یادداشت پژوهشی

تحلیل اطلاعات فوت شدگان تصادف‌های جاده‌ای با روش آنالیز خوشه‌ای

شاهین شعبانی، عضو هیئت علمی، پژوهشکده حمل‌ونقل، تهران، ایران
حسین روزیخواه، کارشناس ارشد راه و ترابری، پژوهشکده حمل‌ونقل، تهران، ایران
مهناز نبیل، کارشناس ارشد آمار، دانشگاه تربیت معلم تهران، تهران، ایران
E-mail: shabani@rahiran.ir

چکیده

در سالهاي اخير تصادفات رانندگي نهمين عامل مرگ و مير در جهان و دومين عامل در كشور ايران بعد از بيماريهاي قلبي و عروقي بوده است. جاي تاسف دارد كه ساليانه تعداد زيادي از انسانها به‌دليل تصادف دچار صدمات متعدد جاني و مالي مي‌شوند. تحليل تصادف‌ها به عنوان ابزاري كارآمد مي‌تواند اطلاعات ارزشمندي در جهت شناخت علل تصادف و راهكارهاي متناسب با آن ارايه كند. تحليل تصادف در دو سطح پروژه و شبكه قابل انجام است. در اين مقاله تحليل در سطح شبكه(استانها) و به منظور دستيابي به يك ساز و كار مناسب براي ارتقاي توان مديريت كلان در ارايه راهكارهاي طبقه‌بندي شده و متناسب با شرايط هر منطقه انجام پذيرفته است. به‌دليل آنكه هر يك از استانهاي كشور شرايط خاصي از نظر رفتار راننده، وضعيت زيرساختها و اقدامات پيشگيرانه قبل و اقدامات امدادي پس از حادثه دارند، اين امر لزوم دسته‌بندي استانهاي مشابه به منظور اولويت‌بندي اجراي اقدامات ايمني و اتخاذ تدابير لازم در جهت كاهش تعداد كشته‌هاي تصادفات را مشخص مي‌كند. همچنين بررسي گروهي استانهايي كه وضعيت نزديك به همي دارند كمك مي‌كند تا بتوان اقدامات يكساني را براي حل مشكلات ايمني در دستور كار سياست گذاران قرار داد. يكي از روشهاي مورد استفاده براي دسته‌بندي، استفاده از آناليز خوشه‌اي است. در تحقيقي كه نتايج آن در اين مقاله ارايه شده است، آناليز خوشه‌اي با استفاده از ۶ روش براي استانهاي كشور بر اساس اطلاعات تصادفات سال ۸۳ اخذ شده از سازمان پزشكي قانوني كشور انجام شده است. متغيرهاي مورد نياز براي انجام آناليز خوشه‌اي با بررسي اپيدميولوژي اطلاعات تصادفات حاصل شده است و شامل هشت گروه اصلي تعداد كشته، جنسيت، سن، سطح تحصيلات، شغل، محل فوت، وضعيت كاربر و علت نهايي فوت و مجموعاً با احتساب زير مجموعه متغيرهاي ذكر شده ۶۴ متغير است. اين آناليز در چهار خوشه انجام شده است. در انتها نيز به كمك روش چهره‌هاي چرنوف اعتبار نتايج به‌دست آمده از آناليز خوشه‌اي بررسي شده است. نتايج به‌دست آمده از اين تحقيق حاكي از آن است كه برخي از استانها را مي توان در يك دسته قرار داد ولي برخي ديگر مانند استان اصفهان را بايد به صورت ويژه مورد بررسي قرار داد.

واژه‌های کلیدی: اپیدمیولوژی، تصادف، آنالیز خوشه‌ای، فاصله اقلیدسی، فاصله بلوک شهری، چهره‌های چرنف

۱. مقدمه

تصادف‌ها یکی از شایع‌ترین حوادثی هستند که زندگی انسانها را تهدید می‌کنند و سالانه جان بسیاری از آنها را به مخاطره می‌اندازند[۱]. هم اکنون بیش از ۱/۲ میلیون نفر در جهان سالیانه به علت تصادف جان خود را از دست می‌دهند. بانک جهانی پیش‌بینی کرده است که تا سال ۲۰۲۰، تصادف‌های جاده‌ای ششمین عامل مرگ و میر و سومین عامل معلولیت‌های مادام‌العمر

*تاریخ دریافت: ۸۶/۹/۷ – تاریخ پذیرش: ۸۸/۲/۹

Archive of SID
www.SID.ir

از بین صد عامل شناخته شده مرگ و میر در جهان باشد[۲]. کشور ما یکی از کشورهایی است که بیشترین موارد مرگ ومیر و جراحت ناشی از تصادف را دارد. ارقامی که در سالهای اخیر از طرف مراجع ذیصلاح اعلام شده است در نمودار شکل (۱) نمایش داده شده است[۳]. متأسفانه تصادفها در حال حاضر دومین علت مرگومیر در کشور محسوب می‌شوند و روزانه جان حدود ۷۵ نفر را می‌گیرند و این امر خود هزینه‌های مستقیم و غیرمستقیم زیادی را به جامعه تحمیل می‌کند، به طوری که در تحقیقی که اخیرا توسط پژوهشکده حمل‌ونقل وزارت راه و ترابری انجام شده است، هزینه خسارات ناشی از تصادفها در سال۱۳۸۳ بالغ بر ۷۰۰۰ میلیارد تومان برآورد شده است[٤].

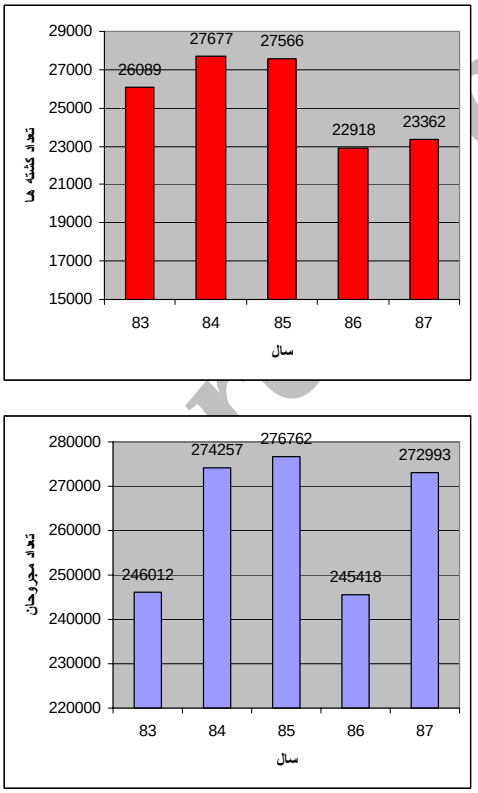
سابقه استفاده از روش آنالیز خوشه‌ای چندان زیاد نیست، ولی به دلیل سودمندی این ابزار، استفاده از آن در حوزه ایمنی و تحلیل تصادفات رو به گسترش است. به عنوان نمونه در مقاله‌ای جهت پیش‌بینی تواتر تصادفها برای حمل‌ونقل کالاهای خطرناک از این روش بهره‌گیری شده است[۵]. در مقاله‌ای دیگر در خصوص آنالیز خوشه‌ای بر سنسورهای ارزیابی خصوصیات رفتاری رانندگان این روش بکار برده شده است[٦]. همچنین این روش برای تعیین قطعات پر حادثه راهها می‌تواند بکاربرده‌شود[۷]. در علوم پزشکی مرتبط با تصادف‌های جاده‌ای نیز جهت هدفمند کردن مطالعات و تحلیل اطلاعات و جامعه آماری، آنالیز خوشه‌ای بسیار سودمند است. مانند مطالعاتی که در خصوص صدمات زانوی کودکان به علت تصادفات جاده‌ای صورت گرفته است[۸].

