Statistical Sciences Extension
Vol. 2, No. 2, Spring & Summer 2017, (P 23-32)

گسترهٔ علوم آماری
سال دوم، شماره دوم، بهار و تابستان ۱۳۹۶، (ص ۲۳-۳۲)

ارزیابی ارزش تغذیه‌ای میوه‌ها و سبزیجات ایرانی بر اساس تجزیه و تحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای و جایگزینی آنها

معصومه حسین‌پور*[۱]، علیرضا فخارزاده جهرمی[۲]
۱. کارشناسی ارشد، ریاضی کاربردی، دانشگاه پیام نور شیراز
۲. استاد، ریاضی کاربردی، دانشگاه صنعتی شیراز و بنیاد نخبگان استان فارس

تاریخ دریافت: ۱۳۹۶/۰۳/۱۷ تاریخ پذیرش: ۱۳۹۷/۰۲/۲۳

The Nutrition Value Assessment and Replacement of Iranian Fruit and Vegetables Based on the Principal Component Analysis and Cluster Analysis

M. Hoseinpour*[1], A.R. Fakharzadeh Jahromi[2]
1. Masters Student, Applied Mathematics, Payame Noor University
2. Professor, Mathematics, Shiraz University of Technology & Fars Elite Foundation

Received: 2017/06/07 **Accepted:** 2018/05/13

چکیده
علی‌رغم توصیه مؤکد پزشکان و متخصصان تغذیه به مصرف میوه‌ها و سبزیجات، کمتر به چگونگی انتخاب میوه‌ها و سبزیجات در وعده‌های روزانه اشاره شده است. در این تحقیق، بر اساس رویکردی نوین، به طبقه‌بندی میوه‌ها و سبزیجات بر اساس ویژگی‌های مشابه آنها با توجه به میزان برخی مواد مغذی موجود در آنها پرداخته شده است. نظر به اینکه یکی از اهداف ارائهٔ رژیم غذایی، داشتن تنوع و تعادل تغذیه‌ای است، طبقه‌بندی میوه‌ها و سبزیجات بر اساس ارزش غذایی آنها، علاوه بر ایجاد تعادل تغذیه‌ای در مصرف این گروه مهم غذایی، یافتن جایگزین‌های مناسب را ممکن می‌سازد. بر این اساس برای برآورده شدن نیازهای تغذیه‌ای با استفاده از تجزیه و تحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای، میوه‌ها و سبزیجات در چندین گروه طبقه‌بندی می‌شوند؛ برای انجام این طبقه‌بندی از نرم‌افزار متلب استفاده خواهد شد.

واژگان کلیدی
تجزیه و تحلیل مؤلفه‌های اصلی، آنالیز خوشه‌ای، طبقه‌بندی، میوه و سبزیجات ایرانی، رژیم غذایی.

Abstract
Despite the physicians and nutritionists recommendations to consume the fruits and vegetables, they point much less on how to choose fruits and vegetables in daily meals. In this study, with a novel approach, fruits and vegetables have been classified on the basis of their similar characteristics and the amount of certain nutrients. One of the goals of diet is the diversity and balance of nutrition; therefore the classification of fruits and vegetables on the basis of their nutritional value makes balancing the nutritional intake of these important food groups and also finding a suitable replacement for them feasible. In order to meet the nutritional needs, using principal component analysis and cluster analysis, fruits and vegetables are classified into several groups. MATLAB software was used for doing this classification.

Keywords
Principal Component Analysis, Cluster Analysis, Classification, Iranian Fruit and Vegetable, Diet.

* **نویسنده مسئول:** معصومه حسین‌پور
*Corresponding Author: mah_h1979@yahoo.com

مقدمه

سیستم غذایی در سراسر جهان ناپایدار است و بنا بر شرایط آب و هوایی، اجتماعی، اقتصادی و بحران سیاسی یا مالی بسیار شکننده است. توجه به رژیم غذایی پایدار علاوه بر مقابله با این مشکلات، به کاهش واردات مواد غذایی، تولید و کشت هدفمند محصولات غذایی به‌صورت فصلی و محلی و پالایش و پردازش محصولات غذایی می‌انجامد. از طرفی سیستم‌های غذایی که از اواسط قرن بیستم توسعه پیدا کرده است به تولید گازهای گلخانه‌ای در حجم بالا و تغییرات اکوسیستم (مانند از دست دادن تنوع زیستی، جنگل‌زدایی، فرسایش خاک، آلودگی توسط مواد شیمیایی، کمبود آب) منجر شده است [۱]؛ در نتیجه تبیین راه‌کارهای جدید برای رژیم غذایی پایدار در کشورهای در حال توسعه، برای اطمینان از کیفیت و امنیت غذایی یک نیاز فوری و ضروری است؛ بخش مهمی از این راه‌کارها، توصیه به مصرف نکردن میوه و سبزیجات خارج از فصل است. مصرف میوه‌ها و سبزیجات فصلی نیازمند بررسی محتوای تغذیه‌ای آنها و طبقه‌بندی میوه‌ها و سبزیجات مشابه با هدف جایگزینی و معادل‌سازی آنها در فصول مختلف سال است. مصرف محصولات فصلی و محلی باعث بهبود طعم محصولات و کیفیت غذایی به برای برداشت محصول در اوج رشد و رسیدگی است. از سوی دیگر میوه‌ها و سبزیجات خارج از فصل با کاستی‌هایی همراه است [۲]؛ مصرف بیش از حد مجاز آفت‌کش‌ها در زمان کاشت، امکان مسمومیت مزمن و حاد را در بر خواهد داشت [۳]. مداخلهٔ مصنوعی در زمان کشت حتی با فرض بی‌خطر بودن و نگهداری در سردخانه‌ها به کاسته شدن ارزش غذایی و افزایش قیمت محصولات می‌انجامد [۴]. با توجه به طعم ضعیف‌تر و کم‌تر میوه‌ها و سبزیجات خارج فصل و در نتیجه کاهش محتوای تغذیه‌ای آن‌ها، بررسی محتوای تغذیه‌ای میوه‌ها و سبزیجات برای اطمینان از تعادل تغذیه‌ای افراد بسیار مهم است [۵]. مهم‌ترین اقدام برای ترویج مصرف میوه‌ها و سبزیجات محلی و فصلی، ارائهٔ جایگزین‌هایی از محصولات در تمام فصول سال است؛ برای تحقق این هدف باید میوه‌ها و سبزیجات بر طبق محتوای تغذیه‌ای مشابه طبقه‌بندی شوند. گرچه تاکنون شیوه‌ها و الگوهایی برای این امر ارائه شده است؛ اما با توجه به تفسیر نوع کشت‌ها و ذائقه‌ها این مطالعات کافی نیست؛ مثلاً دنگ و همکاران در سال ۲۰۱۵ خوشه‌بندی سلسله مراتبی متراکم شونده با استفاده از روش پیوند میانگین، ارزیابی مواد مغذی میوه‌ها و سبزیجات و طبقه‌بندی آنها را مطالعه کردند [۶]. آنان به جنبه‌های بومی و فصلی توجه ویژه‌ای داشتند. بر این اساس برای برآورده شدن نیازهای تغذیه‌ای و مصرف نکردن محصولات خارج فصل، در این تحقیق با استفاده از شیوه تجزیه‌وتحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای در علم آمار بنا بر ارزش‌گذاری‌های جدید تغذیه‌ای میوه‌ها و سبزیجات، این طبقه‌بندی انجام می‌شود. البته پیش از آن در سال ۲۰۰۸، الگوهای مختلف مصرف میوه و سبزی در فرانسه بررسی شد؛ خوشه‌بندی بر طبق ویژگی اجتماعی و میزان درآمد افراد توسط برتال و همکاران انجام پذیرفت [۷]. هم‌چنین طبقه‌بندی میوه‌ها و سبزیجات و تعیین سطح ترکیبات غذایی آنها بر اساس خانواده گیاهان، رنگ و میزان آنتی‌اکسیدان میوه‌ها و سبزیجات توسط پنینگتون و همکاران در سال ۲۰۰۹ انجام شد [۸]. پاتراس و همکاران در سال ۲۰۱۱، طبقه‌بندی بر اساس فعالیت آنتی‌اکسیدان‌های میوه‌ها و سبزیجات با استفاده از پارامترهای کیفیت آنها، نظیر رنگ و میزان رطوبت را مطالعه کردند [۹].

