

بسمه تعالی

**گروه‌بندی خانوارها بر اساس شاخص‌های اقتصادی در سرشماری**

**عمومی نفوس و مسکن ۱۳۹۰**

## فهرست مطالب

### فصل اول: کلیات

- ۱-۱- مقدمه
- ۲-۱- مقدمه‌ای بر روش‌های گروه‌بندی و خوشه‌بندی
  - ۱-۲-۱ روش‌های توسعه‌ی انسانی
  - ۲-۲-۱ تحلیل مولفه‌های اصلی
  - ۴-۲-۱ روش‌های فازی
  - ۵-۲-۱ روش‌های تصمیم‌گیری چندشاخصه
  - ۶-۲-۱ الگوریتم‌های طبقه‌بندی و روش‌های یادگیرنده
- ۳-۱- تجربیات سایر کشورها در زمینه‌ی خوشه‌بندی
  - ۴-۱- معرفی داده‌های مورد استفاده
  - ۱-۴-۱ سرشماری عمومی نفوس و مسکن ۱۳۹۰
  - ۲-۴-۱ طرح آمارگیری هزینه و درآمد خانوار
  - ۳-۴-۱ طرح گردآوری اطلاعات اقتصادی-اجتماعی خانوار
  - ۵-۱- تعریف مفاهیم
  - ۶-۱- اهداف طرح
  - ۷-۱- چشم‌انداز فصل‌های آتی

### فصل دوم: پیش‌بینی درآمد خانوارها در سرشماری عمومی نفوس و مسکن ۱۳۹۰

- ۱-۲- مقدمه
- ۲-۲- روش‌های پیش‌بینی درآمد
  - ۱-۲-۲ معادله‌ی Mincer
  - ۲-۲-۲ جانپهی بادرنگ
  - ۳-۲-۲ جانپهی نزدیک‌ترین همسایه
- ۴-۲-۲ برازش مدل خطی تعمیم‌یافته بر روی داده‌های طرح آمارگیری هزینه و درآمد خانوار ۱۳۹۰
  - ۳-۲-۲ ارزیابی مدل منتخب
  - ۴-۲-۲ نتیجه‌گیری

## فصل سوم: خوشه‌بندی خانوارها

۱-۳ مقدمه

۲-۳ بررسی روش‌های خوشه‌بندی و انتخاب روش مناسب

۱-۲-۳ روش‌های خوشه‌بندی

۲-۲-۳ انتخاب روش خوشه‌بندی مناسب

۳-۳ کاربرد روش خوشه‌بندی K-Means بر روی داده‌های آزمایشی و ارزیابی آن

۴-۳ نتیجه‌گیری

## فصل چهارم: نتیجه‌گیری

۱-۴ مقدمه

۲-۴ ارزیابی روش‌های گروه‌بندی اقتصادی و انتخاب روش مناسب

۴-۳ کار کاربردی

۴-۴ آسیب‌شناسی

## فصل اول

### کلیات

#### ۱-۱ مقدمه

تصمیم‌گیری بهینه و برنامه‌ریزی در هر جامعه‌ای، مستلزم در دست داشتن آمار و اطلاعات دقیق است. اطلاعات در جهان امروز جزء سرمایه‌ی ملی محسوب می‌شود. بدون وجود اطلاعات دقیق و جامع، حرکت در راه تهیه‌ی برنامه‌های توسعه، کاری غیرممکن است. برنامه‌های توسعه‌ای که مبتنی بر آمارهای دقیق نباشند برای کشورها مفید نخواهند بود.

یکی از منابع اطلاعاتی دقیق و جامع، سرشماری عمومی نفوس و مسکن است. سرشماری عمومی نفوس و مسکن گسترده‌ترین و مهم‌ترین منبع اطلاعاتی کشور محسوب می‌شود. اما در سرشماری‌ها، پرسش‌های زیادی در مورد افراد هر خانوار، محل سکونت آن‌ها، امکانات و تسهیلات خانوار پرسیده شد که پاسخ این سوال‌ها مشخص‌کننده‌ی ویژگی‌های اجتماعی و اقتصادی خانوار است. اما از آنجایی که سرشماری بزرگترین آمارگیری کشور است تعداد داده‌های گردآوری‌شده زیاد و کارکردن با داده‌های خام به راحتی امکان‌پذیر نیست. لذا ضروری به نظر می‌رسد خانوارهای سرشماری‌شده بر اساس متغیرهای سرشماری به گروه‌هایی تقسیم شوند که هر کدام از این گروه‌ها نماینده الگوی اقتصادی مشخصی هستند. پس از تعیین گروه‌بندی، گروه‌های هدف از جمله بیماران را می‌توان به این گروه‌ها منتسب کرد تا بر اساس این گروه‌بندی، برای بیماران برنامه‌ریزی و سیاست‌گذاری کرد.