در این تحقیق برای اولین بار در کشور روش آنالیز خوشه‌ای برای تحلیل اطلاعات تصادف مورد استفاده قرار گرفته است. به همین منظور با هماهنگی با سازمان پزشکی قانونی کشور و با همکاری نیروهای این سازمان مستقر در شهرستانها و استانها، کلیه اطلاعات قربانیان تصادفها در سال ۱۳۸۳ جمع‌آوری و سپس مورد بررسی اپیدمیولوژیک قرار گرفت. نتایج حاصل از آن به عنوان متغیرهای آنالیز خوشه‌ای بکار گرفته شد.

در این بررسی کمیتهای آماری تعداد کشته، محل فوت، وضعیت کاربر، جنسیت، شغل و سایر کمیتهای آماری مورد نیاز، به تفکیک برای هر یک از استانها و به کمک نرم افزارهای آماری بررسی شد و سپس کمیت‌های آماری مذکور در ۸ گروه کلی شامل تعداد کشته، جنسیت، سن، سطح تحصیلات، شغل، محل فوت، وضعیت کاربر و علت نهایی فوت طبقه‌بندی شد که با تقسیم هر گروه به زیر گروههای مربوطه مجموعاً ٦٤ متغیر برای استفاده در آنالیز خوشه‌ای انتخاب گردید. این آنالیز با شش روش شامل روشهای سلسله مراتبی پیوند متوسط، پیوند کامل، ward و تقسیم‌بندی شده و روشهای غیر سلسله مراتبی k-mean و P.A. Medoids در چهار خوشه انجام شد.

هدف از آنالیز خوشه‌ای تصادفها، شناسایی گروههای همسان تصادف در مطالعاتی است که به منظور شناسایی راهکارها و اقدامات طبقه‌بندی شده کاهش تصادف در سطح کلان انجام می‌پذیرد. با توجه به اینکه این گروهها خصوصیات مشترک بیشتری دارند، می‌توان ایده‌های کاهش تصادف را در آنها به وجود آورد و راهکارهای نیل به این هدف را تدوین کرد. آنالیز خوشه‌ای عموماً تکنیکی است که پیش از آنکه هدف گرا باشد، موضوع گرا است و می‌تواند به عنوان یک روش مناسب بکار می‌رود[۹].

در انتها با استفاده از روش چهره‌های چرنوف اعتبار سنجی نتایج تحلیل خوشه‌ای مورد بررسی قرار گرفته است.



شکل۱. تعداد کشته‌ها و مجروحین تصادفات کشور ایران در سالهای اخیر[۳]

www.SID.ir

۲. تحلیل خوشه‌ای

خوشه‌بندی یک روش تحلیل چند متغیری است. در تحلیل چندمتغیری مشاهدات شامل اندازه‌های همزمان روی چندمتغیرند. از جمله اهدافی که در روشهای چندمتغیری‌و بالاخص خوشه‌بندی دنبال می‌شوند به موارد زیر می‌توان اشاره‌کرد[۱۱]:

۱ – کاهش داده‌ها یا آسان سازی ساختاری ۲– دسته‌بندی کردن.

تحلیل خوشه‌ای روشی است که در آن هیچ فرضی در مورد تعداد دسته‌ها یا ساختار آنها در نظر گرفته نمی‌شود و دسته‌بندی کردن بر اساس مشابهت‌ها یا عدم مشابهت‌ها (فواصل) انجام می‌شود. در واقع ورودیهای این روش اندازه‌های مشابهت(فواصل) یا داده‌هایی هستند که از آن می‌توان شباهتها را محاسبه کرد[۱۲].

هدف اصلی تحلیل خوشه‌ای این است که دسته‌های طبیعی اقلام یا متغیرها را جستجو کنیم. برای این منظور ابتدا باید مقیاسی کمّی پیدا شود که به وسیله آن تشابه بین اشیاء اندازه‌گیری گردد. راههای زیادی برای اندازه‌گیری شباهت بین زوجهای اشیاء وجود دارد، از جمله استفاده از دو معیار فاصله اقلیدسی و فاصله بلوک شهری[۱۱].

فاصله اقلیدسی و فاصله بلوک شهری بین دو مشاهده p بعدی $\underline{x}_i = [x_{i1}, x_{i2}, ..., x_{ip}]$ و $\underline{x}_j = [x_{j_1}, x_{j2}, ..., x_{jp}]$ به ترتیب از روابط (۱) و (۲) استفاده می‌شود[۱۲]:

(۱) فاصله اقلیدسی(E)
$$d_{(i,j)} = \sqrt{\sum_{f=1}^{p} (x_{if} - x_{jf})^2}$$

(۲) فاصله بلوک شهری(M)
$$d_{(i,j)} = \sum_{f=1}^{p} \left| x_{if} - x_{jf} \right|$$

در تحلیل خوشه‌ای علاوه برموضوع تعیین فاصله(شباهت‌ها) موضوع مهم دیگری که باید در نظر گرفته شود عبارتست از انتخاب روش خوشه‌بندی بر اساس فاصله. همان طور که قبلاً اشاره شد روشهای تحلیل خوشه‌ای به دو دسته کلی سلسله مراتبی و غیر سلسله مراتبی تقسیم می‌شود که در ادامه به هر یک از آنها اشاره می‌شود.

۲-۱ خوشه‌بندی سلسله مراتبی

خوشه‌بندی سلسله مراتبی با یک سری از ترکیب کننده‌های متوالی (روشهای سلسله مراتبی متراکم) یا با یک سری از تقسیم‌بندی‌های متوالی(روشهای سلسله مراتبی تقسیم‌بندی شده) انجام می‌شود[۱۱].