بر اساس آن‌چه که در این مقدمه ذکر شد، هدف اصلی این تحقیق ارائه یک دسته‌بندی از میوه‌ها و سبزیجات فصلی است برای مصرف رژیم غذایی به طوری که با سلیقه افراد نیز تطابق داشته باشد. همچنین این دسته‌بندی چنان باشد که بتواند برای تامین مواد مغذی مورد نیاز بدن هر شخص استفاده شود و از این لحاظ کمبودی از نظر تغذیه ایجاد نشود. بدین ترتیب متخصصین علم تغذیه ابزاری قوی و کاربردی را برای ارائه یک رژیم غذایی بومی و مطابق سلیقه افراد در دست خواهند داشت. برای انجام این امر از آنالیز خوشه‌ای استفاده خواهد شد.

تعاریف و مفاهیم آغازی

برای آشنایی بیشتر خوانندگان گرامی با اهداف این پژوهش ابتدا برخی از مفاهیم و تعاریف ارائه می‌شوند. این اطلاعات در قالب چهار تعریف در این بخش ارائه می‌شوند. دو مورد اول مربوط به معرفی مواد مغذی و دسته‌بندی آنها و دو مورد دیگر مرتبط با خوشه‌بندی است. ارزش تغذیه‌ای غذاها

یا همان مواد مغذی به دو شکل درشت مغذی‌ها و ریز مغذی‌ها به شرح زیر معرفی می‌شوند.

تعریف ۱: درشت مغذی‌ها شامل کربوهیدرات‌ها، پروتئین‌ها و چربی‌ها هستند که به‌عنوان منابع اصلی انرژی و مؤثر در ساختارهای بدنی با وظایف مهم از قبیل جایگزینی سلول‌ها و بافت‌های مرده، ساخت بافت‌های ماهیچه‌ای و نظایر آن به کار گرفته می‌شوند [۱۴].

تعریف۲: ریز مغذی‌ها شامل همهٔ ویتامین‌ها و مواد معدنی هستند. هر چند ریز مغذی‌ها تولید انرژی نمی‌کنند؛ اما برای کارکرد صحیح بدن ضروری هستند. سلول‌های بدن برای شکل‌گیری بر حسب نوع عملکردشان برای رشد، انجام عملیات حیاتی، ترمیم بافت‌هایی که پیوسته در حال فرسایش و بازسازی هستند و نیاز به مواد مغذی دارند و غذاهای مختلف این مواد را در اختیار سلول‌ها قرار می‌دهند. هر یک از مواد مغذی کار خود را انجام می‌دهد و حضور آنها برای سوخت و ساز بدن الزامی است. گرچه ریز مغذی‌ها مانند کلسیم، روی، آهن، ویتامین A و نظایر آن به میزان خوراکی‌ها و مواد غذایی شناخته شده نیستند؛ اما کمبود آنها در مقایسه با درشت مغذی‌ها ارزش و اهمیت کمتری ندارد؛ بلکه به دلیل تنوع‌شان، در صورت کمبود هر یک از آنها خسارت‌ها و آسیب‌های زیادی به افراد می‌رسد [۱۴].

تعریف ۳: خوشه‌بندی، فرآیند سازماندهی مجموعه‌ای از داده‌ها در تعدادی گروه است؛ به‌طوری که مشاهدات در یک گروه، شباهت زیادی به یکدیگر و نسبت به سایر گروه‌ها تفاوت‌های عمده‌ای دارند. با فرض این‌که داده‌ها ویژگی‌های مشخصی داشته باشند که بر طبق آن‌یک گروه را از گروه دیگر قابل تشخیص کند، نقطه شروع در این فرآیند انتخاب شیوه‌ای برای بیان ویژگی‌های مشاهدات است که توسط آن خوشه‌بندی انجام شود [۱۰].

تعریف ٤: ساده‌ترین و رایج ترین شیوهٔ خوشه‌بندی، روش خوشه‌بندی سلسله مراتبی است؛ فرآیند کار شامل دنباله‌ای از ادغام دو گروه در هر مرحله است. به این صورت که ابتدا N داده در N گروه در یک سمت خوشه‌بندی است و در پایان فرآیند یک گروه با N داده در انتهای دیگر طبقه‌بندی قرار دارد؛ روش خوشه‌بندی متراکم شونده، در اصل گروه‌بندی پایین به بالا است؛ چرا که در ابتدا N خوشه تک عضوی و در انتها یک خوشه N عضوی وجود دارد [۱۰]؛ شکل ۲ این روش را به‌خوبی نشان می‌دهد.