## ۱-۲ مقدمه‌ای بر روش‌های گروه‌بندی و خوشه‌بندی

از آنجایی که هدف از این پژوهش خوشه‌بندی خانوارها براساس وضعیت اقتصادی و رفاهی می‌باشد، در این بخش به دنبال ارائه روش‌هایی هستیم که براساس آن می‌توان خانوارهای کشور را با استفاده از داده‌های سرشماره عمومی نفوس و مسکن و به کمک سایر طرح‌های نمونه‌ای مرکز آمار ایران همچون طرح هزینه و درآمد خانوار طبقه‌بندی نمود. به عبارتی سوال مورد بحث نحوه استفاده از اطلاعات موجود در رتبه‌بندی خانوارها و یا به طور ساده‌تر خوشه‌بندی خانوارها می‌باشد. در نتیجه لازم است یا با استفاده از اطلاعات موجود در خصوص ویژگی‌های مختلف خانوار یک شاخص محاسبه و براساس آن خانوارها طبقه‌بندی شوند و یا با استفاده از روش‌های یادگیرنده و الگوریتم‌های موجود در مباحث طبقه‌بندی خانوارها را خوشه‌بندی نمود.

در این خصوص سوالاتی مطرح می‌شود که به فراخور هر فصل در این مطالعه پاسخ داده خواهند شد:

اول اینکه بهترین و کاربردی‌ترین روش به مجموع تجمیع و یا ترکیب ویژگی‌های مختلف چیست؟

دوم اینکه تفاوت و مزایای روش‌های بر مبنای الگوریتم و روش‌های بر مبنای شاخص‌های ترکیبی چه می‌باشد؟

سوم اینکه کدام ویژگی‌های خانوار بهترین نماینده‌های وضعیت خانوار می‌باشند و لازم است در تجمیع یا پیاده‌سازی

الگوریتم بکاربرده شوند؟

چهارم اینکه چگونه می توان از روش انتخابی برای تشخیص و تعیین طبقات استفاده نمود؟

پاسخ سوالات اول و دوم در این بخش و پاسخ سوالات سوم و چهارم در بخش های بعد ارائه خواهد شد. لذا در این بخش مبانی نظری و به عبارتی روش های موجود بر پایه مطالعات تجربی انجام شده در حوزه های مشابه یعنی مطالعات مربوط به رتبه بندی و طبقه بندی براساس ویژگی های مختلف استخراج شده است.

### ۱-۲-۱ شاخص های ترکیبی یا تجمیع شده

مسئله اساسی در این مطالعه خوشه بندی خانوارها با استفاده از ویژگی های مختلف خانوار در حوزه اقتصادی می باشد. یعنی استفاده از شاخصی ترکیبی و یا تجمیع شده به جای یک معیار در طبقه بندی خانوارها. ایده ترکیب شاخص ها بعد از ۱۹۹۰ و پس از انتشار گزارش توسعه انسانی محبوب الحق ارائه دهنده ایده شاخص توسعه انسانی بوده است. همانگونه که آمارتیا سن در سخنرانی خود با عنوان "یک دهه توسعه انسانی" در سال ۱۹۹۹ و در واقع یک دهه پس از انتشار گزارش توسعه انسانی این سوال را مطرح نموده است که: "چرا گزارش توسعه انسانی با این سرعت توانست در دنیایی که ده ها و شاید قرن ها طول می کشد تا یک ایده نو آنطور که شایسته است به رسمیت شناخته به سرعت مورد پذیرش قرار گرفت؟ چرا ایده توسعه انسانی اینچنین موفقیتی را در دوره معاصر بدست آورد؟" به عقیده وی گزارش توسعه در سطحی از درک اجتماعی مطالب قابل ارائه زیادی را برای افرادی که درک بالایی از این موضوعات دارند، ارائه داده است. به عقیده وی گزارش توسعه انسانی با سرعتی بیش از انتظار مورد پذیرش عموم واقع شد. یعنی بیش از آنچه که یک دهه پیش انتظار آن می رفت، مورد قبول واقع شد. در واقع باید این سوال مطرح شود که چرا چنین اتفاقی رخ داد؟ این سوال، منشا به وجود آمدن سوالات اساسی دیگری است. سوالاتی همچون؛ محاسبات توسعه انسانی در حقیقت چه کاری انجام می دهد؟ خصوصیات ویژه و ویژگی های شناخته شده آن چیست؟ در محاسبات توسعه انسانی به جای تمرکز بر برخی معیار های سنتی و حاشیه ای از پیشرفت اقتصادی (همچون تولید ناخالص داخلی سرانه)، برآوردهای سیستمی از متغیرهایی که در چگونگی زندگی انسان ها در هر اجتماعی دارای اهمیت است ارائه شده است. (از جمله وضعیت آموزشی و مراقبت های بهداشتی). لذا در ارزیابی توسعه تصور چندگانه ای از پیشرفت به وجود می آید. از آنجایی که زندگی انسان ها به صورت های گوناگون و با راه های مختلف تغییر شکل داده و دچار نقصان شده است، اولین وظیفه ای که از این منظر به وجود می

