انشو شناسایی قطبهای همگنن از استانهای کشور بر مبنای شاخصهای بهداشتی و جمعیتی به کمـک تکنیـک آماری خوشهبندی فازی

نویسندگان: دکتر غلامرضا بابایی ۱، دکتر سقراط فقیهزاده ۱ و آوات فیضی ^۳

- ۱. دانشیار گروه آمارزیستی دانشکده علوم پزشکی دانشگاه تربیت مدرس
- ۲. دانشیار گروه آمارزیستی دانشکده علوم پزشکی دانشگاه تربیت مدرس
 - ۳. دانشجوی دوره دکتری آمار زیستی دانشگاه تربیت مدرس

چکیده

تکنیک آماری خوشهبندی که در مورد دادههای چند متغیره قابلیت کاربرد دارد با بهکارگیری اصول و روشهای خاص خود، منجر به شناسایی گروههای واقعی مشابه و همگن از افراد، اشیا و بهطور کلی آزمودنیها بر مبنای ویژگیهای ثبت یا اندازهگیری شده در صورد آنها میشود. به طور کلی سه دسته اصلی از روشهای خوشه بندی عبارتند از: سلسله مراتبی، غیرسلسله مراتبی و فازی که اخیراً روش خوشهبندی فازی به علت ویژگیهای برجستهاش که منجر به ارائه اطلاعات بیشتر از ساختار مشاهدات و رفع نواقص موجود در دو روش دیگر میشبود، کاربرد وسیعی پیدا کرده است. در مقاله حاضر با بهرهگیری از تکنیک خوشهبندی فازی، اقدام به منطقه بندی کشور یا در واقع شناسایی استانهای همگن به لحاظ شاخصهای بهداشتی و جمعیتی شده است. در این راستا، چهار دسته شاخص بهداشتی و جمعیتی، شامل ۱) شاخصهای تسهیلات ــــتي و رفـــــاهي خــــ ۲) شاخصهای تنظیم خانواده، ۳)" شاخصهای باروری و زایمان زنان، ۴)" شاخصهای اجتماعی و بهداشتی کودکان برای شناسایی استانهای همگن مورد استفاده قرار گرفته است. این شاخصها از طرح DHS (طرح سنجش شاخصهای بهداشتی و جمعیتی کشور) که توسط وزارت بهداشت و

دوماهنامه علمي -

درمان و آموزش پزشکی اجرا گردیده، استخراج شده است. بر مبنای هر کدام از دسته شاخصهای یاد شده، قطبهای همگن از استانها به کمک تکنیک خوشهبندی فازی شناسایی گردیدهاند.

نتایج این تحقیق در باب راهبردهای عملی درمورد اجرا وارائه خدمات بهداشتی، متناسب با وضعیت قطبهای همگن شناسایی شده، برای دست اندرکاران امور بهداشتی و درمانی کشور مفید خواهدىود.

واژههای کلیدی: خوشهبندی فازی، شاخصهای بهداشتی و جمعیتی، انفکاک و فشیردگی خوشهها، DHS

دوماهنامه علمي - پژوهشي دانشور پزشكي /دانشگاه شاهد /شهريور ۲۰ سال يازدهم / شماره ۱۲

