

## تحلیل خوشه‌ای توزیع امکانات بهداشتی و درمانی استان‌های ایران

مریم میرزایی<sup>۱\*</sup>، تابان باغفلکی<sup>۲</sup> و منیره معنوی<sup>۳</sup>

<sup>۱</sup> پژوهشکده آمار، مرکز آمار ایران، تهران، ایران

<sup>۲</sup> دانشکده علوم ریاضی، دانشگاه تربیت مدرس، تهران، ایران

<sup>۳</sup> مرکز تحقیقات عوامل اجتماعی مؤثر بر سلامت، دانشگاه علوم پزشکی سمنان، سمنان، ایران

mmaryammirzaie73@gmail.com

**چکیده:** از آنجایی که بهداشت و درمان نقش بسیار مؤثری در برخورداری از رفاه اجتماعی ایفا می‌کند، یکی از اهداف اساسی سیاست‌گذاران باید ایجاد تسهیل در دسترسی افراد به خدمات بهداشتی و درمانی در جامعه باشد. از این رو نحوه توزیع امکانات بهداشتی و درمانی از فاکتورهای بسیار مهم توسعه یک کشور محسوب می‌شود. در همین راستا در این مقاله سعی شده است در رابطه با نحوه توزیع این امکانات در سراسر کشور بحث شود و استان‌هایی که از لحاظ میزان برخورداری از امکانات بهداشتی و درمانی مشابه هستند، شناسایی شوند. همچنین در رابطه با استان‌هایی که بیش‌ترین، متوسط و کمترین برخورداری از این امکانات را دارا هستند نیز بحث می‌شود. برای این منظور از روش  $k$ -میانگین و روش‌های صوری صورت چرنف و خوشه‌بندی سلسله مراتبی استفاده شده است. با به کارگیری این روش‌ها استان تهران با اختلاف معناداری بیش‌ترین امکانات بهداشتی و درمانی را دارد. استان‌های خراسان رضوی، اصفهان، فارس، آذربایجان شرقی، خوزستان و مازندران بعد از تهران بیش‌ترین امکانات بهداشتی و درمانی را دارند و به لحاظ برخورداری از امکانات ذکر شده مشابه هستند.