آید، تصدیق این امر است که محرومیت در ابعاد مختلف و به اشکال گوناگون باید در یک چارچوب عمومی و جامع در جوامع مختلف بررسی شود. چارچوب ارائه شده در خصوص ابعاد محرومیت لازم است متقاعد کننده و دارای روابط منطقی بوده و نباید از چندگانگی هایی که در طبیعت محرومیت ها وجود داشته و در جستار نامناسب معیارهای موفقیت و شکست یک فرد وجود دارد، چشم پوشی نمود.

به عقیده آمارتیا سن اقتصاددان برنده نوبل تعیین مفهوم اصلی ارزیابی اجتماعی و تشخیص آنچه باعث تبدیل توسعه انسانی به آنچه که هست و آنچه که - هر چند ناقص - در گزارش و تحلیل ها به آن دست یافته است شده است، در تعیین علت موفقیت آن بسیار حائز اهمیت می باشد. یکی از عوامل کلیدی موفقیت این گزارش ویژگی چندگانگی و آزادی نسبت به ارتباطات چندگانه می باشد. هدف از بیان این مطلب توسط آمارتیاسن استفاده از چند معیار توسط محبوب الحق در بیان توسعه به جای یک معیار بوده است. شاخص توسعه انسانی را محبوب الحق، اقتصاد دهن برجسته پاکستانی، بنیان نهاده است و پس از او، جامعه شناسان و اقتصاددانان برجسته ای چون آمارتیا سن، متفکر هندی در پیش برد و بسط آن نقش داشته اند. این مفهوم به عنوان راهبردی جهانی از طرف دفتر برنامه توسعه سازمان ملل متحد مطرح شده است که بر مبنای آن، توسعه انسانی، هدف نهایی توسعه اقتصادی است و البته بهترین وسیله ممکن برای پیشبرد توسعه نیز است. هدف از توسعه انسانی ارتقاء سطح زندگی انسان است و بر اساس رویکرد توسعه انسانی، نمی توان صرفاً سرانه تولید ملی را شاخص اصلی سطح توسعه دانست چراکه هدف آن است که قابلیت ها و توانایی های افراد به گونه ای رشد یابد تا زندگی انسان خلاقانه و همراه با رضایت باشد. به همین دلیل است که شاخص های عملیاتی شده در گزارش های توسعه انسانی سالانه دفتر برنامه توسعه سازمان ملل، شامل سه حوزه سلامت، آموزش و درآمد است که در مجموع، شاخص ترکیبی توسعه انسانی را می سازد و عددی بین صفر و یک را در بر می گیرد. هر چه این عدد به صفر نزدیک تر باشد، نشان از توسعه کمتر کشورها و هر چه به یک نزدیک تر باشد، نشان از توسعه بیشتر آنها دارد. در واقع محبوب الحق رویکرد جمع گرایی<sup>۱</sup> و یا تجمیع<sup>۲</sup> شاخص ها را در تدوین شاخص توسعه انسانی به کار گرفت. پس از آن بود که شاخص های ترکیبی برتری خود را نسبت به شاخص های برگرفته از یک متغیر به اثبات رساندند.