۱- مقدمه

معمولاً دسته بندی و شناسایی گروهای همگن و مشابه از افراد، اشیا و بهطور کلی آزمودنی ها براساس بررسی وجوه تشابه و تمایز آن ها صورت می گیرد، بدین صورت که برای هر یک از آزمونی ها مشخصه های متعددی وجود دارد که آنها را می توان براساس قرابتی که از لحاظ این ویژگیها باهم دارند در دستههای همگن و مشابه گروه بندی کرد [۱]. چنین کاری می تواند به صورت ذهنی صورت گیرد، اما در حالتی که تعداد آزمودنی ها و صفات اندازه گیری شده در مورد آنها زیاد باشد مطمئناً نتایج مفید و قریب به واقعیت را به دست نخواهد داد [۲]. تکنیک آماری خوشه بندی، بهویژه خوشهبندی به روش فازی که در مورد دادههای وسيع و حجيم قابل كاربرد است چنين مقايسه هايي را برای شناسایی آزمودنیهای همگن ازطریت محاسبات ریاضیاتی و منطقی انجام داده، منجر به شناسایی و انجام یک گروهبندی واقعی میشود [۳]. سه روش اصلی خوشه بندی که در علم آمار بسط و گسترش داده شدهاند عبارتند از: سلسله مراتبي، غيرسلسله مراتبي و فازی. دو روش اول در دسته بندی آزمودنی های مشابه، منجر به تشکیل نتایج قطعی می شوند، بدین معنا که اعضایی که داخل یک خوشه قرار می گیرند، صرفاً فقط به آن خوشه تعلق می گیرند و لاغیر، درحالی کـه وجـه تمایز و ویژگی برجسته خوشهبندی فازی این است که با ارائه یک سری ضرایب تعلق موجب می شود که اعضای داخل یک خوشه بتوانند به خوشههای دیگر نیز تعلق داشته باشند؛ اما شدت تعلق آنها به خوشههای دیگر نسبت به میزان تعلق به خوشهای که در آن قرار گرفتهاند کم تر است. با چنین عملکردی، این روش، یک مشکل اساسی مرتبط با روش های خوشهبندی

قطعی، یعنی اشتباه ردهبندی کردن آزمونیها (misclassification) را از بین میبرد [۴].

در این مقاله با به کارگیری روش خوشه بندی فازی C-means و جدیدترین روش تعداد بهینه و ارزیابی اعتبار خوشه ها، استانهای همگن به لحاظ شاخصهای بهداشتی و جمعیتی شناسایی و دسته بندی گردیده اند. شاخصهای مورد استفاده جهت تشکیل خوشه های همگن و مشابه از استانهای کشور با مراجعه به طرح که درمهر ماه ۱۳۷۹ توسط وزارت بهداشتی و جمعیتی کشور) که درمهر ماه ۱۳۷۹ توسط وزارت بهداشت، درمان و آموزش پزشکی اجرا شد [۵] استخراج گردیده که در قالب چهار دسته مشخصه های بهداشتی و جمعیتی مناطق شهری ۲۸ استان و شهر تهران (جدا از استان مناطق شهری ۲۸ استان و شهر تهران زیر هستند:

شاخصهای تسهیلات بهداشتی و رفاهی خانوارها شامل ۷۳ متغیر؛

۲. شاخصهای باروری و زایمان زنان شامل ۴۹ تغیر؛

۳. شاخصهای تنظیم خانواده شامل ۱۳۱ متغیر؛

۴. شاخصهای اجتماعی و بهداشتی کودکان شامل۱۰۶ متغیر.

بر مبنای هر یک از دسته شاخصهای فوقالذکر، استانهای مشابه وهمگن شناسایی شده اند. دسته بندی استانها بر مبنای این شاخصها منجر به امکان بررسی دقیق تر وضعیت هر یک از خوشهها و شناسایی نقاط قسوت وضعف استانهای واقع در آنها از نظر شاخصهای مورد بررسی و در نتیجه امکان برنامه ریزی دقیق تر و مناسب تر، می گردد و به فراخور وضعیت هر استان، زمینه ارائه و اجرای برنامههای خدمات بهداشتی و درمانی را فراهم می آورد.