**کلید واژه‌ها:** امکانات بهداشتی و درمانی، خوشه‌بندی، سلسله‌مراتبی، صورتک چرنف،  $k$ -میانگین.

### ۱- مقدمه

یکی از مهم‌ترین تکنیک‌های داده‌کاوی خوشه‌بندی است. اولین بار ایده‌ی آن در دهه ۱۹۳۵ میلادی (۱۳۱۴ خورشیدی) ارائه شد که امروزه با پیشرفت‌ها و جهش‌های عظیمی که در آن پدید آمده است در کاربردها و جنبه‌های مختلفی حضور یافته است [1]. روش‌های خوشه‌بندی مختلفی مانند خوشه‌بندی  $k$ -میانگین،  $k$ -نماینده، سلسله‌مراتبی، فازی، زیرفضا و ... ارائه شده‌اند که هر کدام دارای نقاط ضعف و قوت مختص به خود هستند. به دلیل آنکه در خوشه‌بندی اطلاعات مربوط به برجسب کلاس وجود ندارند، به این فرایند یادگیری بی‌ناظر یا بدون ناظر نیز گفته می‌شود [2]. الگوریتم  $k$ -میانگین یکی از مشهورترین الگوریتم‌های خوشه‌بندی داده‌ها می‌باشد که به دلیل پیاده‌سازی آسان و سرعت عملکرد، محبوبیت زیادی یافته است [3]. خوشه‌بندی در زمینه‌های متنوعی مانند مهندسی (بینش محاسباتی، یادگیری ماشین، تشخیص الگو، مهندسی مکانیکی، مهندسی برق) [4]، علوم کامپیوتر (وب کاوی، دسته‌بندی اسناد و یا دسته‌بندی مشتریان به سایت‌ها، تحلیل پایگاه داده فضایی، بازیابی اطلاعات، گردآوری مستندات متنی و پردازش تصویر، تقسیم‌بندی تصاویر پزشکی و یا ماهواره‌ای) [5]، علوم پزشکی (ژنتیک، زیست شناسی، میکروب شناسی، روان پزشکی و داده‌های بیان ژن کاربرد دارند [6]، [۱] انواع الگوریتم‌های  $k$ -میانگین و پیشرفت‌های اخیر که در این حوزه صورت گرفته است را بررسی کرده‌اند و اثربخشی این الگوریتم‌ها را بر اساس تجزیه و تحلیل تجربی روی مجموعه‌های داده مختلف مورد بررسی قرار داده‌اند. [۷] با مقایسه بین دو الگوریتم  $k$ -میانگین و CLARA روی مجموعه داده گل زنبق، دریافتند الگوریتم خوشه‌بندی CLARA عملکرد بهتری دارد. در ایران نیز پژوهش‌هایی در رابطه با خوشه‌بندی و کاربردهای آن در زمینه‌های مختلف صورت گرفته است؛ نمونه پژوهش با عنوان «خوشه‌بندی و برجسب‌زنی کاربران وبسایت با استفاده از روش‌های داده کاوی» صورت گرفته است که در این پژوهش خوشه‌های بازدیدکنندگان شناسایی شدند و یک تقسیم‌بندی از کاربران بر اساس رفتارشان به دست آمد [۸]. پژوهش دیگری برای بهبود مدیریت ارتباط با کاربران در بانک پارسیان انجام شده است که در آن با استفاده از داده کاوی، روابط بین سرویس‌های مورد علاقه هر خوشه از کاربران کشف گردید تا در به کارگیری استراتژی مناسب برای ارائه سرویس مناسب به کاربران استفاده شود [۹]، [۱۰] با استفاده از الگوریتم خوشه‌بندی  $k$ -میانگین به بررسی شناسایی ساعات کاری ادارات در شهر تهران در شرایط همه‌گیری ویروس کرونا پرداختند. هدف از این مقاله خوشه‌بندی استان‌های کشور بر مبنای میزان برخورداری از امکانات بهداشتی مانند تعداد تخت‌های بیمارستانی در بخش‌های مختلف (مانند سوختگی، روانی، نوزادان، مراقبت‌های ویژه و غیره)، تعداد کلینیک‌ها، درمانگاه‌ها و بیمارستان‌های دولتی، خصوصی و غیره جهت شناسایی استان‌های مشابه است. در این مقاله از روش ناپارامتری  $k$ -میانگین و روش صوری صورتک چرنف و روش خوشه‌بندی سلسله‌مراتبی استفاده شده است. داده‌های مورد استفاده در این پژوهش از مرکز آمار ایران سال ۱۳۹۵ اخذ شده است. این مقاله به این صورت سازماندهی شده است که بخش اول شامل مقدمه، بخش دوم روش تحقیق که شامل روش‌های خوشه‌بندی است، بخش سوم یافته‌های تحقیق شامل توصیف و تحلیل داده‌ها و بخش آخر نتیجه‌گیری است.

### ۲- روش تحقیق

## ۱-۲- تحلیل خوشه‌ای

تحلیل خوشه یا به صورت ساده خوشه‌بندی، فرایندی است که در آن مجموعه‌ای از اشیا (یا مشاهدات) به زیرمجموعه‌هایی بخش‌بندی می‌شوند. هر یک از زیرمجموعه‌ها خوشه نامیده می‌شود و اشیاء هر خوشه به یکدیگر شبیه هستند و همچنان متفاوت با اشیاء دیگر خوشه‌ها است. مجموعه‌ای از خوشه‌ها که نتیجه‌ی یک تحلیل خوشه است را می‌توان به عنوان خوشه‌بندی شناخت. ممکن است روش‌های خوشه‌بندی متفاوت، خوشه‌های مختلفی را برای داده‌های همگن و یکسانی تولید کنند. بنابراین از آنجا که خوشه‌بندی می‌تواند ما را به کشف گروه‌های ناشناخته در داده‌ها هدایت کند، یک فرایند مفید محسوب می‌شود. به دلیل آنکه در خوشه‌بندی اطلاعات مربوط به برچسب کلاس وجود ندارند، به این فرایند یادگیری بی‌ناظر یا بدون ناظر نیز گفته می‌شود. به همین دلیل خوشه‌بندی شکلی از یادگیری با کمک مشاهده محسوب می‌شود. در این حال، شرط اساسی برای تشکیل خوشه‌ها این است که خوشه‌ها افزای از جامعه یا نمونه باشند. یعنی هر داده تنها در یک خوشه قرار گیرد و از طرفی اجتماع تمام خوشه‌ها، برابر کل نمونه یا جامعه مورد نظر باشد. در خوشه‌بندی هیچ دسته از پیش تعیین شده‌ای وجود ندارد و داده‌ها صرفاً براساس تشابه گروه‌بندی می‌شوند و عناوین هر گروه نیز توسط کاربر تعیین می‌گردد.