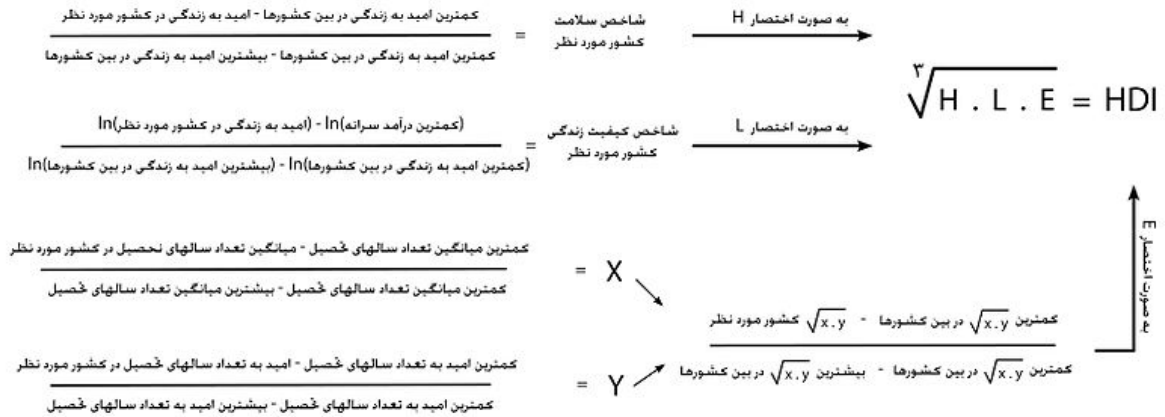
## ۱-۲-۲ شاخص توسعه انسانی

<sup>۱</sup> Plural concept  
<sup>۲</sup> aggregation

شاخص توسعه انسانی یک شاخص ترکیبی است و از ۴ شاخص امید به زندگی در هنگام تولد، درآمد سرانه کشور، میانگین تعداد سال‌های تحصیل کشور و امید به تعداد سال‌های تحصیل کشور تشکیل می‌شود؛ که به ترتیب معیارهای سلامت، سطح زندگی و آموزش را تشکیل می‌دهند. در محاسبه شاخص توسعه انسانی تا قبل از سال ۲۰۱۰ از یک تابع ضربی استفاده می‌شده است. همانطور که عنوان شد شاخص توسعه انسانی شاخصی کلی است که از ترکیب شاخص‌های جزئی بدست می‌آید، لذا بعد از نرمال نمودن متغیرها ابتدا لازم است که دو شاخص جزئی که برای موضوع آموزش در نظر گرفته شده است را با هم ترکیب و شاخص کلی آموزش را بدست آورد. همانگونه که ذکر شد شاخص توسعه انسانی در پی اندازه‌گیری سه ظرفیت اساسی «کسب دانش»، «دسترسی به امکانات مادی لازم برای یک زندگی بهتر» و «برخورداری از عمری طولانی توأم با سلامتی» است. این شاخص برای اندازه‌گیری ظرفیت اول از نرخ باسوادی بزرگسالان بر اساس دو سوم ارزش کلی و نرخ ترکیبی ثبت‌نام در مقاطع مختلف تحصیلی بر اساس یک سوم ارزش کلی و برای اندازه‌گیری ظرفیت دوم از قدرت خرید واقعی افراد که بر اساس دلار تعدیل شده و برای اندازه‌گیری ظرفیت سوم از امید به زندگی در بدو تولد استفاده می‌کند. شاخص توسعه انسانی میانگین وزنی ساده از نماگرهای یاد شده بوده است. مقدار شاخص مذکور بین یک حداقل (صفر) و یک حداکثر (یک) قرار می‌گیرد. کشورهای دارای شاخص توسعه انسانی پایین‌تر از ۰/۵۰۰ به عنوان کشورهای با توسعه انسانی پایین و کشورهای واقع در فاصله ۰/۵۰۰ و ۰/۷۹۹ کشورهای با توسعه انسانی متوسط و کشورهای با شاخص ۰/۸۰۰ تا یک به عنوان کشورهای با توسعه انسانی بالا محسوب می‌شوند. برای محاسبه حداقل و حداکثر شاخص توسعه انسانی، حداقل‌ها و حداکثرها برای سه عنصر شاخص مذکور در نظر گرفته می‌شدند و آن‌ها، حداقل‌هایی بودند که در طول ۳۰ سال گذشته مشاهده شده و حداکثرهایی هستند که می‌توان در طول ۳۰ سال آینده تصور کرد. حداقل برای باسوادی بزرگسالان و نرخ ترکیبی ورود به مقاطع تحصیلی صفر و حداکثر برای آن‌ها ۱۰۰ درصد است. حداقل برای سلامتی، امید به زندگی ۸۵ سال است. حداقل سطح درآمد سرانه ۱۰۰ دلار و حداکثر برای آن ۴۰۰۰۰ دلار است.

بعد از سال ۲۰۱۰ به جای میانگین وزنی ساده از میانگین هندسی برای تجمیع شاخص‌ها استفاده شده است که نحوه محاسبه به صورت زیر می‌باشد:





آنچه در خصوص شاخص توسعه انسانی ارائه شد نشان می دهد وضعیت اقتصادی خانوار نیز همچون توسعه را نمی توان با متغیر های مالی خانوار همچون درآمد و یا هزینه تعیین نمود و در نظر گرفتن این متغیرها به تنهایی نشان دهنده وضعیت واقعی خانوارهای کشور در نقایسه با یکدیگر نمی باشد. لذا علت طرح مبانی نظری شاخص توسعه انسانی بهره گیری از ایده مورد استفاده در تجمیع شاخص ها به منظور محاسبه شاخص توسعه انسانی می باشد. البته تفاوت هایی در شاخص توسعه انسانی با هدف این مطالعه دارد. به عنوان مثال، عموماً شاخص توسعه انسانی برای گروهی از افراد در سطح بین المللی، ملی، استانی و حتی منطقه ای برآورد شده و مورد بررسی قرار داده می شود و به راحتی براساس میزان شاخص میتوان کشورها و یا مناطق را رتبه بندی یا خوشه بندی نمود. اما هدف در این مطالعه خوشه بندی خانوارهاست و لازم است ویژگی های هر یک خانوارها به طور مجزا در نظر گرفته شود و با استفاده از روش شناسی محاسبه میانگین هندسی شاخص ها، یک شاخص ترکیبی محاسبه و براساس آن خانوارها رتبه بندی و یا خوشه بندی شوند. لازم به ذکر است همچون روش های مورد استفاده در تمام مطالعات تجربی این روش نیز مزایا و معایبی به دنبال دارد. اما اولین مسئله قابل بحث به عبارتی ایراد در استفاده از این روش وزن های مورد استفاده در محاسبه میانگین حسابی (روش قدیمی) و یا میانگین هندسی است. زیرا این وزن ها نشان دهنده اهمیت هر یک از ویژگی های خانوار در مقایسه با سایر ویژگی های می باشد. به عنوان مثال مقایسه تعداد اتاق ها و نوع سوخت مورد استفاده در پخت و پز، کدام یک اهمیت بیشتری در وضعیت رفاهی خانوار دارد؟ اهمیت هر یک به چه نسبت است. البته روش هایی نیز وجود دارد که در آن وزن ها براساس روش به طور خودکار و براساس داده ها استخراج می شود که در ادامه توضیح داده خواهد شد. ثانیاً در محاسبه این شاخص لازم است ویژگی های

قابل اندازه گیری خانوارها مورد استفاده قرار داده شود که می تواند مقادیر عددی واقعی اختیار کند، به عبارتی در این روش لازم است از متغیرهای کمی استفاده نمود. اما اغلب اطلاعات موجود در طرح هزینه و درآمد خانوار کدگذاری شده است به عبارتی اطلاعات کیفی است که تبدیل به مقادیر کمی شده است. لذا لازم است به منظور استفاده از این روش شاخص سازی های صورت گیرد، که گاه تنها موجب پیچیدگی محاسبات و دور شدن از هدف اصلی مطالعه می شود. اما مزیت این روش سادگی محاسبات آن می باشد. همچنین تعیین وزن ها اگر با بررسی های کارشناسانه دقیق و براساس واقعیت های موجود صورت گیرد احتمالاً می تواند به واقعیت نزدیک تر باشد تا اینکه وزن ها صرفاً براساس برنامه ریزی های ریاضی و یا آماری صورت پذیرد.

### ۱-۲-۳ تحلیل مولفه های اصلی ۳

فیلمر و پریچت<sup>۴</sup> در سال ۱۹۹۸ در مطالعه ای به منظور طبقه بندی خانوارها از نظر ثروت بدون داشتن هزینه های خانوار از روش تحلیل مولفه های اصلی استفاده نموده اند. آنها به این منظور از سه دسته متغیر استفاده نموده اند. ۸ متغیر مرتبط با مالکیت کالاهای مصرفی بادوام خانوار (همچون ساعت مچی یا دیواری، دوچرخه، رادیو، تلویزیون، چرخ خیاطی و یخچال فریزر) گروه دوم ۱۲ متغیر می باشد که توصیف کننده ویژگی های منزل مسکونی خانوار است که سه مورد در خصوص سرویس های بهداشتی مورد استفاده خانوار، سه مورد در خصوص آب آشامیدنی خانوار، دو مورد در خصوص اتاق ها، دو مورد مصالح ساختمانی، یک مورد در خصوص منبع اصلی نور و نهایتاً منبع انرژی مورد استفاده برای پخت و پز. گروه سوم شامل یک متغیر در خصوص متراژ زمین در اختیار خانوار که آیا از ۶ هکتار بیشتر هست یا نه. در مطالعه آنها از این متغیرها به منظور به دست آوردن شاخص دارایی های خانوار که تقریبی از "ثروت" خانوار است با استفاده از روش تحلیل عوامل اصلی به کار برده شده است.

لیور و همکاران (۲۰۰۹)<sup>۵</sup> نیز از این روش برای محاسبه شاخص ترکیبی کیفیت زندگی استفاده نموده اند. در محاسبات آنها شاخص های به دو دسته جغرافیایی و غیر جغرافیایی دسته بندی شده اند.

<sup>۴</sup> Principle component analysis

<sup>۵</sup> Filmer, Deon and Pritchett, "Estimating Wealth Effects without Expenditure Data -- or Tears: An Application to Educational Enrollments in States of India", September 1, 1998  
Joshua J., Gail Pacheco and Stephanié Rossouw. "Do Non-Economic Quality of Life Factors Drive. Lewer<sup>o</sup>  
IZA DP No. 4385, August 2009 Immigration?".