روشهای سلسله مراتبی متراکم با اقلام تکی شروع می‌شود. از این رو در ابتدا به تعداد اقلام، خوشه وجود دارد. سپس اقلامی که بیشترین شباهت را دارند، دسته بندی می‌شوند. دسته‌های اولیه بر اساس شباهتهایشان ظاهر می‌شوند. سرانجام وقتی شباهتها کاهش می‌یابند تمام زیر دسته‌ها به یک خوشه تبدیل می‌شوند[۱۱].

روشهای سلسله مراتبی تقسیم‌بندی شده در جهت معکوس روش فوق عمل می‌کنند. ابتدا تمامی اقلام در یک خوشه قرار دارند، سپس به دو زیر دسته چنان تقسیم می‌شوند که اقلام در یک دسته از ویژگیهای کاملا متفاوتی نسبت به اقلام دسته دیگر برخوردار باشد. سپس این زیر دسته‌ها به زیر دسته‌های مشابه تقسیم می‌شوند و فرآیند ادامه می‌یابد تا تعداد خوشه‌ها(دسته‌ها) برابر با تعداد اقلام شود. به عبارت دیگر تا زمانی‌که هر شیئ یک دسته بسازد، نتایج هر دو روش متراکم و تقسیم‌بندی شده را می‌توان به شکل یک نمودار دو بعدی بنام دندروگرام نمایش داد[۱۱].

شاخه‌ای از روشهای سلسله مراتبی متراکم، روشهای پیوندی است که به نوبه خود به روشهای پیوند تکی(کمترین فاصله یا نزدیک ترین همسایه)، پیوند کامل(بیشترین فاصله یا دورترین همسایه) و پیوند متوسط (فاصله متوسط) تقسیم می‌شوند[۱۳].

پیوند تکی وقتی نتیجه می‌شود که دسته‌ها بر اساس فاصله بین نزدیک‌ترین اعضایشان ترکیب شوند. پیوند کامل وقتی پیش می‌آید که دسته‌ها بر اساس فاصله بین دورترین اعضایشان ترکیب شوند. برای پیوند متوسط فاصله بین زوجهای مجموعه‌های مربوط ترکیب می‌شوند[۱۳].

چون پیوند تکی خوشه‌ها را با کوتاه‌ترین پیوند بین آنها به هم وصل می‌کند، بنا براین، این روش خوشه‌هایی را که خوب از هم جدا نشده‌اند نمی‌تواند تمیز دهد. از سوی دیگر پیوند تکی یکی از چند روش خوشه‌بندی است که می‌تواند خوشه‌های غیربیضوی را توصیف کند[۱۳].

شاخه دیگر روشهای سلسله مراتبی روش ward است. در روش ward ادغام کردن دسته‌ها بر اساس کمترین افزایش در مجموع مربعات درون هر دسته انجام می‌شود[۱۲].

به طور کلی مراحل مربوط به الگوریتم خوشه‌بندی با روش سلسله مراتبی متراکم برای دسته‌بندی N شیئ(اقلام یا متغیرها) به قرار زیر است[۱۱]:

Archive of SID
www.SID.ir

۱- با N خوشه هر یک شامل یک ماهیت و یک ماتریس متقارن از فاصله‌ها (یا مشابهت‌ها)، $D=\{d_{ik}\}$، شروع می‌شود.
۲- ماتریس فاصله برای نزدیک‌ترین(مشابه‌ترین) زوج خوشه‌ها جستجو شود. فرض می‌شود d_{UV} فاصله بین مشابه‌ترین خوشه‌های Uو V باشد.
۳- خوشه‌های U و V ترکیب می‌شوند.خوشه جدید به نام(UV) نامگذاری شود. درایه‌های ماتریس فاصله (الف) با حذف سطرها و ستونهای متناظر با خوشه‌هایU و V و (ب) با افزودن یک سطر و ستون که فواصل بین خوشهٔ (UV) و خوشه‌های باقیمانده را می‌دهد، بهنگام شود.
٤- مراحل ۲ و ۳ جمعا N-1 دفعه تکرار می‌شود. در خاتمه تمام اشیاء در یک خوشه خواهند بود. همانندی خوشه‌های ترکیب شده و سطوح(فواصل یا مشابهت‌ها) که در آنها ترکیب‌ها به وقوع پیوسته‌اند، ثبت می‌شود.
بهنگام کردن ماتریس فاصله مرحله ۳ به انتخاب نوع پیوند(تکی، کامل، متوسط) بستگی دارد. فرمول محاسبه به تفکیک روشها به شرح زیر است:

(۳) پیوند تکی
$$d_{(uv)w} = \min\{d_{uw}, d_{vw}\}$$

(٤) پیوند کامل
$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\}$$

(۵) پیوند متوسط
$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W}$$

در رابطه(۳) d_{ik} فاصله بین اشیاء در خوشه(uv) و شیئ k در خوشه w، و $N_{(UV)}$ و N_W به ترتیب تعداد اشیاء در خوشه‌های (uv) و w است.

۲-۲ خوشه‌بندی غیر سلسله مراتبی

روشهای خوشه‌بندی غیر سلسله مراتبی برای دسته‌بندی اقلام به جای متغیرها به مجموعه‌ای از k خوشه طراحی شده است. تعداد خوشه‌ها، یعنی k، ممکن است از پیش مشخص شده باشد یا به عنوان بخشی از روش خوشه بندی تعیین شود[۱۳].
روشهای غیر سلسله مراتبی از افراز اولیه اقلام به دسته‌ها یا یک مجموعه اولیه نقاط تغذیه‌ای که اساس خوشه‌ها را می‌سازد، شروع می‌کند. انتخابهای خوب برای پیکربندی‌های اولیه باید عاری از ناآرایی‌های معلوم باشد. یک راه شروع این است که نقاط تغذیه را از میان اقلام موجود به صورت تصادفی انتخاب کنیم یا این که اقلام را به دسته‌های اولیه به تصادف افراز کنیم. یکی از روشهای متداول غیر سلسله مراتبی روش k-میانگین است. در این روش در ساده‌ترین شکل به صورت زیر عمل می‌شود**[۱۱]**:
۱- اقلام به k خوشه اولیه افراز می‌شود.
۲- از فهرست اقلام یک قلم انتخاب و به خوشه‌ای که مرکز(میانگین) آن نزدیک‌ترین است، نسبت داده می‌شود. مرکز برای خوشه‌ای که قلم جدیدی را دریافت می‌کند و برای خوشه‌ای که قلمی را از دست می‌دهد، دوباره محاسبه می‌شود.
۳- مرحله دوم تکرار می‌شود. تا این که تخصیص دوباره‌ای صورت نگیرد.
نوع دیگری از روشهای خوشه بندی غیر سلسله مراتبی روش Partitioning around Medoids است. در این روش، مرکز به جای میانگین با میانه محاسبه می‌شود. یکی از مزیتهای این روش این است که مرکز هر خوشه را هم به‌دست می‌آوریم و این مرکز را می‌توان به عنوان نماینده خوشه در نظر گرفت**[١٤]**.

