOM, or Kohonen network, is a type of artificial neural network used for the visualisation and analysis of high-dimensional data. The data is described by a finite set of models, which are associated with neurons or nodes in the network. The closer the nodes in the network, the more similar the models. The dataset consisted of 290 Thai food dishes and 8 nutritionist-recommended nutrients that influence diabetes. Food items were grouped by their characteristics (such as noodles, rice, and fried food) and were also grouped by a nutritionist into normal food, limited food, and avoidable food based on dietary recommendations for diabetics. The two-stage analysis first constructed and trained the SOM before clustering the SOM using the k-means approach. The resulting clusters contained foods that provided similar amounts of the eight nutrients. The system scored well amongst the nutritionists who were invited to evaluate it.

Kim et al. [39] employed network-based approaches in their analysis of FCDBs. Network-based approaches and average linkage hierarchical clustering methods were applied to consolidate foods that had almost identical nutrient contents, and then to connect foods with similar nutrient contents. The aim was to determine the nutritional balance of food items to aid the design of healthy diets. Many individual nutrients were identified that instrumentally contributed to the nutritional balance. However, pairs of nutrients could also impact the nutritional balance of food while the individual nutrients alone may not. This additional complexity was quantified using Pearson correlations between nutrients (across foods). The network-based approach provided an overview of the relationships between foods and food groups and the relationships between nutrients.

References

1. Greenfield, H.; Southgate, D.A.T. Food Composition Data. Production Management and Use, 2nd ed.; Food and Agriculture Organization of the United Nations: Rome, Italy, 2003.
2. Elmadfa, I.; Meyer, A.L. Importance of food composition data to nutrition and public health. Eur. J. Clin. Nutr. 2010, 64 (Suppl. S3), S4–S7.
3. Reedy, J.; Subar, A.F.; George, S.M.; Krebs-Smith, S.M. Extending Methods in Dietary Patterns Research. Nutrients 2018, 10, 571.
4. Zhang, F.; Tapera, T.M.; Gou, J. Application of a new dietary pattern analysis method in nutritional epidemiology. BMC Med. Res. Methodol. 2018, 18, 119.
5. Zhao, J.; Li, Z.; Gao, Q.; Zhao, H.; Chen, S.; Huang, L.; Wang, W.; Wang, T. A review of statistical methods for dietary pattern analysis. Nutr. J. 2021, 20, 37.
6. Granato, D.; Ares, G. Mathematical and Statistical Methods in Food Science and Technology; John Wiley & Sons, Ltd.: West Sussex, UK, 2014.
7. Cruz, A.G.; Cadena, R.S.; Alvaro, M.B.V.B.; Sant'Ana, A.S.; Oliveira, C.A.F.; Faria, J.A.F.; Bolini, H.M.A.; Ferreira, M.M.C. Assessing the use of different chemometric techniques to discriminate low-fat and full-fat yogurts. LWT-Food Sci. Technol. 2013, 50, 210–214.
8. da Silva Torres, E.A.F.; Garbelotti, M.L.; Moita Neto, J.M. The application of hierarchical clusters analysis to the study of the composition of foods. Food Chem. 2006, 99, 622–629.
9. Kim, J.H.; Kim, W.C.; Kim, J. A practical solution to improve the nutritional balance of Korean dine-out menus using linear programming. Public Health Nutr. 2019, 22, 957–966.
10. Rudelt, A.; French, S.; Harnack, L. Fourteen-year trends in sodium content of menu offerings at eight leading fast-food restaurants in the USA. Public Health Nutr. 2014, 17, 1682–1688.
11. Colchero, M.A.; Guerrero-López, C.M.; Molina, M.; Unar-Munguía, M. Affordability of food and beverages in Mexico between 1994 and 2016. Nutrients 2019, 11, 78.
12. Yarbrough Al-Bander, S.; Nix, L.; Katz, R.; Korn, M.; Sebastian, A. Food chloride distribution in nature and its relation to sodium content. J. Am. Diet. Assoc. 1988, 88, 472–475.
13. Khan, A.S. Processes in ranking nutrients of foods in a food data base. Nutr. Health 1996, 11, 59–72.
14. Nikitina, M.A.; Chernukha, I.M.; Uzakov, Y.M.; Nurmukhanbetova, D.E. Cluster analysis for databases typologization characteristics. News Natl. Acad. Sci. Repub. Kaz. Ser. Geol. Tech. Sci. 2021, 2, 114–121.

15. Davis, D.R.; Epp, M.D.; Riordan, H.D. Changes in USDA Food Composition Data for 43 Garden Crops, 1950 to 1999. J. Am. Coll. Nutr. 2004, 23, 669–682.

16. Mayer, A.M. Historical changes in the mineral content of fruits and vegetables. Br. Food J. 1997, 99, 207–211.

17. White, P.J.; Broadley, M.R. Historical variation in the mineral composition of edible horticultural products. J. Hortic. Sci. Biotechnol. 2005, 80, 660–667.

