

روشی برای طبقه‌بندی خانوارهای شهری ایران براساس مؤلفه‌های اقتصادی و اجتماعی خانوار

جمشید پژویان*

علی‌اکبر خسروی‌نژاد**

این مقاله با توجه به تفاوت رفتارهای مصرفی خانوارها؛ با استفاده از روش طبقه‌بندی پیشنهاد شده توسط «جنسن و مانریکو»، برای طبقه‌بندی خانوارهای ایرانی براساس ویژگیهای اقتصادی- اجتماعی آنها در نظر گرفته شده است و این شیوه پیشنهادی برای طبقه‌بندی خانوارهای شهری ایران، اجرا نیز شده است. نتایج حاصله بیانگر طبقه‌بندی خانوارهای شهری در پنج گروه مستقل و متمایز است.

کلید واژه‌ها:

طبقه‌بندی، خانوارها، اقتصادسنجی، مؤلفه‌های اقتصادی، مؤلفه‌های اجتماعی درآمد، مخارج خوراکها

* دکتر جمشید پژویان؛ عضو هیأت علمی دانشگاه علامه طباطبائی.

** علی‌اکبر خسروی‌نژاد؛ عضو هیأت علمی دانشگاه آزاد اسلامی واحد تهران مرکزی.

مقدمه

در بسیاری از کشورها و از جمله کشور ما آزادسازی بخش کشاورزی یکی از موضوعات مورد توجه است. مسائل مربوط به بودجه، عدم تعادلهای اقتصاد کلان و هزینه بالای تأمین برنامه‌های آزادسازی همراه با الحاق به سازمان جهانی تجارت و یا ارائه درخواست الحاق به سازمان مذکور، از دلایل اصلی و مهم برای تغییر در سیاستها قلمداد می‌شوند. بدیهی است اجرای این اصلاحات سیاستی منجر به تغییر در بازارها شده و از همه مهمتر باعث تعدیل در قیمت مواد غذایی خواهد شد.

به هر حال، چنین تغییرات قیمتی اثرات نهایی بر روی رفاه مصرف‌کنندگان باقی می‌گذارد. واقعیت‌های موجود بر این امر دلالت دارند که رفتار مصرفی برای گروه‌های مختلف مصرف‌کننده به هنگام تغییرات قیمت متفاوت است. بنابراین تحت این شرایط، هرگونه تحلیل در سطح کلی (در سطح بالایی از هم‌فرونی^۱) چندان مفید نبوده و ممکن است که علائم اشتباهی را به سیاست‌گذاران ارائه دهد.

لحاظ کردن اثرات توزیع درآمد در تحلیلها از جمله؛ تحلیل‌های تقاضا و شاخص‌سازی، بسیار مشکل است. بعضی از محققین غالباً از داده‌های مخارج متوسط به‌عنوان نماینده سطوح مختلف درآمدی استفاده می‌کنند و فرض می‌کنند که خطای اندازه‌گیری کوچک است، این خطاها تنها زمانی حداقل می‌شود که توزیع مخارج و متغیرهای جمعیتی یکسان باشد.^۲

استدلالاتی فوق دال بر آن است که به هنگام ارزیابی و اندازه‌گیری اثرات رفاهی ناشی از سیاست‌گذاریهی تعدیل قیمت‌ها در اثر اجرای سیاست‌های آزادسازی، نیازمند آن هستیم که خانوارها را به‌عنوان گروه‌های اصلی مخاطبین این سیاستها طبقه‌بندی کنیم.

خانوارهای دارای رفتار مصرفی مشابه (یکسان) لازم است که در یک طبقه یا

-
1. Level of Aggregation
 2. Deaton, A. and Mullbauer, J., (1980).

گروه قرار گیرند. روش و متدولوژی این مقاله در طبقه بندی خانوارها مبتنی بر تشابه رفتار مصرفی است که در قالب تابع انگل تجلی می یابد. در این مطالعه، طبقه بندی خانوار در گروه های مستقل و متمایز براساس همسانی در رفتار مصرفی خانوارها با استفاده از آزمون واریانس همسانی برای پسماندهای حاصل از برآزش منحنیهای انگل صورت گرفته است.

پس از مقدمه مروری بر ادبیات طبقه بندی خانوارها را ملاحظه خواهید کرد. سپس در قسمت سوم طبقه بندی خانوار را که شامل روش و مدل است، نوع داده های آماری مورد استفاده و نیز نتایج تجربی خواهیم داشت. همچنین ویژگیهای اقتصادی- اجتماعی در طبقات مختلف نیز در قسمت چهارم مقاله آمده و در پایان، مقاله با جمع بندی و نتیجه گیری به اتمام خواهد رسید.

مروری بر ادبیات طبقه بندی خانوارها

مطالعات معدودی در مورد برآورد سیستمها و کششهای تقاضا برای ارائه تصویری از رفتار مصرف کنندگان و سیاستگذاری در گروه های مختلف خانوارها، در جهان انجام شده است. برای مثال؛ می توان به مطالعات جارکو^۱؛ ساوادوگو و برانت^۲؛ تکلو و جانسون^۳؛ برونو و اکمل^۴؛ جونز و ماستیفل^۵؛ جنسن و مانریکو^۶ اشاره کرد. به غیر از مطالعات جارکو، جنسن و مانریکو، بیشتر این مطالعات یک روش مشخص، رسمی و مبتنی بر پایه های نظری اقتصاد خرد را برای طبقه بندی خانوارها دنبال نکرده است؛ بلکه از طبقه بندی از قبل تعریف شده ای براساس تعاریف مراجع رسمی جمع آوری اطلاعات و آمار استفاده کرده اند. این نوع طبقه بندیها، یک طبقه بندی تک موردی می باشد و ویژه یک امر خاص^۷ تلقی می گردد، بنابراین می توان آن را

1. Jarque (1987).

3. Teklu and Jhson (1988).

5. Jones and Mustiful (1987).

7. ad hoc

2. Savadogo and Brandt (1988).

4. Burney and Akmal (1991).

6. Jensen and Manrique (1988).

طبقه‌بندی برونزا^۱ نامگذاری کرد. در ادامه بحث، ابتدا روش جارکو و سپس روش جنسن و مانریکو ارائه شده و پس از آن نیز محدودیتها یا برتری هریک از این روشها، جهت انتخاب روش مناسب طبقه‌بندی خانوارها آمده‌است.