۳. روش تحقیق

همان طور که قبلا ذکر شد گستره تحقیق، اطلاعات قربانیان ناشی از تصادفات استانهای کشور است که از فرمهای پزشکی قانونی استخراج شده است . البته استان تهران به دلیل عدم دسترسی به اطلاعات آن در این تحقیق لحاظ نشده است. از طرفی با توضیحات ارایه شده در بند ۲ و بررسی ویژگیهای روش‌های متداول در آنالیز خوشه‌ای به استثنای روش پیوند تکی سایر روش‌ها برای انجام تحلیل مناسب تشخیص داده شدند.
به این ترتیب روش تحقیق به شرح ذیل انتخاب گردیده است:
۱- انتخاب متغیرهای مورد نظر از جامعه آماری قربانیان تصادفات به عنوان داده‌های تحلیلی۲- تصمیم‌گیری در خصوص استفاده از داده‌های استاندارد شده یا غیر استاندارد۳- محاسبه فاصله بین دو مشاهده (فاصله بین متغیرها در استانهای مختلف که بیان کننده میزان شباهت آنها به یکدیگر است).
٤- خوشه‌بندی استانها با روشهای مختلف ۵- اعتبارسنجی نتایج حاصل از خوشه‌بندی به کمک آنالیز چهره‌های چرنف۵- تجمیع

Archive of SID
www.SID.ir

و مقایسه نتایج حاصل از خوشه‌بندی و ارایه خوشه‌های نهایی. در ادامه به تشریح هر یک از بندهای فوق پرداخته شده است.

٤. انتخاب متغیرها

در این تحقیق، استان‌ها به عنوان مشاهدات در نظر گرفته شده و هر کدام از سطوح اطلاعات کیفی مندرج در فرم اطلاعات پزشکی قانونی به عنوان یک متغیر در نظر گرفته شده است. این متغیرها مجموعاً ٦٤ عدد شامل ٨ گروه اصلی تعداد کشته، جنسیت، سطح تحصیلات، شغل، محل فوت، علت نهایی فوت، وضعیت کاربر و سن متوفی است که پس از بررسی اپیدمیولوژی اطلاعات جمع‌آوری شده به‌دست آمده است. به طور مثال تعداد کشته به ١٢ متغیر شامل تعداد کشته در هر یک از ماههای سال تقسیم شده است. سطح تحصیلات نیز به ٦ متغیر تفکیک شده است. به همین ترتیب با تقسیم هر یک از گروههای مربوطه مجموعاً ٦٤ متغیر حاصل شد که در ادامه به کلیات آنها اشاره می‌شود.

٤-١ بررسی تعداد کشته‌ها[٣]

آمار سال ١٣٨٣حاکی از کشته شدن ٢٦٠٨٩ نفر در کل کشور در اثر تصادفات رانندگی است که استان تهران و خراسان به ترتیب با ٣٢٢١ و ٢٤٦٩ مورد بیشترین، و استان کهکیلویه وبویراحمد و ایلام به ترتیب با ١٦٥ و ١٧٣ مورد کمترین تعداد کشته را داشته‌اند.

همچنین نرخ مرگ و میر به ازای هر یک صدهزار نفر جمعیت ٣٨/٧ نفر بوده است که استانهای سمنان، گیلان و قزوین به ترتیب با ٩٥ ، ٧٧/٤، ٥٦/٦ بیشترین، و استانهای اردبیل،آذربایجان غربی و کهکیلویه و بویراحمد به ترتیب با ٢٣/٣، ٢٤/٢، ٢٥/٥ کمترین نرخ مرگ و میر را داشته‌اند.

٤-٢ بررسی جنسیت

بررسی‌های به عمل آمده نشان داده است که ٨١ درصد کل قربانیان، مردان و ١٩ درصد آنان را زنان تشکیل داده‌اند که از این میان استانهای خراسان و ایلام به ترتیب با ٨/٥ و٠/٦ درصد بیشترین و کمترین سهم را در میان قربانیان مرد و استانهای زنجان و کهکیلویه و بویراحمد به ترتیب با ٢/٣ و٠/١ درصد بیشترین و کمترین سهم را در میان قربانیان زن داشته‌اند.

٤-٣ بررسی سن

این مطالعه نشان داد که افراد ٥٠ سال به بالا به میزان ٢٣/٨ درصد و ٢٤-١٨ سال به میزان ٢١/٩درصد، بیشترین نسبت قربانیان ناشی از تصادفات را در کشورتشکیل می دهند. در گروه سنی کمتر از ١٠ سال، استان اردبیل با ١٢/٤درصد بیشترین و استان بوشهر با ٤درصد، کمترین میزان کشته‌ها را داشته‌اند. در گروه سنی ١٧-١١ سال استان یزد با ١١/٧ درصد، بیشترین و استان مرکزی با ٥/٧ درصد، کمترین کشته را داشته‌اند. در گروه سنی ٢٤-١٨ سال، استان بوشهر با ٣١/٧ درصد، بیشترین و استان کرمانشاه با ١٥/٩ درصد، کمترین نسبت کشته‌ها را داشته‌اند. همچنین در گروه سنی ٥٠سال و بالاتر استان کرمانشاه با ٣٢/٧ درصد، بیشترین و استان بوشهر با١٣/٧ درصد، کمترین کشته ناشی از تصادفات را داشته‌اند.

٤-٤ بررسی سطح تحصیلات

بررسی‌های انجام شده در بین افراد ١٨ سال به بالا نشان داده است که ٢٣/٦ درصد قربانیان تصادف بیسواد،٢٣/٣ درصد دارای تحصیلات ابتدایی،٢١/٢ درصد دارای تحصیلات راهنمایی،١٩/١ درصد دارای تحصیلات متوسطه،٦/٣ درصد دارای تحصیلات دانشگاهی و تحصیلات ٦/٥ درصد افراد نامعلوم بوده است. نسبت قربانیان بیسواد در استان ایلام با ٣٥/٦ درصد، بیشترین و در استان سمنان با ١٢ درصد، کمترین مقدار را در بین استانهای کشور داشته‌اند. قربانیان دارای تحصیلات متوسطه در استان ایلام با٣٢/٩ درصد، بیشترین و در استان کردستان با ١٣/٨ درصد، کمتر از دیگر استانها بوده است. همچنین بیشترین نسبت افراد دانشگاهی کشته شده به استان کهکیلویه و بویراحمد با ١٠/٩درصد و کمترین مقدار به استان سیستان و بلوچستان با ٣/٩ درصد تعلق داشته است.