## ۲-۲- خوشه بندی به روش k- میانگین

الگوریتم k- میانگین یکی از مشهورترین الگوریتم‌های خوشه‌بندی داده‌ها است که به دلیل پیاده‌سازی آسان و سرعت عملکرد، محبوبیت زیادی یافته است. روش خوشه‌بندی k- میانگین اولین بار توسط [11] مطرح شد. الگوریتم k- میانگین با تعیین تابع هدف بر اساس میانگین فاصله اعضای هر خوشه نسبت به میانگین‌شان، عمل می‌کند و به شکلی مشاهدات را در خوشه قرار می‌دهد تا میانگین مجموع مربعات فاصله‌ها در خوشه‌ها، کم‌ترین مقدار را داشته باشد. اگر مشاهدات را  $X$  نمایش داده شود، تابع هدف الگوریتم k- میانگین را می‌توان به صورت زیر نوشت:

$$\sum_{j=1}^k \sum_{x \in c_j} dist(x, \mu_{c_j}) \quad (1)$$

که در آن  $\mu_{c_j}$  میانگین خوشه  $c_j$  و  $dist$  نیز مربع فاصله اقلیدسی است. در این الگوریتم میانگین نقاط موجود در یک خوشه به عنوان نماینده‌ی خوشه معرفی می‌شود. در ابتدا از بین مشاهدات موجود در مجموعه داده‌ها به صورت تصادفی تعداد  $k$  مشاهده به عنوان نمایندگان خوشه‌ها انتخاب می‌شوند. با محاسبه فاصله اقلیدسی میان مشاهدات باقیمانده و این  $k$  مشاهده، هر یک از این مشاهدات به یکی از این  $k$  خوشه تعلق می‌گیرند. الگوریتم k- میانگین با تکرار سعی می‌کند تا اختلاف درون خوشه‌ای را بهبود بخشد. پس از هر تکرار، با کمک مشاهدات موجود در هر خوشه مقدار میانگین جدید خوشه محاسبه می‌شود. همانند قبل مشاهدات با کمک این میانگین‌های به‌نگام‌شده به عنوان نمایندگان و مراکز خوشه‌های جدید در خوشه‌های خود قرار خواهند گرفت. تکرارها تا جایی ادامه می‌یابد که انتساب پایدار شود، بدین معنی که خوشه‌های شکل گرفته در مرحله جاری با آن‌هایی که در مرحله قبل شکل گرفته بود، یکسان باشند و تغییری نکنند.

## ۳-۲- خوشه‌بندی استان‌ها براساس صورتک چرئف

صورتک‌های چرئف توسط آمار شناسی به نام [12] معرفی شد (با توجه به نمودار ۴ سمت چپ). این روش می‌تواند مجموعه داده‌هایی بالغ بر ۱۸ بعد را با کمک اعضای صورت کارتونی نمایش دهد. صورتک‌های چرئف گرایش‌های موجود در داده‌ها را آشکار می‌کند. شکل، اندازه، محل قرارگیری و جهت اعضاء موجود در چهره مانند چشم‌ها، گوش‌ها، دهان و بینی مؤلفه‌هایی هستند که با کمک آنها مقادیر داده‌ها نمایش داده می‌شوند. برای مثال ابعاد می‌توانند با توجه به خصوصیتی مانند اندازه چشم، طول و پهنا و مقدار باز بودن دهان، اندازه مردمک چشم، کجی ابرو، غیر عادی بودن چشم و سر (کج بودن) نشان داده می‌شوند. صورتک‌های چرئف از این قابلیت استفاده می‌کند که انسان تفاوت‌های کوچک در مشخصات صورت را به خاطر می‌سپارد و در یک لحظه سریع بسیاری از این خصوصیات را مقایسه می‌کند، در صورتی که مشاهده جداول بزرگ از داده‌ها خسته‌کننده خواهد بود. فشرده سازی داده‌ها با کمک صورتک‌های چرئف باعث هضم راحت‌تر آنها توسط کاربر می‌شود. در بخش بعد صورتک‌های چرئف برای خوشه‌بندی استان‌ها استفاده می‌شود. در این روش هر استان با یک صورتک نمایش داده می‌شود و استان‌هایی که شباهت زیادی به هم دارند در یک خوشه قرار می‌گیرند، همچنین هم رنگ بودن سر و صورت صورتک نشان دهنده هم گروه بودن استان‌ها است.