جارکو در مقاله خود به تحلیل رفتار مصرف کنندگان مکزیکی در قالب طبقه بندی خانوارها در گروههای مختلف پرداخته است. در مطالعه مذکور که با استفاده از داده‌های مقطعی درآمد و مخارج خانوار سال ۱۹۷۵ - شامل ۴۹۴ خانوار - صورت گرفته‌است، در کنار متغیرهای هزینه‌ای خانوارها به متغیرهایی نظیر: سن و شغل سرپرست خانوار و اندازه خانوار و یا ترکیبهای مختلفی از این متغیرهای اقتصادی اجتماعی توجه شده است. برای تشریح روش جارکو مدل سیستم مخارج خطی تعمیم یافته^۲ زیر را در نظر بگیرید:

$$p_j q_j = p_j Q_j + \theta_j \mu (x^* - p^1 Q) \quad (1)$$

$$s = (1 - \mu)(x^* - p^1 Q) \quad j = 1, \dots, n$$

که در آن $p_j, q_j, Q_j, \theta_j, \mu, \mu, x^*$ و s به ترتیب قیمت کالای زام، میزان مصرف کالای زام، حداقل مخارج الزامی^۳ کالای زام، سهم نهایی بودجه‌ای و کالای زام، میل نهایی به مصرف در ارتباط با درآمد و درآمد جاری قابل تصرف و نیز پس انداز می‌باشد. پس از یک سری عملیات جبری، برای ساده شدن مدل از تعاریفی استفاده گردیده و مدل زیر نیز معرفی شده‌است:

$$y_{ij} = x_i^1 \beta_j + \varepsilon_{ij} \quad (2)$$

$$\beta_j = (\beta_{1j}, \beta_{2j}) , y_{(n+1)} = S_i \quad J = 1, \dots, n+1 : \quad i = 1, \dots, m$$

1. Exogenous
2. Extend Linear Expenditure system=ELES
3. Substance

که در آن x_i بردار متغیرهای برونزای قیمت و درآمد، m تعداد خانوارها در گروه درآمدی و ε_i جمله اختلال مربوط به خانوار i ام در مصرف کالای i ام می‌باشد. معادله (۲) را در نظر بگیرید، این معادله برای سادگی، موقتاً اندیس i را که مربوط به کالاهاست حذف می‌نماید. همچنین فرض می‌شود که بردار پارامترهای β میان خانوارها متفاوت است، برای این منظور از اندیس i در β استفاده می‌کنیم که می‌توان نوشت:

$$y_i = x_i' \beta_i + \varepsilon_i \quad (3)$$

معادله (۳) دارای $2N$ پارامتر می‌باشد و این در حالیست که تنها N مشاهده برای برآورد آماری معادله (۳) در اختیار است، بنابراین بمنظور بدست آوردن مدلی قابل تخمین می‌بایست فروضی بر روی نحوه شکل‌گیری β_i وضع شود. چارکو با استفاده از متغیرهایی مانند سن و شغل سرپرست خانوار و سپس خانوار، برای توضیح β_i ها استفاده کرده‌است.

همچنین، وی فرض می‌کند که بردار β_i از رابطه $\beta_i = \gamma_0 + \gamma_1 z_i + u_i$ پیروی می‌کند. که در آن γ_0 و γ_1 به ترتیب بردار و ماتریس ضرایب، z_i بردار ویژگیهای اجتماعی و جمله اختلال می‌باشد. با جایگذاری رابطه β_i ها در معادله (۳)، معادله‌ای برحسب γ_0 و γ_1 بدست می‌آید که می‌توان آن را برآورد نمود و پس از برآورد، β_i های برآورد شده را برای کلیه خانوارها در یک ماتریس (ماتریس B) قرار داد. همچنین خانوارهایی که دارای β_i های یکسانی هستند، باید در کنار هم و در یک طبقه قرار گیرند؛ یعنی طبقه‌بندی در یک فضای k بعدی ($k=2n$) صورت بگیرد.

معیارهای مختلفی برای گروه‌بندی N واحد آماری نسبت به K متغیر وجود دارد. چارکو (۱۹۷۸) نشان داده‌است که معیار مناسب برای طبقه‌بندی حداکثرسازی

«قدرت توضیح‌دهندگی نسبی سراسری»^۱ توسط R^2 نشان داده می‌شود. R^2 عبارتست از اثر مجموع مجذور β های درون هر گروه که براساس واریانس نمونه استاندارد شده‌اند.

برای تعیین تعداد طبقات، ابتدا به متغیر L یا تعداد طبقات، مقادیر ۲، ۳، ۴ تا ۲۰ داده شده‌است. سپس معیار «قدرت توضیح‌دهندگی نسبی سراسری» یا R^2 برای حالت‌های وجود دو طبقه ($L=2$)، سه طبقه ($L=3$) تا بیست طبقه ($L=20$) برآورد و محاسبه شده‌است. برای بدست آوردن حداکثر مقدار R^2 در سطح متناظری از تعداد طبقات (L) از الگوریتم اسپارک (۱۹۷۳) استفاده شده‌است. نتایج نشان می‌دهد که در تعداد طبقات برابر چهارده ($L=14$) مقدار R^2 حداکثر شده‌است.

اگرچه روش طبقه‌بندی جارکو دارای مبانی نظری خوبی می‌باشد، ولی کاربرد عملی آن خصوصاً به هنگام سیاستگذاری محدود است. روش مذکور نیازمند اطلاعات بسیار زیاد بوده و طبقه‌بندی وی منجر به تعداد زیادی طبقات مختلف خانوارها بسته به ویژگیهای اقتصادی - اجتماعی آنها می‌گردد. افزون بر موارد فوق، در بهره‌گیری از سیستم مخارج خطی تعمیم یافته از لحاظ نظری، قیودی چون هموتیک تابع مطلوبیت استخراجی از آن مستتر است که طبقه‌بندی را با مشکل مواجه می‌سازد.