تجزیه و تحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای

نظریه کلی تجزیه و تحلیل مؤلفه‌های اصلی، بر اساس کاهش ابعاد مؤلفه‌های اصلی و ترکیب چند مؤلفهٔ اصلی به‌صورت مؤلفه‌هایی جدید با هدف حفظ اطلاعات اصلی مسئله است [۱۰]. روش ریاضی پردازش اطلاعات در اکثر مواقع به‌صورت ترکیب خطی p مؤلفهٔ اصلی و تبدیل آن به d مؤلفهٔ جدید است. به‌منظور اطمینان از انعکاس اطلاعات مؤلفه‌های اولیه در مؤلفه‌های جدید انتخاب یک روش استخراج مؤلفه‌های اصلی مناسب ضروری است. در این پژوهش در هر مرحله مؤلفه‌هایی که درصد تجمعی واریانس آنها بیش از ۸۵ درصد باشد، به‌عنوان مؤلفه‌های جدید برگزیده می‌شود؛ استفاده از این روش از حذف اطلاعات مفید مؤلفه‌های اولیه جلوگیری می‌کند. مؤلفه‌های جدید به دلیل کاهش تعداد متغیرها برای تجزیه و تحلیل و استفاده به‌عنوان داده‌های ورودی در دیگر روش‌های آماری، از جمله آنالیز خوشه‌ای ساده‌تر خواهد بود و نتایج دقیق‌تری را در برخواهد داشت. لازم به ذکر است انتخاب روش مناسب برای تجزیه‌وتحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای بنا به نوع تحقیقات است [۱۱]. در ادامه ترکیبی از روش کلی ارزیابی و تجزیه و تحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای برای طبقه‌بندی مناسب میوه‌ها و سبزیجات مشابه ارائه می‌شود. مؤلفه‌های اولیه در این تحقیق تعدادی از شاخص‌ها و محتوای تغذیه‌ای میوه‌ها و سبزیجات است که توسط روش ماتریس همبستگی در PCA[1]، کاهش ابعاد داده‌ها به‌منظور بالا بردن سرعت پردازش داده‌ها و افزایش دقت انجام می‌شود. پس از کاهش بعد، با استفاده از خوشه‌بندی سلسله مراتبی متراکم شونده، میوه‌ها و سبزیجات فصلی و بومی به گروه‌های مشابه تقسیم می‌شوند.

1. Principal Component Analysis

معرفی روش تجزیه‌وتحلیل مؤلفه‌های اصلی

فرض کنید n نوع میوه و سبزی که هر کدام حاوی P ماده مغذی است، به‌عنوان داده‌های اصلی انتخاب می‌شود. در این صورت هر میوه یا سبزی به‌عنوان یک نقطه با بعد P، در نظر گرفته و بردار x_i به شرح زیر برای میوه یا سبزی iام تعریف می‌شود:

$$x_i=\left(x_i(1),x_i(2),\ldots,x_i(p)\right).$$

در نتیجه مواد مغذی همهٔ میوه‌ها و سبزیجات را می‌توان توسط ماتریس $x=\left(x_i(j)\right)_{n\times p}$ بیان کرد. آنگاه ماتریس استاندارد داده‌های ورودی برای سادگی محاسبات توسط تفریق میانگین از مقدار داده‌های هر سطر نتیجه می‌شود؛ بنابراین ماتریس استاندارد داده‌های اولیه به شرح زیر خواهد بود:

$$z=\left(z_i(j)\right)_{n\times p}=\begin{pmatrix} z_1(1) & \cdots & z_1(p)\\ \vdots & \ddots & \vdots\\ z_n(1) & \cdots & z_n(p)\end{pmatrix}, \quad (۱)$$

جایی که $z_i=x_i-\bar{x}_i$ که $\bar{x}_i$ میانگین مؤلفه‌های سطر i ام است. محاسبهٔ ضریب همبستگی داده‌ها با استفاده از ماتریس استاندارد از رابطهٔ زیر حاصل می‌شود:

$$r_{ij}=\frac{\operatorname{cov}(z_i,z_j)}{\sqrt{\operatorname{var}(z_i).\operatorname{var}(z_j)}},$$

جایی که $\operatorname{cov}(z_i,z_j)$، کوواریانس استاندارد داده‌های x_i و x_j است و $\operatorname{var}(z_i)$ واریانس z_i است؛ بنابراین ماتریس ضرایب همبستگی را می‌توان به‌صورت زیر ارائه کرد:

$$R=\begin{bmatrix} r_{11} & \cdots & r_{1p}\\ \vdots & \ddots & \vdots\\ r_{n1} & \cdots & r_{np}\end{bmatrix}.$$

با توجه به متقارن بودن ماتریس R، می‌توان آن را به‌صورت ماتریس قطری Λ به‌صورت زیر نیز نمایش داد:

$$\Lambda=U^{T}RU=\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \lambda_n\end{bmatrix},$$

جایی که $U=[u_1,u_2,\ldots,u_n]$ یک ماتریس متعامد و شامل بردارهای ویژه متناظر با مقادیر ویژه $\lambda_1,\lambda_2,\ldots,\lambda_n$ از ماتریس R، ماتریس همبستگی داده‌های ورودی است؛ به‌طوری‌که $\lambda_1>\lambda_2>\ldots>\lambda_n$ [۱۰].

برای انتخاب تعداد مناسبی از مؤلفه‌های اصلی از روش درصد تجمعی واریانس استفاده می‌شود؛ بر این اساس رابطهٔ $a_i=\sum_{j=1}^{m}\lambda_j/\sum_{j=1}^{n}\lambda_j$ که در آن $i=1,2,\ldots,n$ به ازای $m<n$، برای انتخاب تعداد مؤلفه‌های اصلی به کار می‌رود [۱۰]. برای جلوگیری از حذف اطلاعات اصلی، مؤلفه‌هایی که درصد واریانس تجمعی آنها بیش از ۸۵ درصد باشد به‌عنوان مؤلفه‌های جدید انتخاب می‌شوند. مؤلفه‌هایی که طبق شرایط یاد شده انتخاب شده‌اند، به عنوان پارامترها و داده‌های ورودی روش آنالیز خوشه‌ای استفاده می‌شوند [۶].

مدل‌سازی آنالیز خوشه‌ای

در ابتدای این مدل‌سازی، هر میوه و سبزی به‌عنوان یک خوشه در نظر گرفته می‌شود. برای طبقه‌بندی n میوه و سبزی منتخب، باید معیاری برای تعیین فاصلهٔ بین آنها تعریف شود. در این بخش فاصلهٔ بین داده‌های iام و jام مسئله به‌صورت فاصله اقلیدسی در زیر تعریف می‌شود:

$$d_{ij}=d(z_i,z_j)=\sqrt{\left(z_i(1)-z_j(1)\right)^2+\left(z_i(2)-z_j(2)\right)^2+\ldots+\left(z_i(d)-z_j(d)\right)^2}.$$

سپس در ماتریس Y، فاصلهٔ اقلیدسی دودویی بین داده‌ها به‌صورت زیر ذخیره می‌شود:

$$Y=\left(d_{ij}\right)_{n\times d}=\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n}\\ d_{21} & \cdots & \cdots & d_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ d_{n1} & \cdots & \cdots & d_{nn}\\ & & & \end{pmatrix};$$

جدول ۱. مواد مغذی میوه‌ها و سبزیجات ایرانی.