همچنین صورتک چرئف به عنوان نماد نقشه برای نمایش تغییرات اجتماعی و اقتصادی برای لس آنجلس مورد استفاده قرار گرفت (با توجه به نمودار ۴ سمت راست) [13]، در همان زمان صورتک‌های چرئف برای روشن نمودن تفاوت‌های منطقه‌ای ایالات متحده به کار برده شد، به این صورت که برای هر منطقه شاخص‌های نرخ سواد (اندازه چشم)، درآمد (انحنای دهان)، میزان قتل (عرض بینی) و درجه حرارت (اندازه سر) ثبت شده است که هر کدام از این شاخص‌ها یک عضو صورت را شکل می‌دهد و در نهایت برای هر منطقه یک صورت روی نقشه رسم شده است [14]. صورتک‌های چرئف نامتقارن توسط [15] جهت تعمیم تکنیک اولیه ارائه شد (با توجه به نمودار ۴ وسط)، از آنجا که سمت راست و چپ در صورتک‌های چرئف قرینه هستند، در رسم فضا هدر می‌رود. در صورتک‌های چرئف نامتقارن با دو برابر کردن تعداد مشخصات مربوط به صورت، اجازه نمایش تا ۳۶ بعد نیز داده می‌شود. بعدها صورتک چرئف به عنوان یک جایگزین برای نمایش داده‌های ناحیه‌ای توسط [16] ارائه شد. در برنامه کاربردی، نماد چهره نه تنها به عنوان نماینده برای نمایش داده‌های چند متغیره بلکه به عنوان یک نماد برای نمایش امتیازات حاصل از تحلیل مؤلفه‌های اصلی است، که برای رتبه‌بندی استفاده می‌شود. کاربرد دیگر صورتک چرئف استفاده از آن برای خوشه‌بندی و رده‌بندی داده‌ها است ([17]، [18]).

## ۴-۲- خوشه‌بندی سلسله‌مراتبی

برخلاف تکنیک‌های خوشه‌بندی مبتنی بر افراز که تعداد خوشه‌ها به عنوان پارامتر ورودی توسط کاربر مشخص می‌شود، در تکنیک‌های سلسله‌مراتبی مجموعه داده‌ها و معیاری جهت ارزیابی تشابه به عنوان ورودی معین می‌شوند. الگوریتم‌های خوشه‌بندی سلسله‌مراتبی که یک فرایند پایین به بالا را طی می‌کنند، کارشان با قرار دادن هر نمونه داده در یک خوشه مجزا شروع می‌شود. با ادغام خوشه‌ها الگوریتم تا جایی پیش می‌رود که یا کلیه نمونه‌ها در یک خوشه قرار گیرند و یا شرط از پیش تعیین شده‌ای به عنوان پایان اجرا مشخص شده باشد. در طرف دیگر الگوریتم‌های سلسله‌مراتبی قرار دارند که یک استراتژی بالا به پایین را دنبال می‌کنند. در این روش ابتدا کلیه نمونه‌ها در یک خوشه قرار می‌گیرند. پس از آن با کمک یک معیار تشابه در چند مرحله به صورت سلسله‌مراتبی این خوشه‌ها به خوشه‌های کوچکتر تقسیم می‌شوند. این الگوریتم را می‌توان تا جایی که هر نمونه در یک خوشه قرار گیرد، ادامه داد یا اینکه شرطی جهت پایان الگوریتم معین نمود. معمولاً فرایند خوشه‌بندی سلسله‌مراتبی توسط یک نمودار با نام دندروگرام نمایش داده می‌شود.