جنسن و مانریکو در مطالعه خود ضمن طبقه‌بندی خانوارها به برآورد تقاضا برای کالاهای خوراکی در گروه‌های مختلف خانوارهای اندونزیایی پرداخته‌اند. داده‌های مورد استفاده آنها بررسی اقتصادی و اجتماعی ملی^۲ کشور اندونزی برای سالهای ۱۹۸۱، ۱۹۸۴ و ۱۹۸۷ بوده‌است؛ بگونه‌ای که در مطالعه آنها، فقط از مشاهدات متعلق به مناطق شهری اعم از افراد و خانوارهای ساکن و یا غیرساکن در جاوا^۳ استفاده شده و بطور کلی ۳۷۰۵ مشاهده از خانوارهای نماینده در مناطق مذکور در سه دوره آمارگیری پیش‌گفته مورد استفاده قرار گرفته‌است.

1. Overall Relative Explanatory Power
2. National Social and Economic Survey (SUSENAS)
3. Java

آنها با استفاده از داده‌های آماری پیش گفته و براساس آزمون همسانی واریانس پسماندهای حاصل از برآزش منحنی انگل خانوارها را در گروه‌های مختلف درآمدی طبقه‌بندی کرده‌اند. بطوریکه در مرحله نخست خانوارها به طبقات مختلف تقسیم شده‌اند، در منحنی انگل کالاهای خوراکی و زیرگروه منتخب آن، برای هر یک از طبقات برآزش می‌شود. در مرحله بعد با استفاده از پسماندهای حاصل از برآزش منحنیهای انگل واریانس‌های جملات پسماند برای هر طبقه محاسبه و آزمون گلدفلد-کوانت در مورد همسانی واریانس‌های جملات پسماند معادلات انگل میان دو طبقه متوالی صورت گرفته است. در صورتی که واریانس هر دو طبقه متوالی با هم برابر بود، مشاهدات آن دو طبقه به یک طبقه یکسان تعلق دارند. در غیر اینصورت در طبقات شکسته شده، فرآیند فوق تکرار می‌شود تا نتیجه آزمون، به یکسانی واریانس‌ها حکم دهد و مشخصات متعلق به هر گروه تعیین گردد. این فرآیند برای هر سال تکرار شده‌است. نتیجه مطالعه جنسن و مانریکو طبقه‌بندی خانوارهای شهری اندونزی در چهار طبقه می‌باشد.

روش پیشنهادی جنسن و مانریکو، صرفنظر از اینکه محدودیتهای روش پیشنهادی جارکو را ندارد، بکارگیری آن در عمل آسانتر بوده و بسته به هدف مطالعه، چون اندازه‌گیری اثرات رفاهی اصلاحات قیمتی، یارانه‌ای و مالیاتی، طبقه‌بندی مناسبتری را ارائه می‌دهد.

طبقه‌بندی خانوارهای شهری

روشهای مرسوم فعلی برای طبقه‌بندی خانوارها در ایران، یعنی طبقه‌بندی خانوارها در دهکهای مختلف (براساس درآمد یا مخارج کل) و سپس اندازه‌گیری اثرات تغییرات قیمت، مصرف، فقر، و ... برای خانوارها، اگرچه نسبت به حالتی که در آن تحلیل با داده‌های متوسط و گروه‌بندی شده برای کل خانوارها صورت می‌پذیرد، برتری دارد، ولی دارای نقاط ضعفی است که در زیر به آن اشاره می‌شود.

طبقه‌بندی در دهکهای درآمدی یا هزینه‌ای دارای دو ضعف می‌باشد. اولاً؛ از آنجا که متغیر درآمد (یا مخارج کل) را مبنای طبقه‌بندی قرار می‌دهد، اثرات متغیرهای دیگری مانند متغیرهای اجتماعی را که مؤثر بر رفتار مصرفی مؤثر است حذف می‌نماید و دوم؛ تعیین تعداد طبقات بصورت برونزا می‌باشد که معمولاً در این روش ده طبقه در نظر گرفته می‌شود که ممکن است تعداد ده طبقه با توجه به هدف مطالعه، تناسب لازم را میان ویژگیهای طبقات با نمونه اصلی، ایجاد نکند. از اینرو وقتی که نمونه استخراج شده از جامعه به زیرنمونه‌های متمایز طبقه‌بندی شوند، تعداد خانوارها موجود در هر طبقه (دهکها) کاهش می‌یابد و کفایت آماری نتایج حاصل از هرگونه برازش و نتیجه‌گیری بر روی دهها رازیر سوال می‌برد.

روش و مدل

در این مقاله با الهام از روشهای ارائه شده در مطالعات جازکو، جنسن و مانریکو بخصوص با تأکید بر روش ارائه شده در مطالعه دوم، به طبقه‌بندی درونزای خانوارهای شهری ایران اقدام شده است. این نوع طبقه‌بندی که براساس لحاظ کردن ویژگیهای اقتصادی- اجتماعی خانوارها در قالب برآورد منحنیهای انگل قرار دارد، به لحاظ اینکه بطور همزمان از اطلاعات مذکور در مدلسازی و طبقه‌بندی استفاده می‌شود، اصطلاحاً به آن روش طبقه‌بندی درونزا می‌گویند. روش مذکور براین استدلال متکی است که خانوارهای موجود در یک گروه، دارای رفتار مصرفی مشابه هستند. بنابراین با استفاده از برآورد منحنی انگل که متأثر و یا تابعی از متغیر درآمد و جمعیتی است، آزمون برابری واریانسها برای پسماندهای معادلات برازش شده طبقه‌بندی صورت می‌پذیرد. شکل کلی مدل مورد برآورد به صورت زیر است:

$$EF = f(E; Z) \quad (4)$$

که در آن EF مخارج صرف شده برای کالای خوراکی؛ E کل مخارج و Z بردار

متغیرهای اجتماعی خانوار می‌باشد. با در نظر گرفتن اندیس i برای خانوار و افزودن جمله اختلال u_i به مدل خواهیم داشت:

$$EF_i = f(E_i, Z_i, u_i) \quad (5)$$

بطوریکه EF_i و E_i به ترتیب بردارهای ستونی N عنصر از مخارج خوراکی و مخارج کل خانوارها و Z_i نیز یک بردار ستونی در برگیرنده مؤلفه‌های اجتماعی چون سن، جنس و شغل سرپرست خانوار، وضعیت تحصیلی اعضای خانوار، ترکیب جنسی اعضای خانوار و ... می‌باشد.