18. Ahuja, J.K.C.; Goldman, J.D.; Perloff, B.P. The effect of improved food composition data on intake estimates in the United States of America. J. Food Compost. Anal. 2006, 19, S7–S13.

19. Marles, R.J. Mineral nutrient composition of vegetables, fruits and grains: The context of reports of apparent historical declines. J. Food Compost. Anal. 2017, 56, 93–103.

20. Chu, C.-M.; Lee, M.-S.; Hsu, Y.-H.; Yu, H.-L.; Wu, T.-Y.; Chang, S.-C.; Lyu, L.-C.; Chou, F.-J.; Shao, Y.-P.; Wahlqvist, M.L. Quality assurance with an informatics auditing process for Food Composition Tables. J. Food Compost. Anal. 2009, 22, 718–727.

21. Pennington, J.A.T.; Fisher, R.A. Food component profiles for fruit and vegetable subgroups. J. Food Compost. Anal. 2010, 23, 411–418.

22. Pennington, J.A.T.; Fisher, R.A. Classification of fruits and vegetables. J. Food Compost. Anal. 2009, 22, S23–S31.

23. Nguyen, P.K.; Lin, S.; Heidenreich, P. A systematic comparison of sugar content in low-fat vs regular versions of food. Nutr. Diabetes 2016, 6, e193.

24. Ispirova, G.; Eftimov, T.; Seljak, B.K. Evaluating missing value imputation methods for food composition databases. Food Chem. Toxicol. 2020, 141, 111368.

25. Liu, C.; Sun, Y.; Li, Y.; Yang, W.; Zhang, M.; Xiong, C.; Yang, Y. The relationship between cold-hot nature and nutrient contents of foods. Nutr. Diet. 2012, 69, 64–68.

26. Xie, A.; Huang, H.; Kong, F. Relationship between food composition and its cold/hot properties: A statistical study. J. Agric. Food Res. 2020, 2, 100043.

27. Windham, C.T.; Windham, M.P.; Wyse, B.W.; Hansen, R.G. Cluster-Analysis to Improve Food Classification within Commodity Groups. J. Am. Diet. Assoc. 1985, 85, 1306–1314.

28. Akbay, A.; Elhan, A.; Ozcan, C.; Demirtas, S. Hierarchical cluster analysis as an approach for systematic grouping of diet constituents on basis of fatty acid, energy and cholesterol content: Application on consumable lamb products. Med. Hypotheses 2000, 55, 147–154.

29. do Prado, S.B.R.; Giuntini, E.B.; Grande, F.; de Menezes, E.W. Techniques to evaluate changes in the nutritional profile of food products. J. Food Compost. Anal. 2016, 53, 1–6.

30. Li, Y.; Bahadur, R.; Ahuja, J.; Pehrsson, P.; Harnly, J. Macro-and micronutrients in raw plant foods: The similarities of foods and implication for dietary diversification. J. Food Compost. Anal. 2021, 102, 103993.

31. Atsa'am, D.D.; Oyelere, S.S.; Balogun, O.S.; Wario, R.; Blamah, N.V. K-means cluster analysis of the West African species of cereals based on nutritional value composition. Afr. J. Food Agric. Nutr. Dev. 2021, 21, 17195–17212.

32. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. Philos. Trans. A Math Phys. Eng. Sci. 2016, 374, 20150202.

33. Howard, M.C. A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? Int. J. Hum. Comput. Interact. 2016, 32, 51–62.

34. Balakrishna, Y.; Manda, S.; Mwambi, H.; van Graan, A. Identifying Nutrient Patterns in South African Foods to Support National Nutrition Guidelines and Policies. Nutrients 2021, 13, 3194.

35. Similä, M.; Ovaskainen, M.-L.; Virtanen, M.J.; Valsta, L.M. Nutrient content patterns of Finnish foods in a food composition database. J. Food Compost. Anal. 2006, 19, 217–224.

36. Westrich, B.J.; Altmann, M.A.; Potthoff, S.J. Minnesota's Nutrition Coordinating Center uses mathematical optimization to estimate food nutrient values. Interfaces 1998, 28, 86–99.

37. Ispirova, G.; Eftimov, T.; Korošec, P.; Seljak, B.K. MIGHT: Statistical methodology for missing-data imputation in food composition databases. Appl. Sci. 2019, 9, 4111.

38. Phanich, M.; Pholkul, P.; Phimoltares, S. Food Recommendation System Using Clustering Analysis for Diabetic Patients. In Proceedings of the 2010 International Conference on Information Science and Applications, Seoul, Korea, 21–23 April 2010; pp. 1–8.

39. Kim, S.; Sung, J.; Foo, M.; Jin, Y.S.; Kim, P.J. Uncovering the nutritional landscape of food. PLoS ONE 2015, 10, e0118697.