روش تحلیل مولفه های اصلی یکی از روش های ترکیب و یا تجمیع چند متغیر و تبدیل آن به یک متغیر می باشد که عموماً در رتبه بندی از آن بهره برداری می شود. یکی از مشکلات در اغلب روش های تجمیع وزن های مورد نظر هر یک از متغیرها و به عبارتی میزان اهمیت هر یک از متغیرها است. در ادبیات موضوعی سه روش برای این مسئله طرح شده است. اول استفاده از وزن های برابر برای همه متغیرها که البته انتخاب این راه نمی تواند منعکس کننده همه واقعیات همانگونه که هست باشد. راه دوم در نظرگرفتن مجموعه ای از وزن هاست. به عنوان مثال قیمت دارایی های مختلف را می توان به عنوان شاخصی از ثروت خانوار در نظرگرفت، البته استفاده از این روش تنها زمانی ممکن است که قیمت انواع دارایی ها در دسترس باشد. راه حل سوم این است که متغیرها تجمیع نشود بلکه از یک مدل رگرسیون ساده استفاده نموده و تمامی متغیرها را در یک مدل وارد نموده با کنترل متغیرهای موثر به جز ثروت اثر ثروت را بدست آورد. این روشی است که مونتگمری و همکارانش<sup>۶</sup> در سال ۱۹۹۷ در مطالعه ای در خصوص زاد و ولد و مرگ و میر استفاده نموده اند. اما در روش تحلیل مولفه های اصلی دلیلی برای نگرانی در خصوص وزن ها وجود ندارد.

تحلیل مولفه های اصلی یکی از روش های تحلیل عاملی<sup>۷</sup> بوده و در واقع تبدیلی در فضای برداری است، که غالباً برای کاهش ابعاد مجموعه ی داده ها مورد استفاده قرار می گیرد. تحلیل مولفه های اصلی در سال ۱۹۰۱ توسط کارل پیرسون<sup>۸</sup> ارائه شد. این تحلیل شامل تجزیه مقادیرهای ویژه ماتریس کواریانس می باشد. تحلیل مولفه های اصلی در تعریف ریاضی یک تبدیل خطی متعامد است که داده را به دستگاه مختصات جدید می برد به طوری که بزرگترین واریانس داده بر روی اولین محور مختصات، دومین بزرگترین واریانس بر روی دومین محور مختصات قرار می گیرد و همین طور برای بقیه. تحلیل مولفه های اصلی می تواند برای کاهش ابعاد داده مورد استفاده قرار بگیرد، به این ترتیب مولفه هایی از مجموعه داده را که بیشترین تاثیر در واریانس را دارند حفظ می کند.

---

“Measuring Living Standards”<sup>۱</sup> Montgomery, Mark, Kathleen Burke, Edmundo Paredes, and Salman Zaidi, 1997. Division, The Population Council. New York, NY. with DHS Data: Any Reason to Worry?” mimeo, Research Factor Analysis<sup>۷</sup>

<sup>۸</sup> Pearson, K. (1901). ["On Lines and Planes of Closest Fit to Systems of Points in Space"](#). Philosophical Magazine 2 (6): 559–572.

با در نظر گرفتن  $\underline{X} = (X_1, X_2, \dots, X_p)^T$  یک بردار تصادفی با ماتریس کواریانس معین نامنفی  $\Sigma$  و  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  مقادیر ویژه  $\Sigma$  باشد. همچنین بردارهای ویژه  $a_1, a_2, \dots, a_p$  یک متعامد به ترتیب متناظر با  $\lambda_1, \lambda_2, \dots, \lambda_p$ ، متغیرهای  $Y_1, Y_2, \dots, Y_p$  تعریف شده در زیر را مولفه های اصلی گویند.

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$$

$$Y_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p$$

$$Y_p = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p$$

را مولفه اصلی  $Y_i$  نام گویند. بردار  $\underline{Y} = (Y_1, Y_2, \dots, Y_p)^T$ ، بردار مولفه های اصلی نام دارد.

در تحلیل مولفه های اصلی پس جمع آوری داده، داده ها تنظیم و تبدیل به داده هایی با میانگین صفر می شود. در مرحله بعد ماتریس کواریانس محاسبه و مقادیر ویژه و بردارهای ویژه ماتریس کواریانس به دست می آید. پس از آن بردار ویژگی ایجاد و انتخاب مولفه های اصلی به منظور کاهش بعد انتخاب می شود. نهایتاً داده های جدید بر اساس اطلاعات قبلی به دست می آید.

به بیان ساده و با توجه به هدف از تحقیق حاضر تابع زیر را در نظر بگیرید:

$$W_i = \sum_{k=1}^m \Phi_k x_{k,i} \quad i = 1, \dots, n.$$

که در آن  $W_i$  شاخص نشان دهنده وضعیت رفاهی خانوار  $i$  و  $x$  ویژگی  $k$  ام خانوار  $i$  ام می باشد.