شماره	نام میوه	GL	car	Ftd	Va	Vb6	Vc	Ve	S	Ca	Ir	Z	Se
۱	زردآلو	۴	۱۱/۱	۲	۱۹۲۶	۰/۵	۱۰	۰/۹	۱	۱۳	۰/۴	۰/۲	۰/۱
۲	هندوانه	۲	۷/۵	۰/۴	۵۶۹	۰	۸/۱	۰/۱	۱	۷	۰/۲	۰/۱	۰/۴
۳	شلیل	۲	۱۰/۶	۱/۷	۳۳۲	۰	۵/۴	۰/۸	۰	۶	۰/۳	۰/۲	۰
۴	توت فرنگی	۲	۷/۷	۲	۱۲	۰	۵۸/۸	۰/۳	۱	۱۶	۰/۴	۰/۱	۰/۴
۵	سیب	۳	۱۳/۸	۲/۴	۵۴	۰	۴/۶	۰/۲	۱	۶	۰/۱	۰	۰
۶	آلبالو	۴	۱۲/۲	۱/۶	۶۴	۰	۱۰	۰/۱	۳	۱۶	۰/۳	۰/۱	۰/۱
۷	گلابی	۲	۱۰/۶	۳/۶	۰	۰	۳/۸	۰/۱	۰	۴	۰	۰	۰/۱
۸	گیلاس	۴	۱۶	۲/۱	۱۲۸۳	۰	۷	۰/۱	۰	۱۳	۰/۴	۰/۱	۰
۹	انگور	۶	۱۸/۱	۰/۹	۶۶	۰/۱	۱۰/۸	۰/۲	۲	۱۰	۰/۴	۰	۰/۱
۱۰	نارنگی	۴	۱۳/۳	۱/۸	۶۸۱	۰/۱	۲۶/۷	۰/۲	۲	۳۷	۰/۲	۰/۱	۰/۱
۱۱	پرتقال	۳	۱۱/۵	۲/۴	۲۲۵	۰/۱	۴۵	۰/۲	۰	۴۳	۰/۱	۰/۱	۰/۵
۱۲	لیمو ترش	۳	۱۰/۷	۴/۷	۳۰	۰/۱	۷۷	۰	۳	۶۱	۰/۷	۰/۱	۰
۱۳	انار	۶	۱۸/۷	۴	۰	۰/۱	۱۰/۲	۰/۶	۳	۱۰	۰/۳	۰/۴	۰/۵
۱۴	خرمالو	۵	۱۸/۶	۳/۶	۱۶۲۷	۰/۱	۷/۵	۰/۷	۱	۸	۰/۲	۰/۱	۰/۶
۱۵	خرما	۳۹	۷۵	۸	۱۰	۰/۲	۰/۴	۰/۱	۲	۳۹	۱	۰/۳	۳
16	هلو	۳	۹۹	۱۵	۳۲۶	۰	۶/۶	۰/۷	۰	۶	۰/۳	۰/۲	۰/۱
17	به	۴	۱۵/۳	۱/۹	۴۰	۰	۱۵	۰	۴	۱۱	۰/۷	۰	۰/۶
18	خیار	۱	۳/۶	۰/۵	۱۰۵	۰	۲/۸	۰	۲	۱۶	۰/۳	۰/۲	۰/۳
19	آلو	۳	۱۱/۴	۱/۴	۳۴۵	۰	۹/۵	۰/۳	۰	۶	۰/۲	۰/۱	۰
20	کاهو	۱	۲/۸	۱/۳	۷۴۰۴	۰/۱	۱۸	۰/۳	۲۸	۳۶	۰/۹	۰/۲	۰/۶
21	هویج	۳	۹/۶	۲/۸	۱۶۷۰۵	۰/۱	۵/۹	۰/۷	۶۹	۳۳	۰/۳	۰/۲	۰/۱
22	زیتون	۱	۳/۸	۳/۳	۳۹۳	۰	۰	۳/۸	۱۵۵۶	۵۲	۰/۵	۰	۰/۹
23	پیاز	۳	۹/۳	۱/۷	۲	۰/۱	۷/۴	۰	۴	۲۳	۰/۲	۰/۲	۰/۵
24	بادمجان	۱	۵/۷	۳/۴	۲۷	۰/۱	۲/۲	۰/۳	۲	۹	۰/۲	۰/۲	۰/۳
25	سیب زمینی	۸	۱۸/۴	۲/۲	۲	۰/۲	۱۹/۷	۰	۶	۱۲	۰/۸	۰/۳	۰/۳
26	فلفل دلمه‌ای	۲	۶/۳	۲/۱	۳۱۳۱	۰/۳	۱۲۸	۱/۶	۴	۷	۰/۴	۰/۳	۰/۱
شماره	**نام میوه**	**GL**	**car**	**Ftd**	**Va**	**Vb6**	**Vc**	**Ve**	**S**	**Ca**	**Ir**	**Z**	**Se**
27	بامیه	۳	۷	۳/۲	۳۷۵	۰/۲	۲۱/۱	۰/۴	۸	۸۱	۰/۸	۰/۶	۰/۷
28	گوجه‌فرنگی	۲	۳/۲	۰/۹	۱۴۹۶	۰/۱	۱۶	۰	۴۲	۵	۰/۵	۰/۱	۰/۴
29	شلغم	۲	۶/۴	۱/۸	۰	۰/۱	۲۱	۰	۶۷	۳۰	۰/۳	۰/۳	۰/۷
30	کلم قمری	۲	۶/۲	۳/۶	۳۶	۰/۲	۶۲	۰/۵	۲۰	۲۴	۰/۴	۰	۰/۷
31	نعناع	۲	۸/۴	۶/۸	۴۰۵۴	۰/۲	۱۳/۳	۰	۳۰	۱۹۹	۱۱/۹	۱/۱	۰
32	ریحان	۱	۲/۷	۱/۶	۵۲۷۶	۰/۲	۱۸	۰/۸	۴	۱۷۷	۳/۲	۰/۸	۰/۳
33	تربچه	۱	۳/۵	۱/۶	۷	۰/۱	۱۴/۸	۰	۳۹	۲۵	۰/۳	۰/۳	۰/۶
34	جعفری	۳	۶/۳	۳/۳	۸۴۲۵	۰/۱	۱۳۳	۰/۷	۵۶	۱۳۸	۶/۲	۱/۱	۰/۱
35	پیازچه	۳	۷/۳	۲/۶	۹۹۷	۰/۱	۱۸/۸	۰/۵	۱۶	۷۲	۱/۵	۰/۴	۰/۶
36	اسفناج	۱	۳/۶	۲/۲	۹۳۷۶	۰/۲	۲۸/۱	۲۲	۷۹	۹۹	۲/۷	۰/۵	۱
37	لوبیا سبز	۳	۷/۱	۳/۴	۶۹۰	۰/۱	۱۶/۳	۰/۴	۶	۳۷	۱	۰/۲	۰/۶
38	باقلا	۸	۱۷/۶	۷/۵	۳۳۳	۰/۱	۳/۷	۱/۲	۲۵	۳۷	۱/۵	۱	۰/۸
39	گل کلم	۲	۵/۳	۲/۵	۱۳	۰/۲	۴۶/۴	۰/۱	۳۰	۲۲	۰/۴	۰/۳	۰/۶
40	سیر	۱۶	۳۳/۱	۲/۱	۹	۱/۲	۳۱/۲	۰/۱	۱۷	۱۸۱	۱/۷	۱/۲	۱۴/۲
41	کرفس	۱	۳/۴	۱/۶	۴۴۹	۰/۱	۳/۱	۰/۳	۸۰	۴۰	۰/۲	۰/۱	۰/۴
42	تره فرنگی	۵	۱۴/۲	۱/۸	۱۶۶۷	۰/۲	۱۲	۰/۹	۲۰	۵۹	۲/۱	۰/۱	۱
43	ازگیل	۳	۱۲/۱	۱/۷	۱۵۲۸	۰/۱	۱	۰	۱	۱۶	۰/۳	۰/۱	۱
44	ذرت	۷	۱۹	۲/۷	۱	۰/۱	۶/۸	۰/۱	۱۵	۲	۰/۵	۰/۵	۰/۶
45	شوید	۳	۷	۲/۱	۷۷۱۷	۰/۲	۸۵	۰	۶۱	۲۰۸	۶/۶	۰/۹	۰
46	شقاقل	۵	۱۸	۴/۹	۰	۰/۱	۱۷	۱/۵	۱۰	۳۶	۰/۶	۰/۶	۱/۸
47	موز	۸	۲۲/۸	۲/۶	۶۴	۰/۴	۸/۷	۰/۱	۱	۵	۰/۳	۰/۲	۱
48	انجیر تازه	۶	۱۹/۲	۲/۹	۱۴۲	۰/۱	۲	۰/۱	۱	۳۵	۰/۴	۰/۲	۰/۲
49	انجیرخشک	۳۰	۶۳/۹	۹/۸	۱۰	۰/۱	۱/۲	۰/۴	۱۰	۱۶۲	۲	۰/۵	۰/۶
50	کیوی	۴	۱۴/۷	۳	۸۷	۰/۱	۹۲/۷	۱/۵	۳	۳۴	۰/۳	۰/۱	۰/۲
51	گردو	۰	۹/۹	۶/۸	۴۰	۰/۶	۱/۷	۱/۸	۲	۶۱	۳/۱	۳/۴	۱۷
52	کدو سبز	۲	۳/۳	۱/۱	۲۰۰	۰/۲	۱۷	۰/۱	۱۰	۱۵	۰/۴	۰/۳	۰/۲
53	ریواس	۱	۴/۵	۱/۸	۱۰۲	۰	۸	۰/۳	۴	۸۶	۰/۲	۰/۱	۱/۱
54	انبه	۵	۱۷	۱/۸	۷۶۵	۰/۱	۲۷/۷	۱/۱	۲	۱۰	۰/۱	۰	۰/۶
55	نخودفرنگی	۵	۱۴/۵	۵/۱	۷۶۵	۰/۲	۴۰	۰/۱	۵	۲۵	۱/۵	۱/۲	۱/۸