به منظور برازش مدل (5) می‌توان تصریح‌های ریاضی مختلفی داشت ولی بنا بر کیفی بودن بردار Z معمولاً از تصریح خطی استفاده می‌شود که در قسمت بعدی به هنگام برازش مدل آنرا تصریح خواهیم کرد.

پس از مشخص کردن بردار Z_i در رابطه (5) تصریح ریاضی مدل و برازش آن برای هر سال خاص و انجام آزمون واریانس همسانی گلدفلد-کوانت، برحسب داده‌های بودجه خانوار، به این سؤال پاسخ داده می‌شود که آیا خانوارهای مورد نمونه دارای رفتار یکسان مصرفی هستند یا خیر؟ در صورتی که پاسخ منفی بود، می‌بایست خانوارها را در گروه‌های مستقل و متمایز طبقه بندی کرد. الگوریتم طبقه بندی خانوارها در این مطالعه به صورت زیر است:

۱. تقسیم کل خانوارها به m طبقه (m عددی صحیح و مثبت، فرد و بزرگتر از ۲

است)؛

۲. برازش مدل (5) به طور جداگانه برای تمامی m طبقه؛

۳. انجام آزمون گلدفلد کوانت به صورت گردشی برای m طبقه به صورت یک در

میان؛ در صورتیکه آزمون گلدفلد کوانت در مرحله سوم حکم بر تفاوت واریانس‌های

میان دو گروه غیر متوالی داد به مرحله چهارم می‌رویم؛

۴. تعیین طبقات فرد به عنوان تعداد طبقات و قالب اولیه طبقات؛

۵. شکستن طبقات زوج به k زیر طبقه و برآزش مدل (۵) برای هریک از آنها؛
 ۶. انجام آزمون گلد فلد کوانت برای k زیر طبقه مرحله پنجم با طبقات زوج
 مجاور آنها؛

۷. قرار دادن مشاهدات k زیر طبقه (مرحله پنجم) به طبقات فرد مرحله هفتم
 براساس نتایج مرحله ششم.

الگوریتم بالا براساس انتخاب مقدار m و k بصورت متناظر تکرار می‌شود، تا
 نتایج منجر به یکسانی واریانس‌ها برای مشاهدات قرار داده شده در هر طبقه گردد.

داده‌های آماری و نتایج تجربی

در این قسمت ابتدا داده‌های آماری بودجه خانوار برای بکارگیری در مدل (۵)
 معرفی می‌شود. سپس به برآورد و ارائه نتایج حاصل از اجرای الگوریتم پیشنهادی در
 قسمت (۱-۳) پرداخته خواهد شد.

داده‌های مورد استفاده در این مطالعه شامل اطلاعات داده‌ای بودجه خانوار
 جمع‌آوری شده توسط مرکز آمار ایران طی سالهای ۱۳۷۶ لغایت ۱۳۸۱ می‌باشد.
 داده‌های مذکور شامل چهار قسمت است؛ قسمت اول؛ خصوصیات اجتماعی اعضای
 خانوار، شامل ده مؤلفه، قسمت دوم؛ مشخصات محل سکونت و تسهیلات و لوازم
 عمده زندگی شامل هفت مؤلفه، قسمت سوم؛ هزینه‌های خوراکی و غیر خوراکی
 خانوار و قسمت چهارم؛ درآمدهای خانوار را دربرمی‌گیرد.

از میان متغیرهای اجتماعی اعضای خانوار یعنی قسمت یکم، مؤلفه‌های
 (متغیرهای) جنسیت اعضای خانوار (فقط سرپرست خانوار)، سن اعضای خانوار، پایه
 یا مدرک تحصیلی اعضای خانوار، وضعیت فعالیت و زناشویی اعضای خانوار، مرتبط و
 اثرگذار بر رفتار مصرفی خانوار (در سطح کالاهای خوراکی) شناخته شد. از مؤلفه‌های
 قسمت دوم؛ یعنی «مشخصات محل سکونت و تسهیلات و لوازم عمده زندگی»، تنها
 مؤلفه اول؛ یعنی «نحوه محل سکونت» مرتبط شناخته شد. فرآیند این انتخاب

براساس برآزش مدل‌های انگل و آزمون t برای ضرایب مؤلفه‌ها و مقایسه مقادیر F و R^2 در برآزشهای مختلف بوده‌است.

از آنجا که بعضی از مؤلفه‌های پیش‌گفته از نوع متغیرهای کیفی هستند، به‌منظور پرهیز از هرگونه مشکل اقتصادسنجی در برآورد مدل، متغیرهای مذکور بصورت متغیر مجازی (با کدهای صفر و یک) وارد مدل می‌شوند. لذا، پس از انتخاب مؤلفه‌های نامزد ورود به مدل، آنها از حالت اولیه دربرگیرنده کدهای یک رقمی ۱ تا ۹ در حالت کلی، به متغیرهای مجازی دربرگیرنده کد صفر و یک تبدیل شده‌اند. نکته مورد توجه آن است که در انتخاب پایه و مدرک تحصیلی ما شاهد یک ماتریس k بعدی (k برابر تعداد گزینه‌های پایه و مدرک تحصیلی) هستیم که برای هر سطر از ماتریس اولیه که سطرهای آن اعضای خانوار می‌باشند، قابل تعریف است و از اینرو، فرآیند کار محاسباتی را بسیار پیچیده می‌نماید. این امر قابل تعمیم برای مؤلفه‌های سن اعضای خانوار، وضعیت فعالیت و وضعیت زناشویی نیز می‌باشد. برای پرهیز از پیچیدگی‌های مترتب در امر برآورد و نحوه ورود داده‌ها به مدل و دریافت اثر صحیح متغیرهای مستقل بر روی متغیر وابسته، مؤلفه سن اعضای خانوار در پنج طبقه و مؤلفه پایه و مدرک تحصیلی نیز در پنج طبقه تقسیم‌بندی شده‌است.