از آنجایی که وزن هر یک از شاخص ها براساس شرایط ابتدایی جوامع قابل تعیین است لذا در اینجا  $W_i$  با انتخاب  $\Phi_k$  به عنوان مقادیر ماکزیمم کننده واریانس  $W_i$  با استفاده از شرایط استاندارد به دست می آید. به عبارتی  $\Phi_k$  با استفاده از روش تحلیل اجزای اصلی محاسبه شده و  $W_i$  جزء اصلی اول می باشد.  $W_i$  به عنوان معیار اندازه گیری رفاه اقتصادی به دست خواهد آمد.

مزایای این روش عبارتند از سادگی در کاربرد، عدم نیاز به تعیین وزن هر یک از شاخص های مورد استفاده در تعیین وضعیت اقتصادی خانوار و قابلیت به کارگیری متغیرهای کمی همراه با متغیرهای کیفی. همچنین تحلیل مولفه های اصلی

یکی از انواع روشهای تحلیل داده های چند متغیره است که هدف اصلی آن تقلیل بعد مساله مورد مطالعه است. یکی از کاربردهای مهم تحلیل مولفه های اصلی، در رگرسیون است. با استفاده از تحلیل مولفه های اصلی می توان تعداد زیادی متغیر توضیحی (متغیر مستقل) همبسته را با تعداد محدودی متغیر توضیحی جدید که مولفه های اصلی نامیده می شوند و ناهمبسته اند، جایگزین نمود. به این ترتیب نه تنها بعد مساله تقلیل می یابد بلکه مساله چند همخطی پیش نمی آید. این روش از تجمیع متغیرها محدودیت هایی را نیز به دنبال دارد که این محدودیت ها ناشی از فروض ساده کننده ای است که در این روش همچون فرض خطی بودن، فرض قابلیت اتکای میانگین و کواریانس از نظر احتمالاتی و نهایتاً این فرض که واریانس شاخصه اصلی داده هاست.

لازم به ذکر است، ایده استفاده از شاخص های جغرافیایی منطقه مورد سکونت خانوار در خوشه بندی خانوارها بر اساس اهداف پژوهش کاربردی و قابل بحث و بررسی است. لذا استفاده از شاخص های نماینده وضعیت جغرافیایی محل سکونت خانوار در بین ویژگی های مورد بررسی خانوارها در هر خوشه به استخراج نتایج صحیح و نزدیک تر به واقعیت کمک خواهد کرد.

## ۱-۲-۴ روش های فازی

یکی از اولین کسانی که از تئوری فازی برای ساختن سنجه های رفاه در ابعاد کلان استفاده نمود، بالیامون<sup>۹</sup> می باشد. وی با استفاده از داده های آموزش، طول عمر و حقوق مدنی و سیاسی به رتبه بندی شاخص های مختلف رفاه برای ۴۸ کشور جهان پرداخت. بالیامون برای به دست آوردن رفاه بین کشورها که شامل قوانین، اهداف و وارد کردن ایده حدود بالایی و پایینی است، از تابع عضویت غیرخطی استفاده کرده است. در انتها با توجه به گذر تدریجی در روش فازی و قابلیت تعریف در روش فازی، اندازه گیری رفاه را به کمک روش فازی پیشنهاد داده است.

در مطالعه ای با عنوان «دست یابی به رفاه فازی در آسیای حوزه اقیانوس آرام» که توسط بالیامون لوتز و مک گیلیوری<sup>۱۰</sup> در سال ۲۰۰۶ ارائه شد، آن ها پس از ساختن مدل برای رتبه بندی فازی مؤلفه های شاخص توسعه انسانی، به برآورد

<sup>۹</sup> Balamoune-Lutz, M. (2004), On the Measurement of Human Well-Being: Fuzzy Set Theory and Sen's Capability Approach, WIDER Research Paper No. 2004/16, World Institute for Development Economics Research, United Nations University, Helsinki.

<sup>۱۰</sup> Balamoune-Lutz, M., Mc Gillivray, (2006), Fuzzy Well-being Achievement in Pacific Asia, Journal of Asia Pacific Economy, 11: 2, PP: 168-177

شاخص توسعه انسانی برای ۱۴ کشور حوزه اقیانوس آرام پرداختند. نتایج نشان داد که تغییراتی در رتبه کشورهای در جهان با استفاده از روش رتبه بندی فازی نسبت به روش سازمان ملل وجود دارد. با توجه به مفاهیم مبهم در تعاریف و محتوای مؤلفه های شاخص توسعه انسانی و با توجه به گذر تدریجی، روش رتبه بندی فازی را در مشخص نمودن شاخص توسعه انسانی پیشنهاد نمودند. ارزیابی دیگری نیز در مقاله ای تحت عنوان «تعیین چند بعدی رفاه بر اساس نظر سن» توسط انریکا چیاپرو<sup>۱۱</sup> در سال ۲۰۰۰ ارائه گردید. وی با تأکید بر دیدگاه آمارتیا سن به عنوان یکی از نظریات مهم در زمینه رفاه، به محاسبه رفاه با روش فازی در کشور ایتالیا پرداخت و علاوه بر مقایسه با تحقیقات تجربی، به مزیت های این روش در تبیین رفاه پرداخت.