لازم به یادآوری است که Y ماتریسی متقارن با عناصر صفر در قطر اصلی است. در مرحلهٔ بعد، فاصلهٔ بین خوشه‌های G_p و G_q از رابطهٔ زیر تعیین می‌شود:

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{i\in G_p,\, j\in G_q} d_{ij}^2, \qquad p \neq q,$$

حال فرض کنید که با ادغام خوشهٔ G_p و G_q، خوشهٔ جدیدی با نام G_r حاصل شود. بر این اساس در مرحلهٔ بعد محاسبه D_{rk}^2 به‌صورت زیر بسط داده می‌شود:

$$\begin{aligned} D_{rk}^2 &= \frac{1}{n_r n_k} \sum_{i\in G_r,\, j\in G_k} d_{ij}^2 \\ &= \frac{1}{n_r n_k} \left(\sum_{i\in G_p,\, j\in G_k} d_{ij}^2 + \sum_{i\in G_q,\, j\in G_k} d_{ij}^2 \right) \\ &= \frac{n_p}{n_r} D_{pk}^2 + \frac{n_q}{n_r} D_{qk}^2. \end{aligned}$$

محاسبه فاصلهٔ بین داده‌ها به‌صورت بالا، روش پیوند میانگین است؛ در این رابطه n_r و n_k تعداد اعضای G_r و G_k هستند [۶].

طبقه‌بندی میوه‌ها و سبزیجات ایرانی

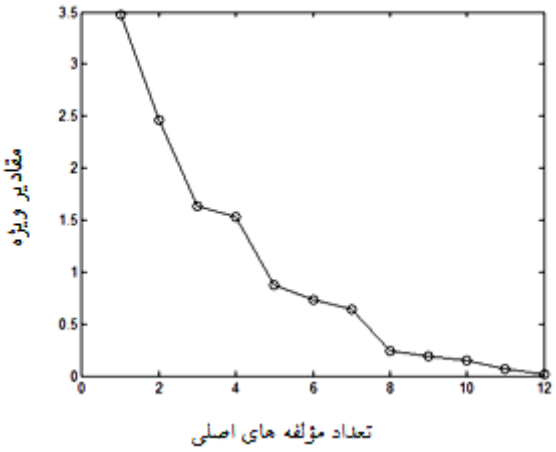
کاهش خطر ابتلا به بسیاری از بیماری‌ها با مصرف میوه و سبزی در رنگ‌ها، شکل‌ها و طعم‌های گوناگون امکان‌پذیر است. در واقع میوه و سبزی تازه در رژیم غذایی برای پیشگیری از بیماری نقش مهمی را ایفا می‌کنند. هر چه رنگ محتویات بشقاب میوه و سبزیجات بیشتر باشد، از نظر ویتامین‌ها و مواد معدنی غنی‌تر خواهد بود؛ به همین دلیل میوه‌ها و سبزیجات ایرانی در این بخش طبقه‌بندی می‌شوند و با استفاده از نتیجهٔ طبقه‌بندی میوه‌ها و سبزیجات از مصرف بیش از حد سبزی و میوهٔ مشابه در رژیم غذایی جلوگیری می‌شود.