<sup>1</sup> Dendrogram

## ۲- یافته‌ها

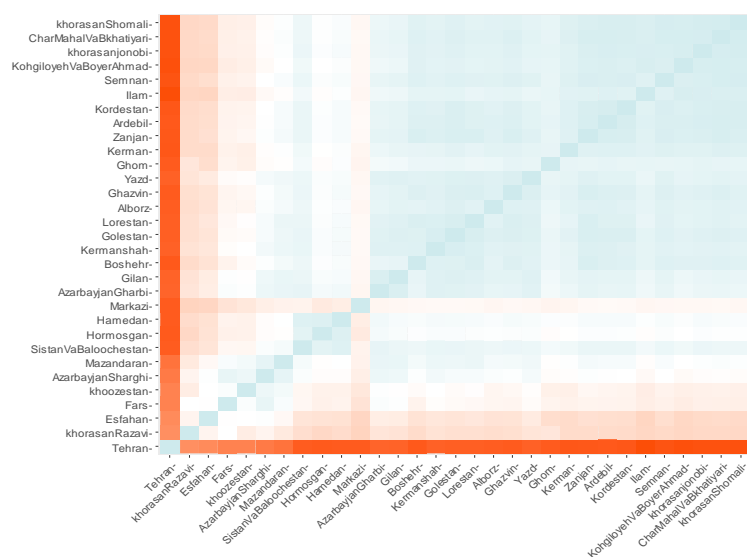
### ۱-۲ توصیف و تحلیل داده‌ها

در این بخش داده‌های بهداشت و درمان اخذ شده از مرکز آمار ایران که حاصل از آخرین سرشماری سال ۱۳۹۵ است، مورد بررسی قرار گرفته است، ساختار این داده‌ها به این صورت است که برای هر استان میزان دسترسی به امکانات بهداشتی و درمانی مانند تعداد تخت‌های بیمارستانی در بخش‌های مختلف (مانند سوختگی، روانی، نوزادان، مراقبت‌های ویژه و غیره)، تعداد کلینیک‌ها، درمانگاه‌ها و بیمارستان‌های دولتی، خصوصی و غیره ثبت شده است. برای شناسایی استان‌های مشابه از نظر برخورداری از این امکانات و سطح برخورداری استان‌ها، داده‌ها به روش ناپارامتری خوشه‌بندی  $k$ - میانگین و دو روش صوری صورتک چرئف و سلسله مراتبی تحلیل می‌شوند. از روی نمودار ۱ می‌توان به میزان تشابه استان‌های مختلف از جهت برخورداری از امکانات بهداشتی و درمانی پی برد. با توجه به شکل می‌توان دید که استان تهران از لحاظ توزیع امکانات بهداشتی و درمانی به هیچ یک از استان‌های دیگر مشابه نیست. استان‌های خراسان رضوی، اصفهان، فارس، خوزستان، آذربایجان شرقی و مازندران بسیار شبیه به هم هستند و سایر استان‌های کشور نیز از لحاظ برخورداری از امکانات بهداشتی و درمانی در یک سطح هستند. با توجه به نمودار ۲ نیز صورتک استان تهران نسبت به سایر استان‌ها بسیار بزرگتر و متفاوت‌تر است و این بیانگر این است که این استان نسبت به سایر استان‌ها از امکانات بیش‌تری برخوردار است. بعد از این استان‌های خراسان رضوی و فارس از امکانات بیش‌تری نسبت به سایر استان‌های کشور برخوردار هستند. استان‌های همدان و هرمزگان نیز صورتک‌های بسیار مشابه‌ای دارند و این بیانگر این است که این دو استان در سطح بسیار نزدیکی به هم هستند. همچنین گیلان و آذربایجان غربی بسیار مشابه به هم هستند. خوشه‌بندی استان‌ها مختلف با توجه به میزان توزیع امکانات بهداشتی و درمانی به روش سلسله‌مراتبی در نمودار ۳ سمت راست آمده است که با توجه به آن می‌توان به طور کلی‌تر استان‌های کشور را به سه خوشه تقسیم نمود.

قبل از خوشه‌بندی به روش  $k$ - میانگین بهتر است تعداد خوشه‌های بهینه را تعیین نمود برای این منظور می‌توان مجموع مربعات فواصل درون خوشه‌ای را به ازای تعداد خوشه‌های مختلف به دست آورد. نمودار ۳ سمت چپ که محور افقی بیانگر تعداد خوشه‌ها و محور عمودی بیانگر مجموع مربعات فواصل درون خوشه‌ای است را رسم کردیم، با توجه به شکل از آنجایی که مجموع مربعات فواصل درون خوشه‌ای زمانی که تعداد خوشه‌ها ۳ است، کمترین مقدار را می‌گیرد و این کمیت از تعداد ۳ خوشه به بعد به مقدار کمی تغییر می‌کند. تعداد خوشه بهینه ۳ در نظر گرفته می‌شود. در نتیجه با اجرای خوشه‌بندی  $k$ - میانگین با ۳ خوشه و با توجه به میانگین‌های حاصل از این روش می‌توان استان‌های کشور را به سه سطح بیش‌ترین امکانات بهداشتی و درمانی شامل شهر تهران، امکانات بهداشتی متوسط شامل شهرهای بزرگ خراسان رضوی، اصفهان، فارس، خوزستان، آذربایجان شرقی و مازندران و کمترین امکانات بهداشتی و درمانی شامل سایر استان‌های کشور، تقسیم می‌شوند (در نمودار ۵ نتایج حاصل از اجرای روش  $k$ - میانگین به صورت نمایش گرافیکی در قالب نقشه ایران مشخص شده است).