٤-٥ بررسی شغل

٢٩ درصد افراد شغل آزاد،١٣/٩ درصد کارگر، ١١/٨ درصد خانه‌دار و ١٠/٤ درصد افراد دانش‌آموز، بیشترین نسبت قربانیان ناشی از تصادف را تشکیل داده‌اند.

نسبت افرادی که شغل آزاد داشته‌اند در استان سیستان و بلوچستان با ٣٧/٢ درصد بیشترین و استان گلستان با ١٤ درصد

Archive of SID

www.SID.ir

کمترین آمار را نشان داده‌اند. نسبت کارگر کشته شده در استان بوشهر با ۳۰/٤ درصد، بیشترین و در استان کهکیلویه و بویراحمد با ۳/٦ درصد، کمترین مقدار را داشته است. نسبت خانم‌های خانه‌دارکشته‌شده در استان ایلام با ۱۹/۱ درصد، بیشترین و در استان بوشهر با۷/٤ درصد، کمترین مقدار را نشان داده‌است. دانش‌آموزان نیز از گروههای مهم در تصادف‌ها محسوب می‌شوند. نسبت دانش آموزان کشته‌شده در استان کهگیلویه و بویراحمد با ۱۵/۲ درصد بیشترین و در استان مرکزی با ۷/۵ درصدکمتر از سایر استانها بوده است.

٤-٦ بررسی محل فوت

بیش از نیمی از فوتی‌ها یعنی۵۱/۱ درصد در محل تصادف، ۳۷/۱ درصد در بیمارستان، ٤ /۱۰درصد در حین انتقال به بیمارستان و ۰/۹ درصد نیزبعد از ترخیص در منزل فوت کرده‌اند. محل فوت۵ /۰ درصد افراد نیز نامعلوم بوده است. بیشترین تعداد کشته‌ها در محل حادثه در استان ایلام با ۷۰/۵ درصد، و کمترین در استان گلستان با ۳۹/۳ درصد، بوده است. بیشترین نسبت کشته‌ها درحین انتقال به بیمارستان در استان کردستان با ۱۹/۵ درصد، و کمترین در استان قم با ٤/۵ درصد اختصاص داشته است. همچنین بیشترین نسبت فوت در بیمارستان مربوط به استان گلستان با ٤٤/۷ درصد، و کمترین مربوط به استان ایلام با ۱۸/۵ درصد بوده است.

٤-۷ بررسی وضعیت کاربر

۳۲ درصد افرادی که درتصادف کشته شده‌اند سرنشین خودرو، ۲۳/۱ درصد موتورسوار، ۲۲/۷ درصد عابرپیاده،۱۹/۲ درصد راننده و ۰/۳ درصد دوچرخه سوار بوده‌اند. همچنین وضعیت۲/۷ درصد افراد نامعلوم گزارش شده است. بیشترین نسبت سرنشینان کشته شده مربوط به استان سمنان، با ۵۱/٦ درصد و کمترین نسبت در استان گلستان با ۲۰ درصد رخ داده است. همچنین استان بوشهر با ٤۰/۸، درصد بیشترین و استان کرمانشاه با ۹/٦ درصدکمترین نسبت کشته‌های موتورسوار را داشته‌اند. نسبت مرگ و میر عابران پیاده نیز در استان کرمانشاه با ۳٤ درصد بیشترین و در استان سمنان با ۷/۳ درصد کمترین مقدار را داشته است.

٤-۸ بررسی علت نهایی فوت

این تحلیلها نشان داده‌اند که علت اصلی مرگ ٦۰/٤ درصد افراد، ضربه به سر، ۱۷/۵ درصد شکستگی‌های متعدد، ۹ درصد علل اشتراکی(یعنی دارای بیش از یک علت)،۷/٦ درصد خونریزی و ۳/۵ درصد سایر علل بوده است.

بررسی به عمل آمده نشان می‌دهد که بیشترین نسبت ضربه به سر در استانهای آذربایجان شرقی با ۷٦/۱ درصد و کمترین آن در استان قم با٤۱ درصد بوده است. بیشترین نسبت فوت ناشی از شکستگی‌های متعدد مربوط به استان قم با ٤۸/۲ درصد، و کمترین آن مربوط به استان ایلام با۷/۲ درصد، بوده است. بیشترین تعداد کشته‌های ناشی از خونریزی مربوط به استان ایلام با ۲۸/۹ درصدو استان قم با ۰/۵ درصد کمترین نسبت را داشته‌اند.

۵. تصمیم در مورد استفاده از داده‌های استاندارد شده یا غیر استاندارد

بعد از انتخاب متغیرها، داده‌ها استاندارد شد تا تحلیل بر اساس آنها انجام شود. در صورتی که از داده‌های غیر استاندارد استفاده شود، متغیرهای کم اهمیت و متغیرهایی که تعداد کشته‌ها را به رده‌های بیشتری اختصاص داده‌اند در یک رده اهمیت قرار می‌گیرند و اهمیت واقعی متغیرها نادیده گرفته می‌شود. در واقع چون داده‌ها استاندارد شده‌اند، می‌توان گفت که هیچ استانی وزن بیشتری نسبت به استانهای دیگر ندارد. به عبارت دیگر استانها با یک مقیاس سنجیده شده‌اند. به عنوان مثال بزرگی و کوچکی استانها در چگونگی خوشه بندی تاثیری نمی گذارند.

٦. محاسبه فاصله بین دو مشاهده

در این تحقیق برای دستیابی به بهترین نتیجه اولا سعی شد چندین روش مختلف در تحلیل خوشه‌ای بکار گرفته شود و ثانیا در هر روش تحلیل چند روش مختلف برای محاسبه فاصله(شباهت‌ها) مورد آزمایش قرار گیرد. به همین دلیل از دو روش فاصله اقلیدسی و بلوک شهری برای محاسبه فاصله بین مشاهدات استفاده شد.

Archive of SID

www.SID.ir

۷- خوشه‌بندی

با توجه به اینکه مسئله خوشه‌بندی استانهای کشور مساله‌ای خاص است. از چندین روش خوشه‌بندی استفاده شد تا اگر برآمدهای حاصل از آنها تقریبا با یکدیگر سازگار باشند، بتوان دسته‌بندی را طبیعی شمرد.

آنالیز خوشه‌ای به روشهای سلسله مراتبی پیوند متوسط، پیوند کامل، ward و تقسیم‌بندی شده، و روشهای غیر سلسله مراتبی K-mean و P.A. Medoids انجام شد.. آنالیز توسط نرم افزار قدرتمند S^+ صورت گرفت.در بین روشهای خوشه‌بندی تنها نتایج روش پیوند تکی به علت مشکلاتی که عنوان شد کنار گذاشته شد.