می‌توان با تجزیه‌وتحلیل مؤلفه‌های اصلی و آنالیز خوشه‌ای، میوه‌ها و سبزیجات را در چندین گروه طبقه‌بندی کرد؛ بر این اساس امکان مصرف میوه‌ها و سبزیجات مشابه در فصل‌های مختلف سال فراهم می‌شود. مزیت این کار در آن است که حتی در صورت بی‌علاقگی بیمار یا حساسیت وی به نوع خاصی از سبزی، می‌توان میوه یا سبزی فصلی با خواص مشابه را جایگزین آن کرد. شاخص‌های مربوط به میوه‌ها و سبزیجات به‌عنوان مؤلفه‌های اصلی در نظر گرفته می‌شوند که مؤلفه ۱ تا مؤلفه ۱۲ به ترتیب فیبر، کربوهیدرات، ویتامین آ، ویتامین ب، ویتامین ث، ویتامین ای، سدیم، کلسیم، آهن، روی، سلنیوم و بارگلیسمی است. لازم به ذکر است که بارگلیسمی، از جمله عوامل افزایش قند خون است و به مقدار کربوهیدرات، نوع قند، نوع نشاسته، نحوهٔ پخت و فرآیند غذاها بستگی دارد. هر چه بارگلیسمی غذاها بیشتر باشد تأثیر آن بر افزایش قند خون بیشتر خواهد بود [۱۲]. میزان شاخص‌های ذکر شده مربوط به میوه‌ها و سبزیجات انتخابی به میزان صد گرم برای هر میوه یا سبزی با استفاده از [۱۳] محاسبه شده است و مواد مغذی مربوط به هر یک از میوه‌ها و سبزیجات در جدول ۱ ارائه شده است؛ در این پژوهش، ۵۵ نوع میوه و سبزی ایرانی با توجه به در دسترس بودن و مصرف رایج آنها برای طبقه‌بندی در نظر گرفته شده است. به‌منظور درک بهتر شکل ۲، توجه به ترتیب شماره‌گذاری آنها در جدول ۱ مفید است؛ بنابراین داده‌های اولیه با توجه به ۵۵ نوع میوه و سبزی و ۱۲ مؤلفه مرتبط با شاخص‌های تغذیه‌ای آنها به‌صورت $X=\left(x_{ij}\right)_{55\times 12}$ بیان می‌شود.

مراحل کاهش بعد مؤلفه‌های اصلی به این صورت است که ابتدا با استفاده از رابطه (۱)، ماتریس استاندارد داده‌های اصلی نوشته می‌شود؛ سپس ضریب همبستگی ماتریس داده‌های استاندارد محاسبه می‌گردد. در مرحلهٔ بعد پس از محاسبه مقادیر ویژهٔ ماتریس ضریب همبستگی و مرتب‌سازی آنها به‌صورت نزولی، مؤلفه‌هایی که درصد تجمعی واریانس آنها بیش از ۸۵ درصد باشد به‌عنوان مؤلفه‌های اصلی جدید در نظر گرفته می‌شوند. نمودار مقادیر ویژه ماتریس ضریب همبستگی در زیر ارائه می‌شود.

در جدول ۲، مقادیر ویژه، نسبت درصد واریانس‌ها و نسبت درصد واریانس تجمعی مربوط به مقادیر ویژه ماتریس ضریب همبستگی مربوط به شاخص‌های میوه‌ها و سبزیجات آمده است. طبق رابطهٔ $a_j = \lambda_j / \sum_1^n \lambda_j$ نسبت درصد واریانس و با استفاده از رابطه $a_j = \sum_1^m \lambda_j / \sum_1^n \lambda_j$ به‌طوری‌که $m < n$، نسبت درصد واریانس تجمعی محاسبه شده است؛ از این مفاهیم برای انتخاب تعداد مؤلفه‌های اصلی استفاده می‌شود.

در این بخش برای انتخاب مؤلفه‌های جدید در روش‌های کاهش بعد، روش درصد تجمعی واریانس به کار گرفته شده است. طبق جدول ۲، هفت مؤلفه از مؤلفه‌های اصلی دارای درصد واریانس تجمعی بیش از ۸۵ درصد هستند؛ این مؤلفه‌ها به‌عنوان مؤلفه‌های اصلی جدید در نظر گرفته می‌شوند. همان‌گونه که توضیح داده شد، از روش تجزیه و تحلیل مؤلفه‌های اصلی برای کاهش بعد، سادگی



شکل ۱. نمودار مقادیر ویژه ماتریس ضریب همبستگی مربوط به میوه‌ها و سبزیجات ایرانی

در محاسبات آینده و تضمین از دست رفتن اطلاعات کم‌تر استفاده می‌شود؛ بنابراین پس از کاهش ابعاد داده‌های اولیه مربوط به میوه‌ها و سبزیجات، از روش آنالیز خوشه‌ای برای طبقه‌بندی آنان بر اساس مؤلفه‌های اصلی جدید، استفاده می‌شود. هر یک از میوه‌ها و سبزیجات به‌عنوان یک خوشه در نظر گرفته می‌شوند؛ فاصلهٔ بین هر خوشه با استفاده از روش پیوند میانگین محاسبه می‌شود [۱۰] و خوشه‌های با فاصلهٔ کمتر در هم ادغام می‌شوند. بر اساس فصل برداشت هر یک از این میوه‌ها و سبزیجات می‌توان نتایج گروه‌بندی را به‌صورت معناداری بیان کرد؛ پس از آن میوه‌ها و سبزیجات هر گروه بنا به فصل برداشت محصول در جدول خوشه‌ها وارد شده است تا امکان یافتن میوه و سبزی جایگزین با خواص مشابه به‌راحتی فراهم شود. نتایج این خوشه‌بندی در جدول ۳ بیان شده است. لازم به یادآوری

جدول ۲: مقادیر ویژه و درصد واریانس تجمعی میوه‌ها و سبزیجات ایرانی.