۲- مواد و روشها

روش تجزیه و تحلیل خوشه بندی فازی برای اولین بار توسط راسپینی (Ruspini) ابداع شد و بعدها توسط بزدک (Bezdek) بسط و گسترش یافت. اصولاً پایهریزی الگوریتم خوشهبندی فازی جهت فائق آمدن بر وضعیت بروز مشكل همپوشي خوشه از طريق وجود آزمودنی هایی که در تعلق آنها به خوشههای مختلف ابهام به وجود مى آيد صورت گرفت. مشكل مزبور يك ایراد اساسی است که در اجرای روشهای خوشهبندی سلسله مراتبی و غیرسلسله مراتبی بروز می کند [۳]. اساس خوشهبندی فازی C-means بر تقسیم m آزمودنی (در این جا ۲۹ استان) به c خوشه از طریق محاسبه و تعیین ضریب عضویت هر آزمودنی به خوشههای مختلف است. در خوشه بندی فازی، مجموعه ای از ضرایب عضویت که شدت تعلق آزمودنی ها به خوشههای مختلف را نشان می دهد وجود دارند. برای مثال اگر خوشهبندی فازی منجر به تشکیل چهار خوشه گردد برای هر آزمودنی واقع در هر یک از خوشهها، چهار ضریب عضویت وجود خواهد داشت که شدت تعلق آن را به هر یک از خوشهها بیان می کند. از این دیدگاه، خوشه بندی سلسله مراتبی و غیرسلسله مراتبی را که منجر به نتایج قطعی می شوند می توان حالت خاصی از خوشهبندی فازی در نظر گرفت که ضریب عضویت برای تعلق به یکی از چهار خوشه مذکور «یک» و برای بقیه «صفر» خواهد بود [۶]؛ بدین معنا که اين عضو قطعاً به آن خوشه تعلق دارد و لاغير. وجه مشخصهٔ خوشهبندی فازی و غیرفازی وجود همین ضرایب عضویت در خوشهبندی فازی و عملکرد خوشهبندی فازی از طریق ضرایب یادشده در رفع كردن مشكل هميوشي خوشهها است.

۱-۲- الگوريتم خوشهبندي فازي

عمومی ترین و کاربردی ترین الگوریتم خوشه بندی فازی، الگوریتم C-means است که در آن، یک تابع فازی، الگوریتم میران $J_m(u,v) = \sum_i \sum_k (u_{ik})^m d^2(x_k,v_i)$ هدف به فرم

 $U = [u_{ik}] \in R^{cn}$ نسبت به ماتریس ضرایب عضویت در یک فرایند تکراری کمینه می گردد.

در تابع هدف فوق u_{ik} ضریب عضویت عنصر X_k از مرکز به خوشه i ام و فاصله اقلیدسی عنصر X_k از مرکز خوشه فازی V_i و i درجه فازی بودن است که خود ضریب عضویت، تابعی معکوس از مجموع فواصل عناصر مختلف از مراکز مختلف خوشهها و همچنین مراکز خوشهها تابعی از مجموع ضرایب عضویت مراکز خوشه تابعی از مجموع ضرایب عضویت الگوریتم در این مقاله، تابع هدف مزبور بر طبق یک الگوریتم در نرم افزار Matlab برنامه نویسی گردید و با تکرار ۱۰۰۰۰ مر تبه و دقت بیش تر از جهت خوشه بندی و شناسایی استانهای همگن براساس هر یک از دسته شاخصهای نامبرده کمینه گردید [V].

ذکر این مطلب ضروری است که خوشهها و ردههایی دارای اعتبار هستند که دارای دو ویژگی اساسی

المنفک از هم باشند، ۲) فشرده باشند (اعضای بسیار مشابه داخل خوشهها قرارگرفته باشید) [۱]. جهت بررسی برقراری این ویژگیها درمورد خوشههای تشکیل شده توسط الگوریتم فازی که پیشتر توضیح داده شد. از جدیدترین میلاک ارائه شده در این خصوص استفاده گردید و با استفاده از آن، تعداد بهینه خوشهها و نیز اعتبار آنها مورد ارزیابی قرار گرفت. جهت اجرای این ملاک در مورد دادهها، الگوریتم این ملاک نیز در محیط نرم افزار هاهاها برنامه نویسی گردید و با تکرار ۱۰۰۰۰ مرتبه ودقت بیش تر از (۱۰-) ملاک مزبوربیشینه شد.

ملاک تعیین تعداد بهینه خوشهها و ارزیابی اعتبار آنها، ترکیبی از دو تابع است که تابع اول (SC_1) نسبت انفکاک به انسجام خوشهها را برمبنای ضرایب عضویت و خواص ساختار هندسی دادهها، و تابع دوم (SC_2) نسبت جدایی خوشهها به فشردگی آنها را صرفاً براساس ضرایب عضویت ارزیابی می کند. ملاک نهایی با استفاده از تفاضل دو تابع بهصورت SC_1 که باید بیشینه گردد تعریف می شود SC_1 .