متغیرهای اقتصادی - اجتماعی مؤثر بر رفتار مصرفی خانوارها - که در این مطالعه در حالت اولیه بکارگرفته شده‌است - به شرح ذیل می‌باشد:

۱. جنسیت سرپرست خانوار، کمیت یک برای مرد کمیت صفر برای زن ($D1$)؛
۲. وضعیت زناشویی سرپرست خانوار، کمیت یک برای دارای همسر و صفر برای بی‌همسر ($D2$)؛
۳. وضعیت فعالیت (اشتغال) سرپرست خانوار، کمیت یک برای شاغل و کمیت صفر برای غیر شاغل ($D3$)؛
۴. نحوه تصرف محل سکونت، این متغیر تحت دو گزینه یک برای ملکی و صفر برای غیر ملکی در نظر گرفته شده‌است. ($D4$)؛

۵. سن سرپرست خانوار ($X1$):

۶. تعداد شاغلین هر خانوار ($X2$):

۷. تعداد اعضای هر خانوار ($X3$):

۸. ترکیب سنی اعضای خانوار که بشرح زیر تقسیم‌بندی شده‌است:

● گروه سنی بین صفر تا ۵ سال ($A1$):

● گروه سنی بین ۶ تا ۱۰ سال ($A2$):

● گروه سنی مرد بین ۱۱ تا ۲۰ سال ($A3$):

● گروه سنی بین ۲۱ و بالاتر ($A4$):

● گروه سنی زن ۱۱ تا ۲۱ سال ($A5$):

● گروه سنی زن ۲۱ تا بالاتر ($A6$).

۹. وضعیت تحصیلی اعضای خانوار که به صورت زیر تقسیم‌بندی شده‌است:

● زیر دیپلم و دیپلم ($G1$):

● لیسانس ($G2$):

● فوق لیسانس ($G3$):

● دکترا و بالاتر ($G4$):

● تحصیلات حوزوی ($G5$).

۱۰. کل مخارج صرف شده روی خوراکیها، آشامیدنیها و دخانیات ($E1$):

۱۱. مخارج کل خانوار (E).

شکل عمومی منحنی انگل بصورت زیر است:

$$EF = F(E, D1, D2, D3, D4, X1, X2, X3, A1, A2, A3, A4, A5, A6, G1, G2, G3, G4, G5) \quad (۶)$$

و تصریح ریاضی آن بصورت زیر خواهد بود:

(۷)

$$E_{vi} = \alpha_0 + \alpha_1 E_i + \alpha_2 A_i + \dots + \alpha_n A_i + \alpha_{n+1} G_i + \dots + \alpha_{n+1} G_i + \alpha_{n+2} G_i + \alpha_{n+3} X_i + \dots + \alpha_{n+6} X_i + \alpha_{n+7} EM_i + u_i$$

$$u_i \sim N(0, \sigma_u^2)$$

u_i : جمله اختلال است که فرض می‌شود کلیه خواص کلاسیک را داراست؛ یعنی دارای میانگین صفر و واریانس σ_u^2 می‌باشد. α ها نیز پارامترهای مدل هستند. ابتدا معادله (۷) در شکل اولیه خود برای هر سال برازش گردید. آزمونهای مختلف نهایی فرم مدل و نوع متغیرهایی که باید در مدل باشد، تعیین گردید و نتایج حاصل از برازش مدلها بر روی کلیه مشاهدات (خانوارها) در هر سال بصورت منفرد حاکی از آن بود که از میان متغیرهای مذکور، سن سرپرست خانوار (X_1) به دلیل همخطی با متغیرهای ترکیب سنی اعضای خانوار، اثری بر نتایج مدل نداشته‌است و بهتر است حذف گردد. همچنین از گروه متغیرهای مربوط به وضعیت تحصیلی، متغیر تحصیلات حوزوی G_5 به دلیل داشتن کمیت‌های زیاد صفر، ضریب آن در تمامی مدل‌های برازش، بسیار کوچک و از نظر آماری بی‌معنی بوده‌است. بنابراین، این متغیر نیز حذف گردید. معادله (۷) با حذف متغیر X_1 و G_5 مبنای برازش مدل‌های انگل در الگوریتم پیشنهادی برای طبقه‌بندی خانوارها قرار گرفت.

فرآیند طبقه‌بندی به اینصورت انجام گرفت که ابتدا کل مشاهدات (خانوارها) هر سال در نظر گرفته شد (برای مثال؛ ۱۴۴۱۰ خانوار در سال ۱۳۸۱) و آزمون گلدفلد-کوانت بر روی مدل صورت پذیرفت. به اینصورت که ابتدا مشاهدات برحسب متغیر مخارج کل (E) از کوچک به بزرگ مرتب گردید، سپس مشاهدات به سه قسمت تقسیم گردید و با آزمون گلدفلد-کوانت وجود پدیده واریانس ناهمسانی در مشاهدات تایید گردید. سپس توسط یک زیربرنامه در محیط اجرایی نرم‌افزار TSP، فرآیند طبقه‌بندی انجام یافت که به شرح زیر می‌باشد:

۱. مشاهدات به هفت طبقه شکسته شد.

۲. طبقات میانی که اعداد صحیح مثبت و زوج به آنها اطلاق گردیده بود، از مشاهدات کنار گذاشته شد.

۳. واریانس رگرسیون‌های طبقات فرد محاسبه شد و به عنوان مبنای مقایسه در آزمون F قرار گرفت.

۴. سپس طبقات حذف شده به ۲، ۳ و ۴ زیرطبقه شکسته شده و برای هر یک از زیرطبقه‌ها واریانس محاسبه گردید.

۵. آماره F برای هر یک از زیرطبقه‌های طبقات زوج با طبقات مجاور آن محاسبه و آزمون گلدفلد-کوانت این زیرطبقات با طبقات مجاور انجام شد. فرضیه صفر و مقابل در حالت کلی بصورت زیر بوده است:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_m^2$$

در بررسی‌های انجام شده به این نتیجه رسیدیم که شکستن مشاهدات در مرحله اول به هفت طبقه واریانس همسانی برای تمامی زیرگروه‌های طبقات زوج حذف شده با طبقات مجاور آنها (طبقات فرد) ارائه نمی‌نماید. لذا کل مشاهدات به نه طبقه شکسته گردید و فرآیند ۱ تا ۵ مجدداً تکرار گردید.

پس از برآورد منحنیهای انگل و انجام آزمونهای واریانس همسانی داده‌ها، خانوارها در هر سال به پنج طبقه تقسیم شده‌اند. از آنجا که تعداد معادلات برازش شده بسیار زیاد است، در این مقاله نتایج حاصل از برازش مدلها و اجرای الگوریتم در قالب آماره‌های F و دامنه مشاهدات در هر سال ارائه شده است. نتایج آماره‌های F محاسباتی و دامنه مشاهدات برای هر یک از زیرطبقه‌ها، در سالهای مختلف مورد مطالعه در جدول (۱) آمده است.