مؤلفه‌ها	مقادیر ویژه	درصد واریانس مقادیر ویژه	درصد واریانس تجمعی
مؤلفه ۱	۳/۴۶۹۹	۲۸/۹۲	۲۸/۹۱۶۱
مؤلفه ۲	۲/۴۶۰۵	۲۰/۵۰	۴۹/۴۲۰۵
مؤلفه ۳	۱/۶۳۳۹	۱۳/۶۲	۶۳/۰۳۶۲
مؤلفه ۴	۱/۵۲۶۹	۱۲/۷۲	۷۵/۷۶۰۴
مؤلفه ۵	۰/۸۷۴۴	۷/۲۹	۸۳/۰۴۷۴
مؤلفه ۶	۰/۷۳۶۷	۶/۱۴	۸۹/۱۸۶۷
مؤلفه ۷	۰/۶۳۷۲	۵/۳۱	۹۴/۴۹۶۷
مؤلفه ۸	۰/۲۳۵۷	۱/۹۶	۹۶/۴۶۰۷
مؤلفه ۹	۰/۱۸۹۴	۱/۵۸	۹۸/۰۳۹۱
مؤلفه ۱۰	۰/۱۵۲۸	۱/۲۷	۹۹/۳۱۲۱
مؤلفه ۱۱	۰/۰۶۷۳	۰/۵۶	۹۹/۸۷۳۲
مؤلفه ۱۲	۰/۰۱۵۲	۰/۱۳	۱۰۰

است که فصل هر یک از میوه‌ها و سبزیجات طبق زمان برداشت آنها در استان فارس در نظر گرفته شده است؛ هر چند منبع معتبر و قابل استناد برای تعیین دقیق فصل برداشت همهٔ محصولات وجود ندارد.

برای طبقه‌بندی میوه‌ها و سبزیجات، ۱۰ خوشه به‌طور تجربی در نظر گرفته شده است. روش دیگر برای تعیین تعداد خوشه‌ها، استفاده از نمودار دندوگرام است [۱۰]. در هر مرحله از فرآیند آنالیز خوشه‌ای، خوشه‌هایی که فاصله کم‌تری دارند در هم ادغام می‌شوند. محاسبه فاصلهٔ بین خوشه‌ها با استفاده از نرم‌افزار متلب $(MATLAB\ R2013a)$ و به روش پیوند میانگین صورت گرفته است؛ برای درک بهتر خوشه‌بندی نمودار دندوگرام در شکل ۲، ارائه‌شده است؛ با استفاده از این نمودار، مراحل طبقه‌بندی نشان داده می‌شود. با رسم خطی موازی محور مربوط به نام میوه‌ها، در فاصلهٔ حدود ۳/۲، ده خوشه ایجاد می‌شود که کاملاً مطابق مؤلفه‌های اصلی جدید در نظر گرفته می‌شوند.

جدول ۳. طبقه‌بندی میوه‌ها و سبزیجات ایرانی با توجه به فصول برداشت.

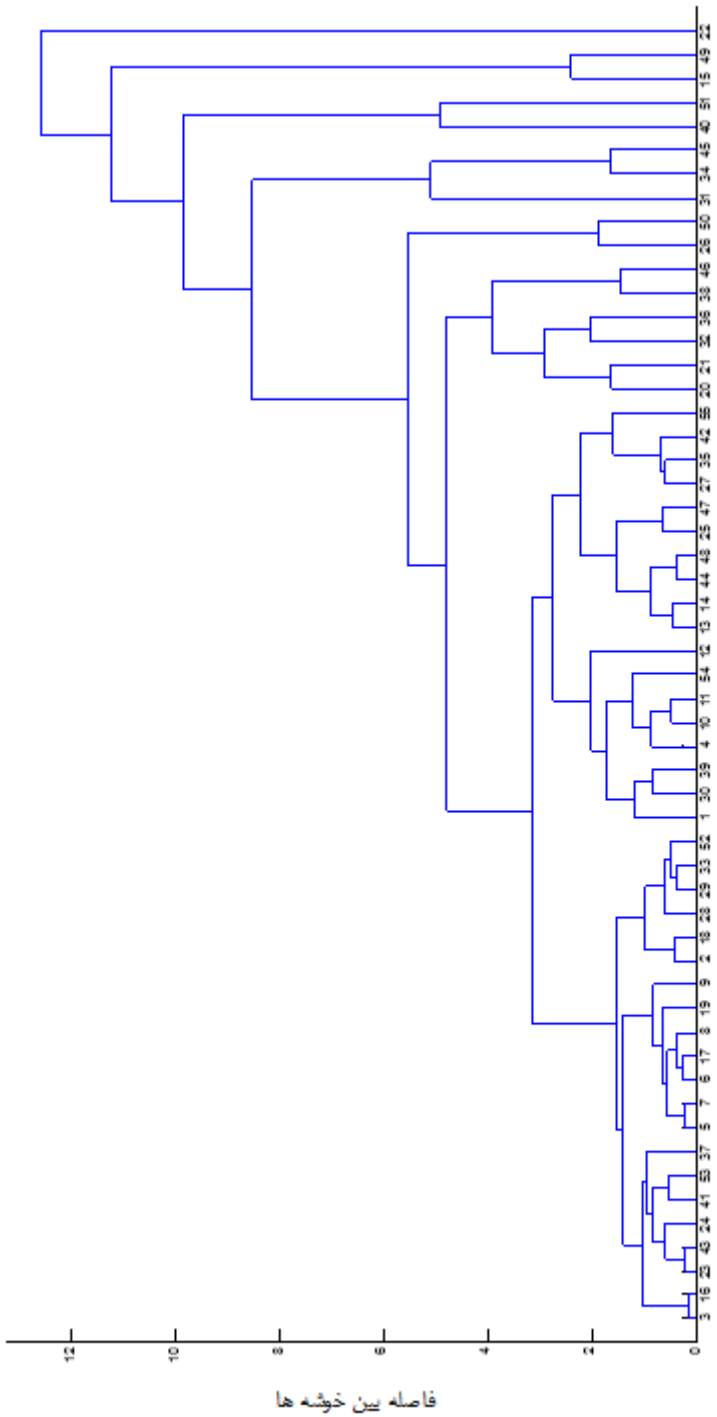
گروه‌ها	فروردین	اردیبهشت	خرداد	تیر	مرداد	شهریور	مهر	آبان	آذر	دی	بهمن	اسفند
گروه ۱	پرتقال پیاز ریواس موز	توت‌فرنگی پیاز ازگیل پیازچه انجیر تازه موز	آلو توت‌فرنگی پیاز ازگیل بادمجان پیازچه سیب‌زمینی انجیر تازه انبه موز	زردآلو آلبالو هلو شلیل آلو خیار هندوانه گوجه فرنگی گیلاس بادمجان لوبیا سبز سیب‌زمینی ذرت انبه کدو سبز	زردآلو آلبالو هلو شلیل آلو انگور خیار هندوانه گوجه فرنگی گیلاس لیمو ترش بادمجان لوبیا سبز سیب زمینی ذرت انبه کدو سبز	هلو شلیل انگور خیار هندوانه گوجه فرنگی سیب گلابی انار لیمو ترش بادمجان بامیه ذرت کدو سبز	خیار هندوانه تربچه گوجه فرنگی سیب گلابی به انار لیمو ترش بامیه خرمالو موز	پرتقال شلغم تربچه انار لیمو ترش بامیه خرمالو موز	پرتقال شلغم تربچه گل کلم تره خرمالو کلم قمری موز	پرتقال نارنگی شلغم تربچه گل کلم تره کلم قمری	پرتقال نارنگی شلغم کرفس کلم قمری	پرتقال نارنگی شلغم کرفس
گروه ۲						گردو						
گروه ۳			نعناع	نعناع		نعناع			تعناع			
گروه ۴								جعفری	جعفری شوید	شوید		
گروه ۵				ریحان		ریحان	ریحان هویج کاهو	اسفناج هویج کاهو	اسفناج هویج کاهو	اسفناج هویج کاهو	هویج کاهو	هویج کاهو
گروه ۶	نخودفرنگی	باقلا نخودفرنگی	باقلا	باقلا				شقاقل	شقاقل	شقاقل		
گروه ۷						فلفل دلمه‌ای	فلفل دلمه‌ای	فلفل دلمه‌ای کیوی	کیوی			
گروه ۸			خرما	خرما	خرما انجیرخشک	خرما انجیرخشک	خرما انجیرخشک	خرما				
گروه ۹					زیتون	زیتون	زیتون	زیتون				
گروه ۱۰			سیر	سیر	سیر							