خوشهبندی استانها بر مبنای شــاخصهــای تســهیلات بهداشتی رفاهی خانوار

با بهره گیری از الگوریتم خوشهبندی فازی C-means در مورد ۷۳ متغیر و شاخص مربوط به تسهیلات بهداشتی و رفاهی خانوار، شناسایی استانهای همگن به لحاظ این شاخصها به تشکیل ۷ خوشه که اعضای هریک ومیزان تعلق آنها به خوشههای مربوط به شرح زیر است منجر گردید:

٣- يافتههاي تحقيق

جهت شناسایی استانهای همگن به لحاظ هر یک از چهار دسته شاخص بهداشتی و جمعیتی که اسم برده شد از الگوریتم خوشهبندی فازی داری C-means استفاده و گردید. تابع هدف فازی و تابع تعیین تعداد بهینه و ارزیابی خوشهها با تکرار ۱۰۰۰ مرتبه و دقت (۱۰- ویشینه گردیدند که نتیجه اجرای آنها در دسته بندی استانها براساس هر یک از دسته شاخصهای بهداشتی و جمعیتی به شرح زیر است:

جدول ۱ نتایج خوشهبندی (قطبهای همگن) استانهای کشور بر مبنای شاخصهای تسهیلات بهداشتی رفاهی خانوار

					1		1						
ميزان	خوشه	ميزان	خوشه	ميزان	خوشه	ميزان	خوشه	ميزان	خوشه	ميزان	خوشه	ميزان	خوشه
تعلق	هفتم	تعلق	ششم	تعلق	پنجم	تعلق	چہارم	تعلق	سوم	تعلق	دوم	تعلق	اول
٠./٣٠۴	مازندران	-/941	چهارمحا	٠/٣٧٨	گیلان	٠/٨۶٣	آ-شرقی	٠/٧٩٣	كرمانشاه	•/۲۶۶	سیستان و	٠/۴	مر کزی
			ل و								بلوچستا		
			بختياري								ن		
•/٢٨٢	آ-غربی	•/ ٣٢٧٣	يزد	•/٣٣•	كهگيلويه	•/497	اصفهان	٠/٢٩۶	خوزس	•/٣۴٣	ايلام	1/097	خراسان
									تان				
٠/٢١٣	فارس	٠/٢١٣	فارس	·/۲۹Y	اردبيل	•/٣۴٩	كردستا	٠/۴٨	كرمان	•/٨١٨	بوشهر	-/۵۸۲	همدان
							ن						
٠/٣٠١	لرستان			•/494	قزوين	٠/۶۵۵	زنجان			•/٧۴۶	هرمزگا	٠/۴٧٨	سمنان
											ن		
٠/٢٩۶	استان تهران			٠/٣٧٩	گلستان	٠/٣٠۴	قم					٠/٣٢٥	شهر تهران

جدول ۲ نتایج خوشهبندی (قطبهای همگن) استانهای کشور بر مبنای شاخصهای باروری و زایمان زنان

ميزان تعلق	خوشه ششم	ميزان تعلق	خوشه پنجم	ميزان تعلق	خوشه چها <i>ر</i> م	ميزان	خوشه سوم	ميزان	خوشه	ميزان	خوشه
						تعلق		تعلق	دوم	تعلق	اول
-/٢۵۵	آ-شرقی	•/۲۴۴	مر کزی	•/٢٢١	كردستان	-/٢٨٩	كرمان	•/٣٢۴	آ-غربی	•/٢۶٩	سیستان و بلوچستان
•/199	چهار محال وبختياري	•/٢۴1	گیلان	-/٢٢	استان تهران (بدون شهر تهران)	·/۵۲Y	خراسان	•/۶٨٩	خوزستان	•/ ٩٣ Y	هرمز گان
+/480	لرستان	•/٢۴۶	مازندران			•/٢٣۴	کهگیلویه وبویراحمد	•/٢١۴	يزد		
•/ ۲۳ Υ	ايلام	•/۲۴•	كرمانشاه			-/٢١٥	بوشهر	+/٢۵۴	قه		
./٢۵۶	زنجان	·/TT1	فارس					•/٣۴٣	گلستان		
•/199	سمنان	+/ ۲ ۴ Y	اصفهان								
·/٣1٩	اردبيل	•/٢۴٨	همدان								
		٠/٢۴۵	قزوين								