جدول (۱): دامنه‌های مشاهده‌ها و طبقه‌بندی آنها، سالهای ۱۳۷۶ تا ۱۳۸۱

طبقه	سال ۱۳۷۶			سال ۱۳۷۷			سال ۱۳۷۸			طبقه
	دامنه	مقدار	طبقه برگزیده	دامنه	مقدار	طبقه برگزیده	دامنه	مقدار	طبقه برگزیده	
درم	۱۷۱۱-۱۶۶۲	۱,۶۲	۳	۸۷۶-۱۰۱۵	۱,۳	۱	۱۳۵۶-۱۶۹۳	۱,۶۵	۱	
	۱۶۶۳-۱۷۵۲	۱,۶۳	۳	۱۰۱۶-۱۳۱۴	۱,۵	۳	۱۰۹۴-۲۰۳۱	۱,۷۴	۳	
	۱۷۵۳-۲۰۶۶	۱,۳۶	۳	۱۳۱۵-۱۵۳۳	۱,۵۵	۳	۲۰۳۲-۱۳۶۹	۱,۶۴	۳	
	۲۰۶۷-۱۳۳۸	۱,۳۹	۳	۱۵۳۴-۱۷۵۲	۲	۳	۱۳۷۰-۲۷۰۷	۱,۳۷	۳	
	۲۵۱۱-۳۸۰۲	۱,۶۴	۳	۱۷۵۳-۱۸۷۷	۱,۲۴	۳	۴۰۶۶-۳۴۰۳	۱,۱۹	۳	
	۳۸۰۳-۴۰۴۹	۱,۳۷	۳	۱۸۷۸-۲۸۶۷	۱۲	۳	۳۴۰۴-۳۷۲۱	۱,۳۸	۳	
	۴۰۵۰-۳۳۸۶	۱,۳۱	۵	۲۸۶۸-۳۰۶۷	۱,۲۴	۵	۳۷۲۲-۵۰۷۹	۱,۳۹	۳	
	۳۳۸۷-۴۶۷۸	۱,۳۱	۵	۳۰۶۸-۳۲۸۵	۰,۵۸	۵	۵۰۸۰-۵۴۱۷	۱,۳۷	۵	
	۵۵۵۱-۴۱۲۲	۱,۱۸	۵	۳۲۸۶-۳۵۹۹	۱,۳۷	۵	۴۶۴۰-۷۱۱۳	۱,۳	۷	
	۴۱۲۳-۴۴۴۴	۱,۲۷	۵	۳۶۰۰-۳۸۱۸	۱,۱۹	۷	۷۱۱۴-۷۵۵۱	۱,۳۲	۷	
مشم	۴۴۴۵-۳۷۲۶	۱,۱۹	۵	۳۸۱۹-۵۰۳۷	۱,۲۵	۵	۷۵۵۲-۷۸۸۹	۱,۳۹	۵	
	۳۷۲۷-۷۰۱۸	۰,۵۸	۷	۵۰۳۸-۵۲۵۶	۰,۸۶	۷	۷۸۹۰-۸۱۲۷	۱,۳۴	۷	
	۸۱۱۹-۸۲۸۴	۱,۱۵	۷	۵۲۵۷-۶۱۳۱	۱,۶۶	۷	۸۱۲۸-۸۳۳۳	۱,۵۲	۷	
	۸۲۸۵-۸۷۲۴	۱,۳۸	۷	۶۱۳۲-۶۵۷۰	۱,۳	۷	۸۳۳۴-۱۰۱۶۱	۱,۵۶	۷	
	۸۷۲۵-۹۰۶۶	۱,۶۷	۷	۶۵۷۱-۶۷۸۹	۱,۷۴	۷	۱۰۱۶۲-۱۰۴۹۹	۱,۳۳	۷	
	۹۰۶۷-۹۲۵۸	۱,۵۷	۷	۶۷۹۰-۷۰۰۸	۱,۲۵	۹	۱۰۵۰۰-۱۰۸۷۷	۱,۵	۷	

ادامه جدول (۱)

طبقه برگزیده	سال ۱۳۸۱		طبقه برگزیده	سال ۱۳۸۰		طبقه برگزیده	سال ۱۳۷۹		طبقه
	مقدار آماره	دامنه		مقدار آماره	دامنه		مقدار آماره	دامنه	
۱	۱,۳۳	۱۶۰-۱۶۰۱	۱	۱,۶	۱۳۸-۱۶۵۸	۳	۱,۶	۱۳۰-۱۶۶۵	دوم
۲	۱,۳۲	۲۰۰-۱۳۰۱	۱	۱,۷	۱۶۵-۱۹۸۱	۲	۱,۶	۱۶۶-۱۹۹۲	
۳	۱,۳۲	۲۶۰-۱۸۰۱	۳	۱,۶	۱۹۹-۲۳۲۰	۲	۱,۳۲	۱۹۴-۲۳۸۹	
۳	۱,۳۱	۲۸۰-۲۲۰۱	۳	۱,۶۶	۲۳۱-۱۶۵۱	۳	۱,۳۹	۲۲۹-۲۶۱۶	
۵	۱,۳۵	۳۸۰-۵۲۰۳	۳	۱,۳۲	۳۸۲-۳۲۱۲	۳	۱,۱۶	۳۹۵-۲۵۱	چهارم
۵	۱,۳۳	۵۲۰-۵۶۰۳	۳	۱,۲۸	۳۲۱-۳۶۴۳	۵	۱,۳۸	۴۵۴-۲۵۷۸	
۵	۱,۳۲	۵۶۰-۶۰۰۳	۳	۱,۲۵	۳۶۴-۳۹۷۳	۵	۱,۲۱	۴۵۷-۲۹۰۵	
۵	۱,۳۳	۶۰۰-۶۴۰۳	۵	۱,۲۱	۳۹۷-۵۳۰۵	۵	۱,۱۲	۴۹۰-۵۳۲۶	
۵	۱,۱۹	۸۰۰-۸۴۰۵	۵	۱,۱۲	۳۹۷-۳۹۶۶	۵	۱,۲۸	۶۵۴-۳۹۶۷	ششم
۷	۱,۳۶	۸۴۰-۸۸۰۵	۵	۱,۱۸	۳۹۷-۳۹۷	۵	۱,۲	۶۸۸-۷۱۹۲	
۷	۱,۲۷	۸۸۰-۹۲۰۵	۵	۱,۱۶	۷۱۵-۶۶۳۸	۵	۱,۲۱	۷۱۵-۷۵۱۱	
۷	۱,۳۳	۹۲۰-۹۶۰۵	۷	۱,۲۸	۶۶۳-۶۶۳۸	۷	۱,۳۲	۷۵۲-۷۸۶۸	
۷	۱,۳۶	۱۱۲۰-۱۱۶۰۷	۷	۱,۱۲	۹۱۹-۹۶۲۰	۷	۱,۱۹	۹۱۵۷-۹۲۸۳	
۷	۱,۲۹	۱۱۶۰-۱۲۰۰۷	۹	۱,۰۹	۹۶۲-۹۵۵۱	۷	۱,۲۱	۹۲۸۳-۹۸۱۰	
۹	۱,۳	۱۲۰۰-۱۲۴۰۷	۷	۱,۱	۹۵۴-۱۰۸۱۲	۷	۱,۷	۹۸۱۱-۱۰۱۳۷	
۹	۱,۶	۱۲۴۰-۱۲۸۰۷	۷	۱,۲۹	۱۰۸۳-۱۰۹۱۲	۷	۱,۵۳	۱۰۱۳۸-۱۰۶۶۶	