۷-۱ تعیین تعداد خوشه‌ها

برای رسیدن به راه حلی مناسب برای تعیین تعداد خوشه‌های مناسب موجنا(۱۹۷۷)[۱] [۱۵]روشی پیشنهاد کرده است که تعداد خوشه‌ها به اندازه نسبی سطوح اتحاد دندروگرام بستگی دارد.

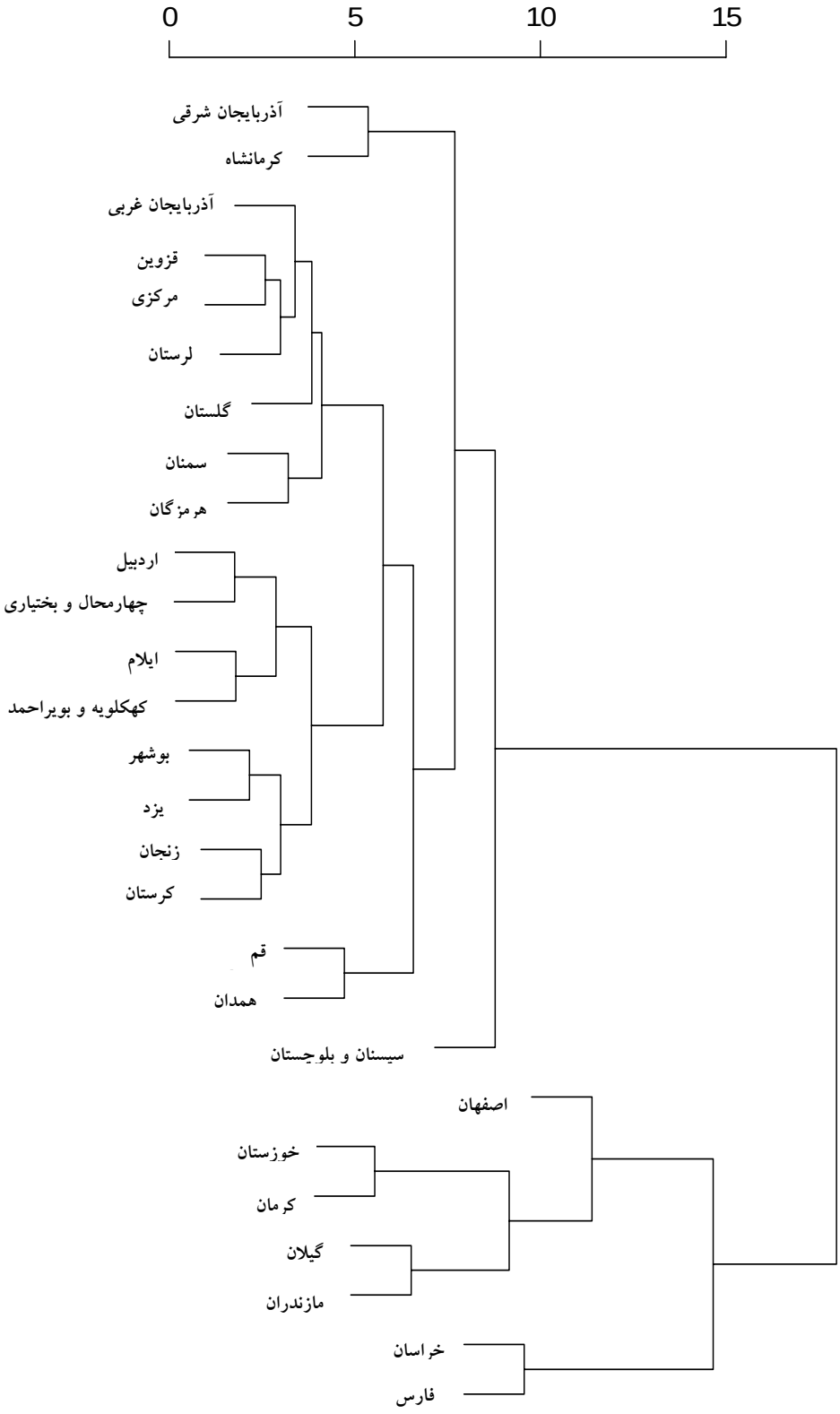
به عبارت دقیق‌تر در دندروگرام اولین مرحله‌ای که $\alpha_{j+1} \langle \overline{\alpha} + ks_\alpha$ باشد، در نظر گرفته می‌شود. تعداد خوشه‌ها برابر با تعداد گروه‌های متناظر با آن مرحله است. در این رابطه $\alpha_0, \alpha_1, ..., \alpha_{n-1}$ سطوح متناظر با داشتن ۱،...، n-1 و n خوشه است و $\overline{\alpha}$ و s_α به ترتیب میانگین و انحراف استاندارد نااریب، و α و k مقادیر ثابت هستند. با وجود آن که میلیگان و کوپر(۱۹۸۵)[۲] [۱۷] مقدار ۱/۲۵ را برای k مقدار مناسبی می‌دانند، موجنا توصیه کرده است که مقدار k باید بین ۲/۷۵ و ۳/۵۰ باشد. در جدول ۷ محاسبه تعداد خوشه‌ها با انواع روشهای خوشه‌بندی ارایه شده است.

تحلیل خوشه‌ای با کمک نرم افزار S^+ انجام شد. خروجی این نرم افزار به صورت دندروگرام ارایه می‌شود. یک نمونه از آن در شکل ۱ آورده شده است. در جدول ۸ نتایج خوشه بندی در چهار خوشه و با روشهای ذکر شده برای استانهای کشور جمع‌بندی و خلاصه شده است.