با خوشه‌بندی ذکر شده در جدول ۳ است. البته اگر خط موازی در فاصلهٔ کمتری از محور مربوط به فهرست میوه‌ها و سبزیجات رسم شود، تعداد بیشتری خوشه نتیجه می‌شود. با استفاده از نتایج به‌دست‌آمده در این بخش علاوه بر شناسایی میوه‌ها و سبزیجات مشابه، امکان جایگزینی آنها به دلیل بی‌علاقگی فرد یا حساسیت وی به نوع خاصی از میوه و سبزیجات یا به دلیل مصرف نکردن میوه‌های غیر فصلی، فراهم می‌شود. به‌عنوان مثال با توجه به سبزیجات مشابه متعلق به گروه ۵ در جدول ۳، در صورت بی‌علاقگی فرد به اسفناج، کاهو گزینه مناسبی برای جایگزینی آن است.

نتیجه‌گیری و بحث

استفاده از نتایج این خوشه‌بندی می‌تواند فواید زیادی داشته باشد که از جمله فواید آن انتخاب میوه‌ها و سبزیجات فصلی مناسب برای تأمین مواد مغذی مورد نیاز بدن هر شخص به‌جای مصرف میوه‌ها و سبزیجات خارج فصل و گران‌تر است. وجود جایگزین برای سبزی یا میوه می‌تواند مشکل بی‌علاقگی کودکان یا حتی بزرگ‌سالان نسبت به نوع خاص سبزی را رفع کند. ترویج مصرف میوه‌ها و سبزیجات فصلی، علاوه بر موارد ذکر شده، فواید دیگری نیز دارد؛ مثلاً بر این اساس، از مصرف بیش از حد مجاز سموم شیمیایی در مرحلهٔ کاشت محصولات خارج فصل جلوگیری می‌شود. در صورت کاشت محصولات به‌صورت فصلی و محلی، با کاهش فاصلهٔ حمل‌ونقل محصولات کشاورزی از محل تولید به محل فروش، ضمن حمایت از کشاورزان و باغداران محلی، تجارت عادلانه دور از ذهن نخواهد بود؛ کاهش آلودگی هوا و کاهش مصرف انرژی از پیامدهای حاصل از کشت محلی و فصلی است.

حفظ محیط زیست و تنوع زیستی گیاهان مختلف، نتیجهٔ تنوع کشت گونه‌های مختلف گیاهی در هر منطقه است.



شکل ۲: نمودار دندوگرام میوه ها و سبزیجات ایرانی.

در این راستا محافظت از خاک، جنگل، آب و به حداقل رساندن آلودگی‌های شیمیایی و حفظ اکوسیستم در دراز مدت از نتایج دست‌یافتنی خواهد بود. ترکیب دانش سنتی و مهارت‌ها و ابزارهای جدید کشاورزی در تولید محصولات محلی و فصلی، به بهبود تغذیه و حفاظت از تنوع زیستی، بهینه‌سازی طعم و کیفیت غذایی و جلب سلیقهٔ افراد خواهد انجامید.

References

[1] Lairon D. Biodiversity and sustainable nutrition with a food-based approach. SUSTAINABLE DIETS AND BIODIVERSITY. 2012.

[2] Fan ZH. Anti-Season Vegetable and Fruit Nutrition Discount. Cancer Frontier. 2010;6:51- 2.

[3] Tan DM. Anti-Season Vegetable and Fruit Problems. Food and Health. 2001; 11:17.

[4] Lv B. How to Eat to Be Safe. Environment. 2011;4:64- 7.

[5] Yang B. Anti-Season Vegetable and Fruit Pay Attention to Eat. Health and Beauty. 2010;5:220- 3.

[6] Deng G, Liang X, Yan B. Fruit and Vegetable Nutrition Value Assessment and Replacement Based on the Principal Component Analysis and Cluster Analysis. Applied Mathematics. 2015; 6: 1620- 9.

[7] Bertail P, Caillavet F. Fruit and vegetable consumption patterns: a segmentation approach. American Journal of Agricultural Economics. 2008;90(3):827-42.

[8] Pennington JA, Fisher RA. Classification of fruits and vegetables. Journal of Food Composition and Analysis. 2009; 22: S23-S31.

[9] Patras A, Brunton NP, Downey G, Rawson A, Warriner K, Gernigon G. Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity. Journal of Food Composition and Analysis. 2011; 24(2): 250-6.

[10] Martinez WL, Martinez AR, Martinez A, Solka J. Exploratory data analysis with MATLAB: CRC Press; 2010.

[11] Chen XJ. Clustering Analysis in Data Mining Research. Computer Technology and Development. 2006; 9:44.

[12] Bas E. A robust optimization approach to diet problem with overall glycemic load as objective function. Applied Mathematical Modelling. 2014; 38 (19): 4926-40.

[13] Nast C. Nutrition Facts http://nut-ritiondata.self.com/facts/snacks/5374/2. Accessed; 2013.

[14] Sullivan J. R. Digestion and nutrition (The human body: how it works): Chelsea House: 2009.