جدول ۳ نتایج خوشه بندی (قطبهای همگن) استانهای کشور بر مبنای شاخصهای تنظیم خانواده

خوشه اول	گیلان، مرکزی، مازندران، فارس، کرمان، خراسان، اصفهان، زنجان، سمنان، یزد، تهران، بـدون شـهر تهران، شـهر
	تهران، قم، قزوین و گلستان
خوشه دوم	آ- شرقی، آغربی، کرمانشاه، خوزستان، سیستان و بلوچستان، کردستان، همدان، چهارمحال وبختیاری، لرستان،
	ایلام، کهگیلویه وبویراحمد، بوشهر، هرمزگان واردبیل

خوشه بندی استان ها بر مبنای شــاخص هــای بــاروری و زایمان زنان

در شناسایی استانهای همگن به لحاظ این شاخصها، الگوریتم خوشهبندی فازی C-means روی ۴۹ متغیر مربوط به وضعیت باروری و زایمان زنان اجرا گردید. نتیجه اجرای این الگوریتم تشکیل شش خوشه بود که ساختار و اعضای آنها به شرح جدول ۲ است.

خوشهبندى استانها برمبناى شاخصهاى تنظيم خانواده بررسی و شناسایی استانهای همگن از نظر شاخصهای تنظیم خانواده که دربردارنده ۱۳۱ متغیر مربوط به اجرای انواع روشهای تنظیم خانواده در مناطق شهری استانهای مختلف کشور بود از طریق اجرای الگوریتم خوشه بندی فازی منجر به تشکیل دو خوشه از استانها به شرح جدول ۳ گردید.

شایان ذکر است، میزان تعلق هریک از استانها به خوشههای مربوط حدود ۰/۵ است.

خوشهبندی استانهای کشور بـر مبنـای شـاخصهـای اجتماعي و بهداشتي كودكان

در شناسایی استانهای همگن از نظر شاخصهای اجتماعی و بهداشتی کودکان در مناطق شهری ۱۰۶ متغیر مربوط به مسائل بهداشتی جمعیتی و مرگ ومیر کودکان از طرح DHS استخراج گردیدو در الگوریتم خوشهبندی فازی مورد استفاده قرار گرفت که نتیجه پیادهسازی الگوریتم در دسته بندی استان ها و شناسایی قطبهای متجانس منجر به تشکیل پنج خوشه از آنها به شرح جدول ۴ گردید.

جدول ٤ نتایج خوشهبندی (قطبهای همگن) استانهای کشور بر مبنای شاخصهای اجتماعی بهداشتی کودکان

ميزانتعلق	خوشه	ميزان تعلق	خوشه چهارم	ميزان تعلق	خوشه	ميزان تعلق	خوشه دوم	ميزان تعلق	خوشه اول
	پنجم				سوم				
٠/٣١٩	كرمانشاه	·/۲1 ·	مرکزی	•/۲۶•	كرمان	٠/٢٠۵	فارس	•/489	آ-غربی
•/٢٨١	خوزستان	-/۲۲۳	گیلان	•/٢٧١	سیستان و بلوچستان	•/٢•٩	بوشهر	-/۵۲۳	كردستان
•/٢٨۶	کهگیلویه و بویراحمد	٠/٢٠٣	مازندران	-/191	خراسان	•/199	زنجان	•/٣۶٩	همدان
•/٢٣٨	اردبيل	-/۲1-	آ-شرقی	•/199	ايلام	٠/٢٠١	تهران بدون شهر تهران	•/٣••	لرستان
		•/٢••	اصفهان	·/۲۱۴	هرمز گان	-/147	گلستان		
		-/٢١٩	چهارمحال و بختیاری						
		-/٢١٩	يزد						
		٠/٢١٣	قم						
		•/19٢	قزوین						