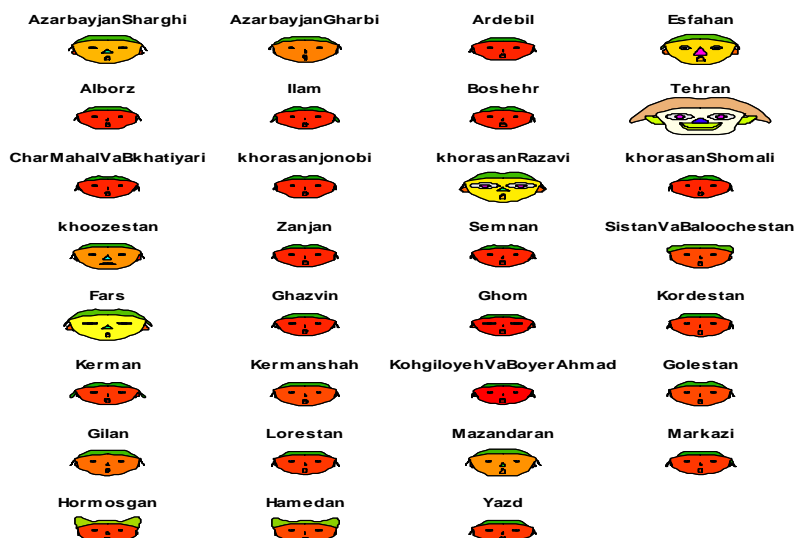
### بحث و نتیجه‌گیری

با توجه به اهمیت بهداشت و درمان و رفاه جامعه سطح برخورداری از امکانات بهداشتی و درمانی یکی از فاکتورهای بسیار مهم توسعه‌یافتگی هر کشور و هر استان محسوب می‌شود. در همین راستا تعیین سطح برخورداری از امکانات بهداشتی و درمانی و شناسایی استان‌های مشابه به موضوعی است که می‌توان با استفاده از آن به نتایج مهمی دست یافت. تعیین تشابه استان‌های مختلف کشور با استفاده از روش‌های خوشه‌بندی  $k$ - میانگین، صورتک چرئف و خوشه‌بندی سلسله‌مراتبی انجام شد. نتایج حاصل از اجرای روش  $k$ - میانگین در نمودار ۵ قابل ملاحظه است. همانطور که مشخص است، استان تهران نسبت به سایر استان‌های کشور از امکانات بهداشتی و درمانی قابل توجهی برخوردار است. استان‌های مازندران، آذربایجان شرقی، خراسان رضوی، اصفهان، فارس و خوزستان بعد از استان تهران بیش‌ترین امکانات بهداشتی و درمانی را دارند.

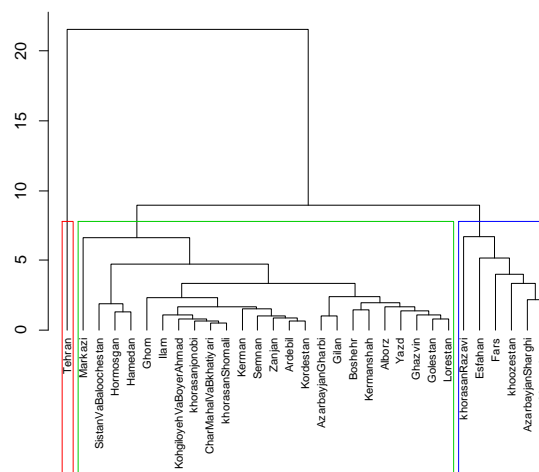
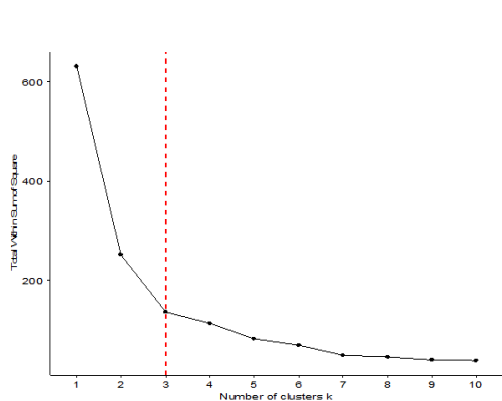


نمودار ۱- نقشه حرارتی استان‌های مختلف کشور بر اساس توزیع امکانات بهداشتی و درمانی.