ویژگیهای اقتصادی و اجتماعی خانوارها در طبقات مختلف

جدول (۲) متغیرهای اجتماعی خانوارهای شهری ایران در طبقات مختلف را ارائه می‌کند. چنانچه تعداد خانوارهای قرار داده شده در هر طبقه به عنوان معیار اولیه در توزیع مشاهدات (خانوارها)، در بین طبقات مدنظر قرار گیرد، ملاحظه می‌شود که داده‌های طبقات میانی (دوم، سوم، چهارم) بیشترین فراوانی را داشته و طبقات ابتدایی و انتهایی نیز، کمترین فراوانی را نشان می‌دهند. طبقات اول تا پنجم برحسب میزان درآمدها، از پائین‌ترین به بالاترین میزان درآمد قرار گرفته‌اند. تعداد افراد در هر طبقه با حرکت از طبقات اول تا چهارم افزایش می‌یابند، ولی در طبقه پنجم با کاهش مواجه می‌گردند. نسبت تعداد شاغلین به کل جمعیت هر طبقه تقریباً در بین طبقات یکسان است.

جدول ۲. متغیرهای اجتماعی خانوارهای شهری ایران
طی دوره (۱۳۸۱-۱۳۷۶)

عنوان متغیر	طبقه	اول	دوم	سوم	چهارم	پنجم	کل
تعداد خانوارها		۹۲۵۶	۱۷۰۰۱	۱۵۹۲۳	۱۷۵۱۶	۸۹۸۶	۶۸۶۹۳
درصد از کل خانوارها		۱۳/۳۷	۲۴/۷۵	۲۳/۲	۲۵/۵	۱۳/۰۸	۱۰۰
تعداد افراد هر طبقه (نفر)		۳۰۸۶۱	۷۳۶۳۰	۷۵۹۸۰	۸۸۳۱۸	۳۵۱۵۷	۳۱۳۹۳۶
تعداد شاغلین (نفر)		۷۳۵۳	۱۸۳۹۱	۱۹۰۳۶	۲۲۵۹۵	۱۲۲۲۳	۷۹۶۰۸
درصد تعداد شاغلین به کل جمعیت همان طبقه		۲۳/۸	۲۴/۶	۲۵/۱	۲۵/۶	۲۷/۱	۲۵/۳
متوسط سن سرپرست خانوار		۳۹/۶	۳۴/۶	۳۴/۹۶	۳۶	۳۷/۶	۳۶/۵۵
تعداد متأهلین (سرپرست خانوار)		۶۶۶۷	۱۵۰۶۶	۱۴۶۳۳	۱۶۳۹۰	۸۳۵۶	۶۱۲۷۹
درصد متأهلین (سرپرست خانوار) به کل خانوارهای هر طبقه		۷۲	۸۹	۹۲	۹۳	۹۳	۸۹
تعداد مالکین هر طبقه		۵۳۶۱	۱۱۱۰۱	۱۱۳۹۷	۱۳۳۳۱	۷۰۷۳	۴۸۴۴۳
درصد مالکین به کل خانوارهای هر طبقه		۵۹	۶۵	۷۲	۷۶	۷۹	۷۰

متوسط کل سن سرپرست خانوار برابر ۴۶/۵ سال می‌باشد که می‌توان گفت توزیع میانگین آن در میان طبقات تقریباً یکنواخت می‌باشد. کمترین میانگین سنی سرپرست خانوار متعلق به طبقه دوم برابر ۴۴/۶ سال و بیشترین میانگین سنی، مربوط به طبقه اول با ۴۹/۶ سال است.

جدول (۳) متغیرهای اقتصادی خانوارهای شهری ایران (درآمد یا مخارج کل) را طی دوره ۱۳۸۱-۱۳۷۶ نشان می‌دهد. براساس شاخص میانگین درآمد (طی دوره ۱۳۷۶ تا ۱۳۸۱) ملاحظه می‌شود که طبقه اول، با درآمدی معادل ۶ میلیون ریال در پایین‌ترین گروه درآمدی جای دارند و پس از آن، طبقه دوم با ۱۱ میلیون ریال؛ طبقه سوم ۱۸ میلیون ریال؛ طبقه چهارم ۲۷ میلیون ریال و بالاخره، طبقه پنجم با میانگین ۴۸ میلیون ریال پردرآمدترین طبقه خانوارهای شهری قلمداد می‌شوند. براساس ضرایب پراکندگی متغیر درآمد درمی‌یابیم که کمترین پراکندگی در طبقه سوم و بیشترین پراکندگی نیز مربوط به طبقه اول می‌باشد؛ اگر چه این ضریب در میان طبقات دوم تا پنجم اختلاف بسیار اندکی دارد.