جدول۷. تحلیل تعداد خوشه‌های لازم

خوشه‌بندی سلسله مراتبی متراکم						
		$\overline{\alpha}$	S_α	K=1.25	$K \in (2.75,3.50)$	
Average	E	۵/۶۱۷۷	٤/۱۳۹۰	٤	۱	۲
	M	۳۰/۳٦۸۸	۲٦/۰۵۳۰	۳	۱	۲
Complete	E	٦/۹۷۱۹	۷/۰۱۸۲	۳	۲	۲
	M	٤۰/۳۹۱۳	٤۸/٦۱۱۰	۳	۲	۲
Ward	E	۸/۱۳٤۲	۱۰/۳٤٤۵	۳	۲	۲
	M	٤۷/٤۰۰٤	۷۲/۰۷۵۸	۳	۲	۲
خوشه‌بندی سلسله مراتبی تقسیم بندی شده						
E		۷/۰۲۰۳	۷/۰۱۵۷	۳	۲	۲
M		٤۰/۲۲۲۷	٤۸/٤۲٤۵	۳	۲	۲

www.SID.ir



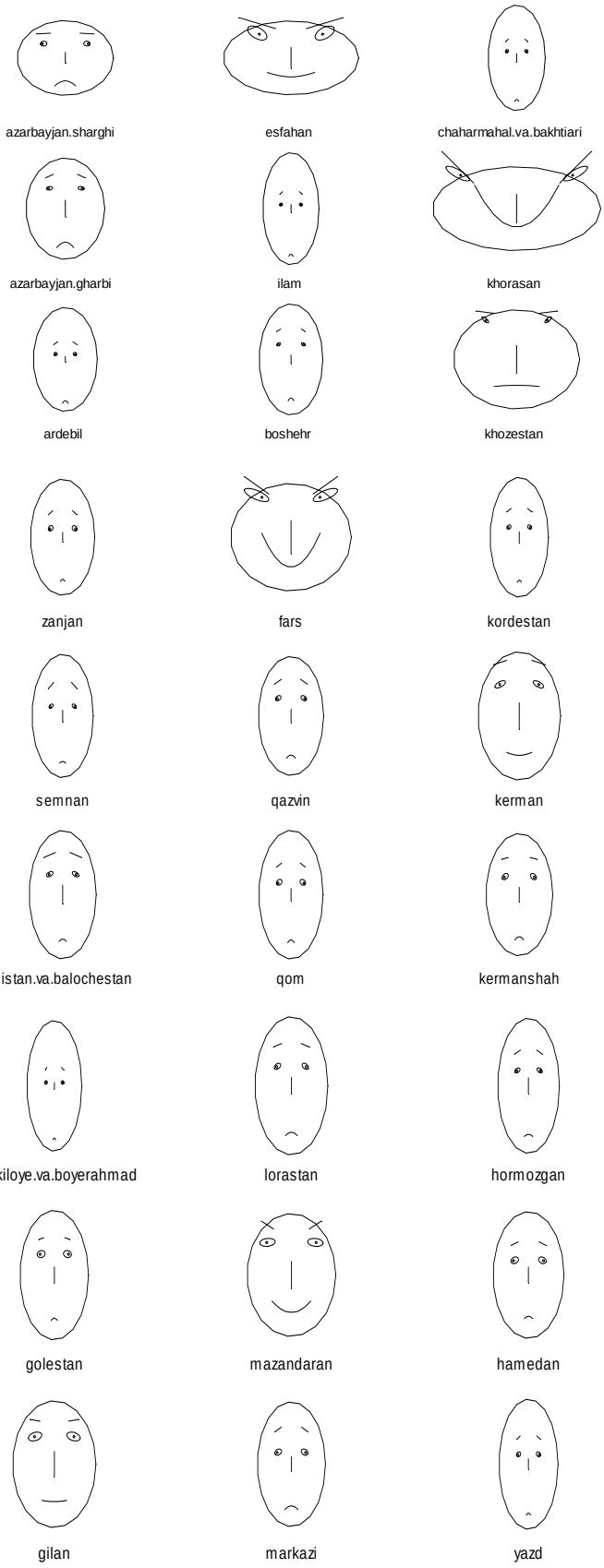
شکل ۱. دندروگرام خوشه‌بندی سلسله مراتبی پیوند متوسط- فاصله اقلیدسی

www.SID.ir

جدول۸ خوشه‌بندی استانها با چهار خوشه

روشهای خوشه‌بندی		فاصله	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ٤
سلسله مراتبی	پیوند متوسط	E	خراسان فارس	خوزستان گیلان کرمان مازندران	اصفهان	بقیه استانها
		M	خراسان فارس	خوزستان گیلان کرمان مازندران	اصفهان	بقیه استانها
	پیوند کامل	E	خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	آذربایجان شرقی کرمانشاه سیستان و بلوچستان	بقیه استانها
		M	خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	اردبیل چهارمحال ایلام کهکلویه	بقیه استانها
	Ward	E	خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	اردبیل چهارمحال ایلام کهکلویه یزد زنجان کردستان	بقیه استانها
		M	خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	اردبیل چهارمحال ایلام کهکلویه یزد زنجان کردستان قم بوشهر	بقیه استانها
	روش تقسیم بندی شده	E	خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	آذربایجا کرمانشاه ن شرقی سیستان و بلوچستان	بقیه استانها
		M	خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	اردبیل چهارمحال ایلام کهکلویه یزد زنجان کردستان قم سمنان بوشهر	بقیه استانها
سلسله مراتبی	k-means		خراسان فارس	خوزستان گیلان اصفهان کرمان مازندران	اردبیل چهارمحال ایلام کهکلویه یزد زنجان کردستان قم سمنان بوشهر	بقیه استانها
	P.A.MEDOIDS		خراسان **فارس** اصفهان	خوزستان گیلان **کرمان** مازندران	اردبیل **چهارمحال** ایلام کهکلویه یزد زنجان کردستان بوشهر	بقیه استانها (**قزوین**)

www.SID.ir



شکل ۲. چهره‌های چرنوف برای استانهای کشور

www.SID.ir

۸. اعتبارسنجی نتایج تحلیل با روش چهره‌های چرنوف[۱][۱۷]

آنالیز چرنوف با استفاده از مشاهدات p بعدی و به صورت چهره دو بعدی که خصیصه‌های آن (شکل چهره، انحنای دهان، بلندی دماغ، اندازه چشمها و غیره) با اندازه‌گیری روی p متغیر تعیین می شوند، برای دسته‌بندی اولیه و دسته‌بندی نهایی حاصل از الگوریتمهای خوشه بندی به کار می‌رود. در چهره‌های چرنوف مقایسه به صورت چشمی انجام می‌شود. شباهت هر دو شیء می‌تواند با مقایسه تک تک اعضای صورت با یکدیگر انجام شود. مزیت دیگر چهره‌های چرنف سادگی درک آن است که نیازی به دانش خاصی ندارد.

در شکل (۲) چهره‌های چرنوف برای استانهای کشور ارایه شده است. شباهت چهره‌های چرنوف در واقع همان شباهت اشیاء در خوشه‌های مختلف در روش تحلیل خوشه‌ای است. همان طور که در شکل(۲) مشاهده می‌شود به عنوان مثال استانهای خوزستان، کرمان، گیلان و مازندران نسبت به سایر استانها از چهره مشابهی برخوردارند و بنابراین در یک دسته قرار می‌گیرند که این نتیجه با نتیجه تحلیل خوشه‌ای ارایه شده در جدول(۸) تا حد زیادی سازگاری دارد و به این ترتیب صحت نتایج تحلیل خوشه‌ای به تأیید می‌رسد.