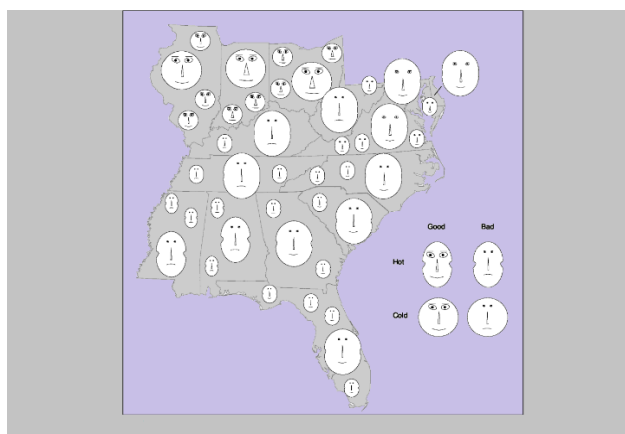
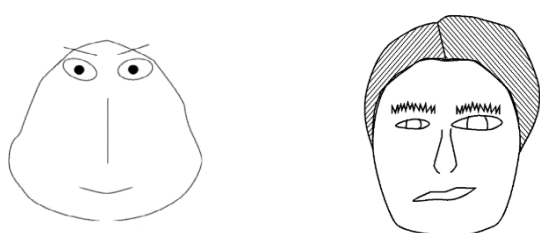
# میرزایی و همکاران، تحلیل خوشه‌های داده‌های بهداشت و درمان ایران



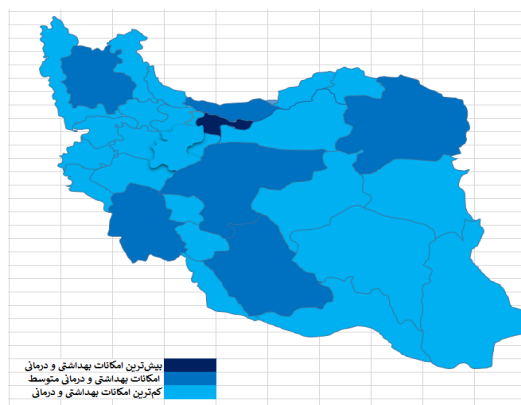
نمودار ۲- صورت چرئف استان‌های مختلف کشور بر اساس توزیع امکانات بهداشتی و درمانی.



نمودار ۳- سمت راست خوشه‌بندی استان‌های مختلف کشور بر اساس توزیع امکانات بهداشتی و درمانی به روش سلسله‌مراتبی، سمت چپ تعداد خوشه‌های بهینه برای خوشه‌بندی استان‌های مختلف کشور بر اساس توزیع امکانات بهداشتی و درمانی به روش k-میانیگین.



نمودار ۴- سمت راست صورتک چرئف به عنوان نماد نقشه برای نشان دادن تغییرات اجتماعی و اقتصادی برای لس آنجلس، وسط صورتک نامتقارن، سمت چپ صورتک چرئف اصلی.



نمودار ۵- نمایش گرافیکی از خوشه‌بندی استان‌های مختلف کشور بر اساس توزیع امکانات بهداشتی و درمانی به روش k-میانیگین.