٤- بحث و نتیجهگیری

در این تحقیق با هدف قطب بندی استان های کشور بر مبنای شاخص های بهداشتی و جمعیتی از روش خوشهبندی فازی جهت ردهبندی استانها برمبنای چهار دسته شاخص ۱) شاخص های تسهیلات بهداشتی رفاهی خانوار، ۲) شاخص های باروری و زایمان و

۳) شاخصهای تنظیم خانواده، ۴) شاخصهای اجتماعی و بهداشتی کودکان استفاده گردید و استانهای همگن به لحاظ هریک از این دسته شاخصها شناسایی شدند. نتیجه این ردهبندی در بخش یافته های تحقیق در جداول مربوط ذکر گردیده است. همان گونه که در جداول ملاحظه می شود هر خوشه دربردارنده چند استان است و برای هر استان ضریب

دوماهنامه علمي - پژوهشي دانشور پزشكي / دانشگاه شاهد / شهريور ۸۳ / سال يازدهم / شماره ۵۲

معضلات آن بهتر و مناسب تر فراهم می آورد. ایس امر گامی مهم درجهت توزیع عادلانه و متناسب تر امکانات و خدمات بهداشتی درمانی که یکی از اجزای اصلی توسعه یایدار و همه جانبه است، محسوب می گردد.

۵– منابع

- Hoppner F., Kalwon F., Kruse R., Runkler T.; Fuzzy cluster analysis, Methods for classification, data analysis and image recognition, Wiley & sons (1999).
- 2. Johnson R., Wichern D., Applied Multivariate statistical analysis, Wiley & sons (1999).
- T.W. Chenge, D.B. Goldfer, L.O. Hall. Fast fuzzy clustering. fuzzy set and system93 (1998), 55.
- Kamel M.S., Selim S.Z., New algorithm for solving the fuzzy clustering problem, Pattern recognition (1994) 27:421-428.
- ۵. کتابنامه سیمای جمعیت و سلامت جمهوری اسلامی ایران
 (آمار و اطلاعات طرح DHS) معاونت سلامت وزارت
 بهداشت، درمان و آموزش پزشکی اجرا مهرماه ۱۳۷۹.
- Krishnapurama R., Keller J., The probabilistic C-Means Algorithm: Insights and Recommendation. IEEE Trans. fuzzy systems (1996) 4:385-395.
- Gath I., Geva A.B. Fuzzy clustering of Elliptic ring shaped clusters. Pattern recognition letters (1995) 16:727-741.
- Zahid N., Limouri M., Essaid A., A new cluster validity for fuzzy clustering, Pattern Recognition, (1999) 32:1089-1097.

عضویتی که میزان و شدت تعلق آن استان را به خوشه مزبور بیان می کند آمده است. با توجه به ماهیت عملکرد خوشهبندی فازی، استانهایی که در خوشههای تشکیل شده قرار گرفتهاند قطعاً و بهطور کامل متعلق به این خوشهها نیستند و می توانند به خوشههای دیگر نیز تعلق داشته باشند که البته شدت تعلق آنها به خوشههای دیگر از شدت تعلق به خوشهای که در آن واقع شدهاند کم تر است (جزئیات فرایند خوشهبندی و واقع شدهاند کم تر است (جزئیات فرایند خوشهبندی و جمعیتی و جمعیتی و جمعیتی و جمعیتی و جمعیتی و جمعیتی در تحقیقی که در دانشگاه تربیت مدرس انجام گردیده تحقیقی که در دانشگاه تربیت مدرس انجام گردیده آمده است).

شناسایی و ردهبندی استانهای همگن به لحاظ شاخصهای بهداشتی و جمعیتی، امکان مطالعه جامع تر و دقیق تر ویژگیهای مزبور را در قطبهای تشکیل شده فراهم می کند. این امر منجر به شناسایی نقاط قوت و ضعف هر یک از قطبها از نظر شاخصهای بهداشتی و جمعیتی خواهد شد که بالتبع امکان برنامه ریزی دقیق تر به تناسب ویژگیهای هر قطب را در جهت اجرای برنامههای بهداشتی و درمانی و رفع