۹. نتیجه‌گیری

روشهای آماری چند متغیره خوشه بندی برای دسته‌بندی استانهای کشور با توجه به متغیرهای حاصل از تحلیل و بررسی اطلاعات قربانیان تصادف بکار گرفته شد و نتایج با روش تحلیل چهره‌های چرنوف مقایسه گردید. بر اساس تحلیل ٦٤ متغیر برای هر یک از استانهای کشور که با ٦ روش سلسله مراتبی و غیرسلسله مراتبی و دو روش محاسبه فاصله انجام شد، نتایج بیشترین مشابهت‌ها در استانهای کشور که با روش چهره‌های چرنوف نیز مورد تأیید قرار گرفت، به شرح ذیل به دست آمد:

- خراسان و فارس؛
- گیلان، مازندران، خوزستان و کرمان؛
- آذربایجان شرقی، کرمانشاه و سیستان و بلوچستان؛
- اردبیل، چهارمحال و بختیاری ، ایلام و کهکلویه و بویراحمد؛
- یزد، زنجان و کردستان؛
- اصفهان؛
- سایر استانها.

استانهای قم، سمنان و بوشهر رفتار مرزی از خود نشان داده‌اند، به عبارت دیگر در خوشه‌های مختلف جابجا شده‌اند، ولی می‌توان شباهت آنها را با یکدیگر زیاد دانست. استان اصفهان نیز رفتار نسبتا مستقلی نسبت به سایر استانها نشان داده است.

خوشه بندی حاصل برای استانهای کشور را می توان برای ارایه راهکارهای مدیریتی جهت شناسایی گروههای همسان تصادفات و افزایش ایمنی بکار برد. همچنین بکاربردن آنالیز خوشه‌ای موجب کاهش هزینه‌های مدیریتی و ساده‌تر شدن برنامه‌های اجرایی می شود. در طرح‌های راهبردی ارتقای ایمنی، تعیین اهداف، راهبردها و اقدامات کاملاً یکسان برای کلیه استانها و مناطقی که دارای مشکلات و خصوصیات رفتاری متفاوت هستند، نمی‌تواند یک راهکار مدیریتی بهینه باشد، بلکه از نتایج تحلیل خوشه‌ای می‌توان اهداف، راهبردها و اقدامات را بر اساس خوشه‌های متفاوت در نظر گرفت و در نتیجه سریع‌تر و با هزینه کمتر به اهداف مورد نظر دست پیدا کرد.

۱۰. سپاسگزاری

شایسته است از سازمان پزشکی قانونی کشور و مسئولین و کارشناسان اداره آمار آن سازمان به ویژه آقایان سعداله مرادی و علی‌خادمی که هماهنگی جمع‌آوری‌و ارایه اطلاعات قربانیان‌ناشی از تصادفات کشور را انجام داده‌اند، تشکر و قدردانی به‌عمل آید.

۱۱. پانویس‌ها

تحلیل خوشه‌ای	Cluster analysis
روشهای سلسله مراتبی	Hierarchical methods
روشهای غیرسلسله مراتبی	Non Hierarchical methods
روشهای سلسله مراتبی متراکم	Agglomerative hierarchical methods
روشهای سلسله مراتبی تقسیم بندی شده	Divisive hierarchical methods
فاصله اقلیدسی	Euclidean distance
فاصله بلوک شهری	Manhattan distance
روشهای پیوندی	Linkage methods
چهره‌های چرنوف	Chernoff faces

Archive of SID
www.SID.ir

۱۱. مراجع

1- Koushki, P.A., Bustan, M.A. and Kartam, N. (2003) "Impact of safety belt use on road accident injury and injury type in Kuwait", Accident Analysis & Prevention Journal, Volume 35.March 2003, pp 237-241.

2- PIARC Technical Committee on Road Safety, C13 (2003) "Road safety manual", PIARC.

۳- سازمان پزشکی قانونی کشور(۱۳۸۳) " مجموعه آمار و اطلاعات قربانیان تصادف"، تهران، سازمان پزشکی قانونی کشور.

٤- آیتی، اسماعیل(۱۳۸۳) "برآورد هزینه تصادفات ترافیکی راههای ایران و نرم افزار مربوطه"، تهران، پژوهشکده حمل‌ونقل.

5-Kana, K., Atsumi, M. and Terushige, O. (2006) "Prediction of accident frequency for dangerous goods using cluster analysis", Sci Technol Energ Mater, Vol.67, No.1, 17-22.

6-Yuepeng, C., Yaping, Z., Yanli, M. and Yuqin, F. (2007) "Cluster analysis on evaluation indicators of driver characteristics", Intelligent Transportation Systems Conference, ITSC, IEEE.

7-Tracz, M. and Nowakowska, M. (1996) "Baysian theory and cluster analysis in the identification of road accident black spot", Transportation and Traffic Theory Congress, France, pp. 261-276.

8-Iakovakis, I. Dessypris, N. Dalamaga, M. and Petridou, E. (2003)"A cluster analysis of road traffic related childhood knee injuries, child: care", Health and Development, Vol.29, Number4, 297-301(5).

9- TRL (2004) "The innovatory analysis of road traffic accident data".

10- Postorino, M.N. and Sarnè, G.M.L. (2002) "Cluster analysis for road accidents investigation", Proceedings of Urban Transport, WIT.

۱۱- جانسون، ریچارد آ. و ویچرن، دین دبلیو. ، مترجم دکتر حسینعلی نیرومند(۱۳۷۹) "تحلیل آماری چند متغیری کاربردی"، مشهد: انتشارات دانشگاه فردوسی مشهد.

12- Data Analysis Products Division MathSoft, Inc (1999) "S-PLUS 2000, Guide to Statistics", Seattle, Washington.

۱۳- مانلی، بی.اف.جی، ترجمه محمد مقدم، سید ابوالقاسم محمدی شوطی، مصطفی آقایی سربرزه(۱۳۷۳) "آشنایی با روشهای آماری چندمتغیره"، تهران: انتشارات پیشتاز علم.

14- Everitt, B.S. and Dunn, G. (1991) "Applied multivariate data analysis", Edward Arnold, London.

15-Mojena, R. (1977) "Hierarchical grouping methods and stopping rules: an evaluation", computer journal, 20, pp.359-63.

16- Milligan, G.W. and cooper, M.C. (1985) "An examination of procedures for determining the number of clusters in a data set", Psychometrical, 50, pp.159-79.

17- Chernoff, H. (1973) "The use of faces to represent points in k-dimensional space graphically", Journal of the American Statistical Association, 68, No. 342, pp. 361-368.

Archive of SID

www.SID.ir