## مراجع:

- [1] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: a comprehensive survey and performance evaluation," *Electronics*, 9(8), PP.1295, 2020.
- [2] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, 31(8), PP. 651-666, 2010.
- [3] D. Pollard, "Strong consistency of k-means clustering," *The Annals of Statistics*, 9(1), PP. 135-140, 1981.
- [4] D. Fisher, L. Xu, J.R. Carnes, Y. Reich, J. Fennes, J. Chen, R. Shiavi, G. Biswas, and J. Weinberg, "Applying AI clustering to engineering tasks," *IEEE Expert*, 8(6), PP.51-60, 1993.
- [5] M. E. Celebi, "Partitional clustering algorithms," Springer, PP. 79-98, 2014.
- [6] K. Y. Yeung, D. R. Haynor, and W. L. Ruzzo, "Validating clustering for gene expression data," *Bioinformatics*, 17(4), PP.309-318, 2001.
- [7] T. Gupta and S. P. Panda, "A comparison of k-means clustering algorithm and clara clustering algorithm on iris dataset," *International Journal of Engineering and Technology*, 7(4), PP. 4766-4768, 2018.
- [8] خاکباز، محمدحسین، خوشه‌بندی و برجسب‌زنی کاربران وب‌سایت با استفاده از روش‌های داده‌کاوی، دانشگاه تربیت مدرس، ۱۳۸۵.
- [9] حسینی باکمان، محمدحسین، به کارگیری تکنیک‌های داده‌کاوی جهت بهبود مدیریت ارتباط با مشتری در صنعت بانکداری، علامه طباطبائی، ۱۳۸۸.
- [10] شباک، اشکان؛ محمدی، عزت‌اله؛ کرامتی، ودود؛ میرزایی، مریم؛ موسوی، محسن، بررسی شناورسازی ساعات کاری ادارات در شهر تهران در شرایط همه‌گیری ویروس کرونا، تهران پژوهشکده آمار، مرکز آمار ایران، ۱۳۹۹.
- [11] J. MacQueen, "Some methods For classification And analysis Of multivariate observations," In *Proceedings Of the fifth Berkeley symposium On mathematical statistics And probability*, 1(14), PP. 281-297, 1967.
- [12] H. Chernoff, "The use of faces To represent points In k-dimensional space graphically," *Journal Of the American Statistical Association*, 68(342), PP. 361-368, 1973.
- [13] R. Gnanadesikan, J. R. Kettenring, and J. M. Landwehr, "Interpreting and assessing the results of cluster analyses," *Bulletin of the International Statistical Institute*, 47(2), PP. 451-463, 1977.
- [14] H. Wainer, "Graphic experiment In display Of nine variables uses faces To show multiple properties Of States," *Newsletter Of the Bureau Of Social Sciences Research*, 13, PP. 2-3, 1979.
- [15] B. Flury and H. Riedwyl, "Graphical representation Of multivariate data By means Of asymmetrical faces," *Journal Of the American Statistical Association*, 76(376), PP. 757-765, 1981.
- [16] D. R. Montello and M. Gray, "Miscommunicating With isoline preference maps: Design principles For thematic maps," *Cartographic Perspectives*, 1(50), PP. 24-33, 2005.
- [17] R. Song, Z. Zhao, and X. Wang, "An application Of the V-system To the clustering Of Chernoff faces," *Computers And Graphics*, 34(5), PP. 529-536, 2010.
- [18] میرزایی، مریم، خوشه‌بندی در رگرسیون بعد بالا، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس، ۱۳۹۸.

# Cluster analysis of the distribution of sanitation and curative facilities of Iran provinces

1<sup>st\*</sup> Mirzaee, 2<sup>nd</sup> baghfalaki, 3<sup>rd</sup> Maanavi

<sup>1</sup> Statistical Research and Training Center, Tehran, Iran

<sup>2</sup> Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran

<sup>3</sup> Social Determinants of Health Research Center, Semnan University of Medical Sciences, Semnan, Iran

mmaryammirzaie73@gmail.com

**Abstract**— Since sanitation and curative plays a very important role in enjoying social welfare, one of the main goals of policy makers should be to facilitate people's access to health services in society. Therefore, the distribution of sanitation and curative facilities is one of the most important factors in the development of a country. In this regard, this paper has tried to discuss how to distribute of these facilities throughout the country and identify provinces that are similar in terms of sanitation and curative facilities. It also discusses the provinces that have the most, average and minimum of these facilities. For this purpose, the k-mean method and visualization methods such as Chernoff face and hierarchical clustering have been used. By using these methods, Tehran province has the most sanitation and curative facilities with a significant difference. The provinces of Khorasan Razavi, Isfahan, Fars, East Azerbaijan, Khuzestan and Mazandaran have the most sanitation and curative facilities after Tehran and are similar in terms of having the mentioned facilities.

**Keywords**— Chernoff face, clustering, hierarchical, k-means, sanitation and curative facilities.