جدول ۳. متغیرهای اقتصادی خانوارهای شهری ایران

طی دوره (۱۳۷۶-۱۳۸۱)

عنوان متغیر	طبقه	اول	دوم	سوم	چهارم	پنجم
تعداد خانوارها		۹۲۵۶	۱۷۰۰۱	۱۵۹۲۳	۱۷۵۱۶	۸۹۸۶
حداقل درآمد (ریال)		۲۷۸۷۲۳	۳۵۳۸۳۹۸	۸۸۲۳۹۱۱	۱۳۱۲۷۷۷۰	۲۲۲۲۸۸۳۳
حداکثر درآمد (ریال)		۱۲۱۲۲۹۲	۱۹۰۰۷۹۳۹	۲۸۵۸۲۳۷۳	۳۵۶۸۳۱۱۱	۹۲۳۹۰۵۷۳
دامنه تغییرات (ریال)		۱۱۸۲۳۵۶۹	۱۳۴۵۹۵۵۱	۱۹۷۵۸۵۶۳	۲۲۵۵۶۳۳۱	۶۹۹۶۱۷۳۰
میانگین درآمد (ریال)		۶۰۳۰۱۲۹	۱۱۲۶۳۸۵۷	۱۷۸۵۷۶۳۵	۲۶۹۳۱۸۵۶	۳۷۹۲۶۷۶۵
انحراف معیار درآمد		۲۶۸۳۵۸۱	۳۵۱۹۷۳۶	۵۰۱۹۶۲۲	۸۰۱۶۵۲۳	۱۵۵۸۲۰۰۵
ضریب پراکندگی درآمد		۰/۳۳	۰/۳۱	۰/۲۸	۰/۲۹	۰/۳۲
چولگی درآمد		۰/۳۱	۰/۲۳	۰/۱۰	۰/۲۸	۰/۲۸

نتیجه‌گیری

در این مقاله ابتدا به ضرورت طبقه‌بندی خانوارها در گروه‌های مستقل و متمایز به منظور شناخت صحیح رفتار مصرف‌کنندگان و ارزیابی سیاست‌های حمایتی، تأمین اجتماعی، پرداخت یارانه و کاهش آن و نیز اعمال مالیات و هرگونه سیاست اصلاح قیمتی اشاره شد.

برای پاسخ به این ضرورت، از روش جنسن و مانریکو در طبقه‌بندی خانوارها استفاده کرده و روشی برای طبقه‌بندی خانوارهای ایرانی با استفاده داده‌های مرکز آمار ایران ارائه می‌گردید. روش مذکور را با استفاده از داده‌های خام بودجه خانوار مرکز آمار ایران طی دوره ۱۳۷۶ تا ۱۳۸۱ برای خانوارهای شهری ایران اجرا نمودیم و خانوارها را در پنج طبقه مستقل و متمایز گروه‌بندی کردیم.

نتایج حاصل از اجرای روش پیشنهادی حاکی از آن است که بیشترین تعداد خانوارها در طبقه چهارم و پس از آن نیز در طبقه دوم قرار دارند. کمترین تعداد خانوارها در طبقه اول قرار گرفته‌اند. طبقات اول تا پنجم، به ترتیب از سطح پائین درآمدها به سوی سمت بالای درآمد چیده شده‌اند. بگونه‌ای که میانگین درآمد سالانه طبقه اول طی دوره مورد بررسی (۱۳۸۱-۱۳۷۶) ۶ میلیون ریال، میانگین درآمد طبقه دوم؛ ۱۱ میلیون ریال و میانگین درآمد طبقه پنجم نیز ۴۸ میلیون ریال بوده‌است؛ به عبارت دیگر طی دوره مورد بررسی میانگین درآمد طبقه پنجم حدود ۸ برابر طبقه اول، $\frac{4}{2}$ برابر طبقه دوم، $\frac{2}{7}$ برابر طبقه سوم و $\frac{1}{8}$ برابر طبقه چهارم بوده‌است.

فهرست منابع:

۱. مرکز آمار ایران، ویرسشنامه آمارگیری از هزینه و درآمدی خانوار (شهری)، سالهای ۱۳۷۶ تا ۱۳۸۱.
۲. مرکز آمار ایران، «آمار خام طرح بودجه خانوار»، سالهای ۱۳۷۶ تا ۱۳۸۱.
3. Burney, N., and Akmal, M. "Food Demand in Pakistan: An Application of the Extended Linear Expenditure System". *Journal of Agricultural Economics*, Vol.42, (1991):185-195.
4. Deaton, A., and Muellbauer, J. "Economics and Consumer Behavior". Cambridge University Press, New York., 1980.
5. Jarque, C.M. "Sample Splitting and Applied Econometric Modeling". *Journal of Business and Economic Statistics*, Vol.23, (1978).
6. Jarque, Carlos M. "An Application of LDV Models of Household Expenditure Analysis in Mexico", *Journal of Econometrics*, Vol.36, (1987):31-53.
7. Jensen, H.H. and Justo Manrique. "Demand for Commodities by Income Groups in Indonesia". *Applied Economics*, Vol.30, (1998):491-501.
8. Jones, E., and Mustiful, B. "Purchasing Behavior of Higher and Lower-Income Shoppers: A look at Breakfast Cereal". *Applied Economics*, Vol.28, (1987):31-53.
9. Pinstrup- Andersen, P., Lodono, N., and Hoover, E. "The Impact of increasing Food Supply on Human Nutrition: Implication for Commodities Priorities in Agricultural Research and Policy". *Agricultural Economics*, Vol.58, (1976):131-42.
10. Pinstrup- Andersen, P., and Caicedo, F. "The potential Impact of change in Income Distribution on Food Demand in Human Nutrition". *American Journal of Agricultural Economics*, Vol.60, (1978):402-15.
11. Savadogo, K., and Brandt, J. "Households Food Demand for Burkina Faso: Implication for food Policy". *Agricultural Economics*, Vol.2, (1988):345-64.
12. Spark, D.M. "Euclidean Cluster Analysis". *Journal of the Royal Statistical Society*, C22, (1973):126-130.
13. Teklu, T., and Johnson, S. R. "Demand System for Cross Section Data: An Application to Indonesia". *Canadian Journal of Agricultural Economics*, Vol.36, (1988):83-101.
14. "Time Series Processor", Version 4.2 Reference Manual October 1991.
15. "Time Series Processor", Version 4.2 User's Manual October 1991.