والریه برنگر و اودری وردیر چوچان<sup>۱۲</sup> در مقاله ای تحت عنوان «معیارهای چندبعدی رفاه: استاندارد زندگی و کیفیت زندگی» به ایجاد دو شاخص فوق با روش تحلیل کلی فازی برای ۱۷۰ کشور دنیا با توجه به نظریات سن در سال ۲۰۰۷ پرداختند و نتایج را با شاخص توسعه انسانی و درآمد سرانه مقایسه نمودند. شاخص مورد استفاده برای استاندارد زندگی و کیفیت زندگی در تحقیق آنان خود شامل ۹ متغیر می باشد که در سه دسته استاندارد تحصیلات، استاندارد سلامت و وسایل رفاهی (کیفیت محیط اطراف) طبقه بندی شده اند منطق فازی به تازگی توانسته در موضوعاتی مثل فقر و رفاه وارد گردد. در مقایسه این دو موضوع نیز فقر پیشینه بیش تری نسبت به رفاه دارد.

### – نظریه مجموعه های فازی

تئوری مجموعه های فازی در سال ۱۹۶۵ توسط پروفیسور لطفی زاده<sup>۱۳</sup> در دانشگاه برکلی با مقاله ای تحت عنوان «مجموعه های فازی» که در مجله اطلاعات و کنترل به چاپ رساند، متولد شد. در سال ۱۹۷۴ که نقطه عطف منطق فازی بود، ابراهیم ممدانی از دانشگاه لندن برای نخستین بار منطق فازی را در زمینه کنترل یک موتور بخار ساده به کار گرفت. از آن زمان به بعد این علم شاهد گسترش روزافزون جنبه های تئوری و عملی آن توسط دانشمندان علوم مختلف بوده است. مهم ترین ویژگی منطق فازی در مقایسه با منطق کلاسیک این است که دانش و تجربه بشر را می

<sup>۱۱</sup> Chiappero Martineti Enrica, (2000), A Multidimensional assessment of well-being based on sen's functioning approach, mimeo, University of Pavia, Italy

<sup>۱۲</sup> Be'rnger, Verdier- Chouchane, (2007), Multidimensional Measures of Well-Being: Standard of Living and Quality of Life Across Countries, World Development, Vol. 35, No. 7, pp: 1259 –1276.

<sup>۱۳</sup> Zadeh, L. A. (1965) Fuzzy sets, Information and Control, 8, pp. 338–343

تواند در قالب روابط ریاضی بیان نماید. این مسئله باعث شده که مسائل موجود در دنیای واقعی را به خوبی بتوان با استفاده از آن مدل سازی نمود.

در مسائل اقتصادی مباحث بسیاری مانند عدالت، اقتصاد زیرزمینی، فقر، رفاه و ... نیز به دلیل عدم قطعیت، عدم صراحت و چندبعدی بودن شامل پیچیدگی زیاد هستند و می توان از منطق فازی استفاده نمود.

در روش رتبه بندی فازی مفهوم می توان مجموعه های فازی را توسط تابع عضویت پیوسته تعریف نمود که بر اساس آن، مقادیر ورودی، درجات مختلفی را کسب می نمایند. درجات عضویت توسط تابع عضویت در بازه  $[0,1]$  مشخص می شود. تغییر تدریجی این مجموعه ها شرایطی را به وجود می آورند که درجه امکان یک مجموعه به آرامی از یک تابع عضویت کامل به تابع عدم عضویت یا برعکس تغییر حالت دهد. برای مثال میزان تحصیلات یا آموزش، موضوعاتی هستند که به تدریج تغییر حالت می دهند، اما هیچ گاه مقادیر صفر یا یک را اکتساب نمی نمایند. به عنوان مثال برای محاسبه یک شاخص چند بعدی با عنوان  $\gamma$  استفاده از روش فازی می توان به روش زیر عمل نمود.

ابتدا درجه توابع  $\mu_h$  و  $\mu_l$  را انتخاب می کنیم. بدین منظور  $\mu_h$  را بالاترین درجه عضویت اکتساب شده توسط مؤلفه  $X_h$  و  $\mu_l$  را پایین ترین درجه عضویت کسب شده توسط مؤلفه  $X_l$  در نظر می گیریم. در ادامه، مقادیر  $\alpha$  و  $\beta$  مورد استفاده را برای نمایندگان سه مؤلفه شاخص  $\gamma$  طبق روابط (۱) و (۲) محاسبه می شود.

نهایت با توجه به تابع عضویت زیر به برآورد شاخص مؤلفه مربوطه برای هر مقطع (کشور، استان و یا خانوار) می پردازیم: